



ITS
Institut
Teknologi
Sepuluh Nopember

TUGAS AKHIR - IS 184853

PERAMALAN KEJADIAN LUAR BIASA PENYAKIT DEMAM BERDARAH DENGUE DI KABUPATEN MALANG MENGGUNAKAN EXTREME GRADIENT BOOSTING (XGBOOST)

FORECASTING OF DENGUE FEVER EXTRAORDINARY EVENT (KLB) IN MALANG DISTRICT USING EXTREME GRADIENT BOOSTING (XGBOOST)

HUMAIRA NUR PRADANI
NRP 0521164000011

Dosen Pembimbing :
Edwin Riksakomara, S.Kom, M.T
Retno Aulia Vinarti, S.Kom., M.Kom., Ph.D

DEPARTEMEN SISTEM INFORMASI
Fakultas Teknologi Elektro dan Informatika Cerdas
Institut Teknologi Sepuluh Nopember
Surabaya 2020



ITS
Institut
Teknologi
Sepuluh Nopember

TUGAS AKHIR - IS 184853

PERAMALAN KEJADIAN LUAR BIASA PENYAKIT DEMAM BERDARAH DENGUE DI KABUPATEN MALANG MENGUNAKAN EXTREME GRADIENT BOOSTING (XGBOOST)

HUMAIRA NUR PRADANI
NRP 0521164000011

Dosen Pembimbing :
Edwin Riksakomara, S.Kom, M.T
Retno Aulia Vinarti, S.Kom., M.Kom., Ph.D

DEPARTEMEN SISTEM INFORMASI
Fakultas Teknologi Elektro dan Informatika Cerdas
Institut Teknologi Sepuluh Nopember
Surabaya 2020



ITS

Institut
Teknologi
Sepuluh Nopember

FINAL PROJECT - IS 184853

FORECASTING OF DENGUE FEVER EXTRAORDINARY EVENT (KLB) IN MALANG DISTRICT USING EXTREME GRADIENT BOOSTING (XGBOOST)

HUMAIRA NUR PRADANI
NRP 0521164000011

SUPERVISOR :

Edwin Riksakomara, S.Kom, M.T
Retno Aulia Vinarti, S.Kom., M.Kom., Ph.D

DEPARTMENT OF INFORMATION SYSTEMS
Faculty of Intelligent Electrical and Informatics Technology
Institut Teknologi Sepuluh Nopember
Surabaya 2020

LEMBAR PENGESAHAN**PERAMALAN KEJADIAN LUAR BIASA PENYAKIT
DEMAM BERDARAH DENGUE DI KABUPATEN
MALANG MENGGUNAKAN EXTREME GRADIENT
BOOSTING (XGBOOST)****TUGAS AKHIR**

Disusun Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer (S.Kom)

pada

Departemen Sistem Informasi
Fakultas Teknologi Elektro dan Informatika Cerdas (ELECTICS)
Institut Teknologi Sepuluh Nopember

Oleh

Humaira Nur Pradani

0521164000011

Surabaya, 14 Agustus 2020

Kepala Departemen Sistem Informasi

Dr. Mujahidin, ST., MT.
NIP. 197010102003121001



LEMBAR PERSETUJUAN

PERAMALAN KEJADIAN LUAR BIASA PENYAKIT DEMAM BERDARAH DENGUE DI KABUPATEN MALANG MENGGUNAKAN EXTREME GRADIENT BOOSTING (XGBOOST)

TUGAS AKHIR

Disusun Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer
pada

Departemen Sistem Informasi
Fakultas Teknologi Elektro dan Informatika Cerdas
Institut Teknologi Sepuluh Nopember

Oleh :

HUMAIRA NUR PRADANI
NRP. 052116400011

Disetujui Tim Penguji

Tanggal Ujian : 2 Juli 2020
Periode Wisuda : 122

Edwin Riksakomara, S.Kom., MT.

(Pembimbing I)

Retno Aulia Vinarti, S.Kom., M.Kom., Ph.D (Pembimbing II)

Ahmad Muklason, S.Kom., M.Sc., Ph.D. (Penguji I)

Faizal Mahananto, S.Kom., M.Eng., Ph.D. (Penguji II)

SURAT PERNYATAAN BEBAS PLAGIARISME

Saya yang bertandatangan di bawah ini:

Nama : Humaira Nur Pradani

NRP : 0521164000011

Tempat / Tanggal lahir : Situbondo / 18 April 1998

Fakultas / Departemen : FTEIC / Sistem Informasi

Nomor Telp / Hp / Email : 081336500664 / hnpradani@gmail.com

Dengan ini menyatakan dengan sesungguhnya bahwa penelitian/makalah/tugas akhir saya yang berjudul:

PERAMALAN KEJADIAN LUAR BIASA PENYAKIT DEMAM BERDARAH DENGUE DI KABUPATEN MALANG MENGGUNAKAN EXTREME GRADIENT BOOSTING (XGBOOST)

Bebas Dari Plagiarisme Dan Bukan Hasil Karya Orang Lain.

Apabila dikemudian hari ditemukan seluruh atau sebagian penelitian/makalah/tugas akhir tersebut terdapat indikasi plagiarisme, maka saya bersedia menerima sanksi sesuai peraturan dan ketentuan yang berlaku. Demikian surat pernyataan ini saya buat dengan sesungguhnya dan untuk dipergunakan sebagaimana mestinya.

Surabaya, 7 Agustus 2020



Humaira Nur Pradani
NRP.0521164000011

Halaman ini sengaja dikosongkan

**PERAMALAN KEJADIAN LUAR BIASA PENYAKIT
DEMAM BERDARAH DENGUE DI KABUPATEN
MALANG MENGGUNAKAN EXTREME GRADIENT
BOOSTING (XGBOOST)**

Nama Mahasiswa : Humaira Nur Pradani
NRP : 0521164000011
Jurusan : Sistem Informasi FTEIC-ITS
Pembimbing 1 : Edwin Riksakomara, S.Kom, M.T
Pembimbing 2 : Retno Aulia Vinarti, S.Kom., M.Kom.,
Ph.D

ABSTRAK

Demam Berdarah Dengue (DBD) merupakan salah satu masalah kesehatan utama di Indonesia karena jumlah kasusnya yang cenderung semakin meningkat dari tahun ke tahun. Penyakit ini juga pernah menjadi Kejadian Luar Biasa hampir di seluruh wilayah Indonesia pada berbagai kurun waktu, termasuk di kabupaten Malang. Peramalan Kejadian Luar Biasa Demam Berdarah sangat penting dilakukan agar Dinas Kesehatan Kabupaten Malang dapat membuat perencanaan yang lebih baik untuk pencegahan terjadinya Kejadian Luar Biasa demam berdarah di Kabupaten Malang. Kabupaten Malang memiliki tiga pembagian daerah berdasarkan letak geografis, yaitu dataran rendah, dataran sedang, dan dataran tinggi, dimana tiap dataran terdapat variabel-variabel yang dapat mempengaruhi peningkatan angka kasus demam berdarah.

Penelitian ini bertujuan untuk menemukan model terbaik serta mengetahui tingkat akurasi dari model tersebut untuk meramalkan Kejadian Luar Biasa (KLB) demam berdarah. Penelitian ini menggunakan metode Extreme Gradient Boosting untuk meramalkan jumlah kasus demam berdarah serta Klasifikasi Berbasis Aturan untuk mengklasifikasi KLB demam berdarah di

Kabupaten Malang. Extreme Gradient Boosting atau XGBoost merupakan salah suatu metode ensemble tree yang menerapkan konsep gradient boosting yang telah dioptimalkan. Sedangkan Klasifikasi Berbasis Aturan merupakan metode klasifikasi berdasarkan aturan-aturan tertentu dalam penentuan kelas.

Dari hasil penelitian yang dilakukan dapat diketahui bahwa skenario pembentukan model mulai dari pembagian jumlah data pelatihan, variabel independen yang digunakan, jumlah fitur lag, dan hyper-parameter dapat mempengaruhi hasil akurasi peramalan. Hasil akurasi terbaik yang mampu dihasilkan XGBoost dalam melakukan peramalan data time series angka kejadian demam berdarah yakni dengan nilai RMSE sebesar 2,532 pada dataran rendah, RMSE sebesar 2,899 pada dataran sedang, dan RMSE sebesar 1,446 pada dataran tinggi. Sedangkan Klasifikasi Berbasis Aturan mampu menentukan daerah KLB atau non KLB dengan tingkat akurasi terbaik sebesar 93,75% di dataran sedang, hasil ini disertai dengan nilai sensitivitas sebesar 100,00% dan spesifisitas sebesar 90,48%. Selain itu, pada dataran rendah didapatkan nilai akurasi, sensitivitas, dan spesifisitas secara berturut-turut sebesar 83,33%, 50,00%, dan 88,46%. Sedangkan pada dataran tinggi ketiga nilai faktor tersebut adalah 86,21%, 66,67%, dan 100,00%.

Kata Kunci : Peramalan, Kejadian Luar Biasa, Demam Berdarah Dengue, Extreme Gradient Boosting, Klasifikasi Berbasis Aturan

**PERAMALAN KEJADIAN LUAR BIASA PENYAKIT
DEMAM BERDARAH DENGUE DI KABUPATEN
MALANG MENGGUNAKAN EXTREME GRADIENT
BOOSTING (XGBOOST)**

Nama Mahasiswa : Humaira Nur Pradani
NRP : 0521164000011
Jurusan : Sistem Informasi FTEIC-ITS
Pembimbing 1 : Edwin Riksakomara, S.Kom, M.T
Pembimbing 2 : Retno Aulia Vinarti, S.Kom., M.Kom.,
Ph.D

ABSTRACT

Dengue Hemorrhagic Fever (DHF) is one of the main health problems in Indonesia because the number of cases tends to increase from year to year. This disease has also been an Extraordinary Event (KLB) in almost all regions of Indonesia at various times, including in Malang regency. Forecasting Extraordinary Occurrence of Dengue Fever is very important so that Malang District Health Office can make better plans for preventing the occurrence of Dengue Fever in Malang Regency. Malang Regency has three regional divisions based on geographical location, namely lowlands, medium plains, and highlands, where each plateau has variables that can affect the increase in the number of dengue cases.

This study aims to find the best model and determine the level of accuracy of the model to predict Dengue Fever Extraordinary Event (KLB). This study uses the Extreme Gradient Boosting method to predict the number of dengue cases as well Rule-Based Classification to classify KLB of dengue fever in Malang Regency. XGBoost is an ensemble tree method that applies the optimized gradient boosting concept, whereas Rule Based Classification is a

classification method based on certain rules in class determination.

From the results of this study it can be seen that the scenario of the model formation starting from the division of the amount of training data, the independent variables used, the number of lag features, and hyper-parameters can influence the results of forecasting accuracy. The best accuracy value that can be generated by XGBoost in forecasting time series data is RMSE of 2.532 in lowlands, RMSE of 2.899 in medium plains, and RMSE of 1.446 in the highlands. Whereas Rule Based Classification is able to determine the KLB or non-KLB area with the best accuracy level of 93.75% in the medium plains, this result is accompanied by a sensitivity value of 100.00% and specificity of 90.48%. In addition, in the lowlands values obtained accuracy, sensitivity, and specificity were 83.33%, 50.00%, and 88.46%, respectively. Whereas in the third plateau the value of these factors was 86.21%, 66.67%, and 100.00%.

Keywords: Forecasting, Extraordinary Event, Dengue Hemorrhagic Fever, Extreme Gradient Boosting, Rule-Based Classification

KATA PENGANTAR

Puji dan syukur penulis tuturkan ke hadirat Allah SWT, Tuhan Semesta Alam yang telah memberikan karunia dan hidayah-Nya kepada penulis sehingga penulis mendapatkan kelancaran dalam menyelesaikan tugas akhir dengan judul:

PERAMALAN KEJADIAN LUAR BIASA PENYAKIT DEMAM BERDARAH DENGUE DI KABUPATEN MALANG MENGGUNAKAN EXTREME GRADIENT BOOSTING (XGBOOST)

Terima kasih penulis sampaikan kepada pihak-pihak yang telah mendukung, memberikan saran, motivasi, semangat, dan bantuan baik berupa material maupun moril demi tercapainya tujuan pembuatan tugas akhir ini. Tugas akhir ini tidak akan pernah terwujud tanpa bantuan dan dukungan dari berbagai pihak yang sudah meluangkan waktu, tenaga dan pikirannya. Secara khusus penulis akan menyampaikan ucapan terima kasih yang sebanyak-banyaknya kepada :

1. Bapak Andi Prahjono dan Ibu Dewi Ramdani selaku kedua orang tua serta Annisa Pradani dan Rahmaidha Arofa Pradani selaku saudara kandung dari penulis yang tiada henti memberikan dukungan dan semangat.
2. Bapak Edwin Riksakomara, S.Kom, M.T dan Retno Aulia Vinarti, S.Kom., M.Kom., Ph.D selaku dosen pembimbing dan ibu Wiwik Anggraeni, S.Si., M.Kom, yang senantiasa meluangkan waktu, memberikan ilmu dan petunjuk, serta memotivasi untuk kelancaran tugas akhir.
3. Bapak Ahmad Muklason, S.Kom., M.Sc., Ph.D dan Faizal Mahananto, S.Kom, M.Eng, selaku dosen penguji yang telah memberikan saran dan kritik yang sangat membangun untuk perbaikan tugas akhir.
4. Seluruh dosen Jurusan Sistem Informasi ITS yang telah memberikan ilmu yang bermanfaat kepada penulis.

5. Firin, Berta, dan Ayun yang selalu saling mendukung untuk menyelesaikan tugas akhir.
6. Aldy dan Rully yang telah membantu saya memahami materi dan kode.
7. Teman-teman seperjuangan dalam lab RDIB yang telah banyak memberikan masukan dalam kehidupan dan tugas akhir.
8. Rekan-rekan ARTEMIS yang telah memberikan banyak kenangan manis dan pahit semasa kuliah.
9. Berbagai pihak yang tidak bisa disebutkan satu persatu yang telah turut serta menyukseskan penulis dalam menyelesaikan tugas akhir.

Penyusunan laporan ini masih jauh dari kata sempurna sehingga penulis menerima adanya kritik maupun saran yang membangun untuk perbaikan di masa yang akan datang. Semoga buku tugas akhir ini dapat memberikan manfaat bagi pembaca.

Surabaya, 23 Mei 2020

Penulis

DAFTAR ISI

ABSTRAK	vii
ABSTRACT	ix
SURAT PERNYATAAN BEBAS PLAGIARISME	v
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xiii
DAFTAR GAMBAR	xix
DAFTAR TABEL	xxiii
DAFTAR KODE	xxvii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Batasan Permasalahan	4
1.4. Tujuan Penelitian.....	5
1.5. Manfaat Penelitian.....	5
1.6. Relevansi	5
BAB II TINJAUAN PUSTAKA	7
2.1. Penelitian Sebelumnya	7
2.2. Landasan Teori	11
2.2.1. Kejadian Luar Biasa	11
2.2.2. Demam Berdarah Dengue	12
2.2.3. Peramalan Time Series	12
2.2.4. Normalisasi Data	13
2.2.5. Uji Korelasi	14
2.2.6. Angka Bebas Jentik (ABJ)	14
2.2.7. Windowing	15
2.2.8. Extreme Gradient Boosting	15
2.2.9. Klasifikasi Berbasis Aturan.....	19
2.2.10. Pengukuran Performa Peramalan	20
a. RMSE	20
b. SMAPE.....	21
2.2.11. Pengukuran Performa Klasifikasi.....	21
a. Akurasi	22
b. Spesifisitas.....	22
c. Sensitivitas.....	22

BAB III METODOLOGI PENELITIAN	25
3.1. Metodologi	25
3.2. Uraian Metodologi.....	25
3.2.1. Studi Literatur.....	25
3.2.2. Persiapan Data	25
a. Pengumpulan Data.....	27
b. Pra-Proses Data.....	27
3.2.3. Tahap Peramalan	28
a. Pemodelan	28
b. Mengoptimalkan Parameter.....	29
c. Pengujian Model Peramalan	29
3.2.4. Tahap Klasifikasi.....	29
a. Penambahan Fitur Penunjang Klasifikasi	29
b. Penentuan Status KLB.....	29
c. Pengukuran Kinerja Klasifikasi	30
3.2.5. Peramalan Periode Mendatang	30
3.2.6. Analisa Hasil dan Penarikan Kesimpulan.....	30
3.2.7. Penyusunan Buku Tugas Akhir	30
BAB IV PERANCANGAN	31
4.1. Persiapan data.....	31
4.1.1. Pengumpulan Data.....	31
4.1.2. <i>Pra-Processing</i> Data.....	32
a. Mengubah Data Harian Variabel Iklim	32
b. Penanganan <i>Missing Value</i>	32
c. Pemilihan Kecamatan Pembentuk Model.....	33
d. Normalisasi Data	33
e. Pemilihan Fitur dengan Uji Korelasi	34
f. Windowing	35
g. Membagi Data Pelatihan dan Pengujian.....	36
4.2. Tahap Peramalan	36
4.2.1. Penentuan Parameter pada Model XGBoost	37
4.2.2. Pembentukan Model Peramalan XGBoost	39
4.2.3. Menguji Model Peramalan XGBoost	40
4.3. Tahap Klasifikasi.....	41
4.3.1. Penambahan Fitur Penunjang Klasifikasi.....	41

4.3.2.	Pelabelan KLB pada Data Hasil Peramalan	42
4.3.3.	Pengukuran Performa Klasifikasi	42
4.4.	Peramalan Data 24 Periode Selanjutnya.....	42
4.5.	Perbandingan dengan Metode/Algoritma Lain	46
BAB V	IMPLEMENTASI	47
5.1.	Persiapan Implementasi.....	47
5.2.	Persiapan Data	47
5.2.1.	Mengubah Data Harian Variabel Iklim	48
5.2.2.	Penanganan Missing Value	48
5.2.3.	Pemilihan Kecamatan Pembentuk Model	50
5.2.4.	Normalisasi Data	51
5.2.5.	Pemilihan Fitur dengan Uji Korelasi.....	52
5.2.6.	Windowing	52
5.2.7.	Pembagian Data Pelatihan dan Data Pengujian...	53
5.3.	Tahap Peramalan	54
5.3.7.	Penentuan Parameter pada Model XGBoost	54
5.3.8.	Pembentukan Model XGBoost.....	55
5.3.9.	Pengujian Model Peramalan.....	57
a.	Pengujian pada <i>test-set</i>	57
b.	Pengujian pada Proporsi Data Lain	58
c.	Pengujian pada Kecamatan Lain	60
5.4.	Tahap Klasifikasi.....	60
5.4.1.	Penambahan Fitur Penunjang Klasifikasi.....	61
5.4.2.	Penentuan Status KLB untuk Data Hasil Peramalan 62	
5.4.3.	Pengukuran Performa Klasifikasi	63
5.5.	Peramalan Data 24 Periode Selanjutnya.....	64
5.5.1.	Peramalan Variabel Independen 24 Periode Selanjutnya.....	64
5.5.2.	Peramalan Kejadian Demam Berdarah untuk 24 Periode Selanjutnya.....	65
5.5.3.	Pelabelan KLB untuk 24 Periode Selanjutnya	67
BAB VI	HASIL DAN PEMBAHASAN.....	69
6.1.	Hasil Pra-Proses Data.....	69
6.1.1.	Hasil Pembentukan Data Iklim.....	69

6.1.2.	Hasil Penanganan <i>Missing Value</i>	69
6.1.3.	Hasil Pemilihan Kecamatan Pembentuk Model ..	71
6.1.4.	Hasil Normalisasi	73
6.1.5.	Hasil Pemilihan Fitur dengan Uji Korelasi.....	73
a.	Dataran Rendah	73
b.	Dataran Sedang.....	75
c.	Dataran Tinggi.....	76
6.1.6.	Hasil Windowing.....	78
6.2.	Hasil Peramalan dan Klasifikasi Dataran Rendah	79
6.2.1.	Hasil Peramalan pada Dataran Rendah.....	79
a.	Hasil Uji Coba Skenario Peramalan	79
b.	Hasil Peramalan pada Kecamatan Selain Pembentuk Model.....	86
6.2.2.	Hasil Klasifikasi pada Dataran Rendah	90
a.	Hasil Uji Coba Skenario Klasifikasi.....	90
b.	Hasil Klasifikasi pada Kecamatan Selain Pembentuk Model.....	93
6.2.3.	Hasil Perhitungan 24 Periode Kedepan pada Dataran Rendah	96
6.3.	Hasil Peramalan dan Klasifikasi Dataran Sedang	98
6.3.1.	Hasil Peramalan pada Dataran Sedang	99
a.	Hasil Uji Coba Skenario Peramalan	99
b.	Hasil Peramalan pada Kecamatan Selain Pembentuk Model.....	105
6.3.2.	Hasil Klasifikasi pada Dataran Sedang.....	110
a.	Hasil Uji Coba Skenario Klasifikasi.....	110
b.	Hasil Klasifikasi pada Kecamatan Selain Pembentuk Model.....	112
6.3.3.	Hasil Perhitungan 24 Periode Kedepan pada Dataran Sedang.....	118
6.4.	Hasil Peramalan dan Klasifikasi Dataran Tinggi.....	120
6.4.1.	Hasil Peramalan pada Dataran Tinggi	120
a.	Hasil Uji Coba Skenario Peramalan	120
b.	Hasil Peramalan pada Kecamatan Selain Pembentuk Model.....	126

6.4.2.	Hasil Klasifikasi pada Dataran Tinggi	129
a.	Hasil Uji Coba Skenario Klasifikasi.....	129
b.	Hasil Klasifikasi pada Kecamatan Selain Pembentuk Model	132
6.4.3.	Hasil Perhitungan 24 Periode Kedepan pada Dataran Tinggi.....	133
6.5.	Hasil Pencarian Satu Model Terbaik untuk Seluruh Kecamatan	135
6.6.	Hasil Perbandingan dengan Metode Lain.....	137
6.6.1.	Perbandingan Performa Peramalan	137
6.6.2.	Perbandingan Performa Klasifikasi	138
6.7.	Kesimpulan Hasil Percobaan.....	139
6.8.	Analisis Hasil Percobaan.....	143
6.8.1.	Analisis Segi Manajerial.....	143
6.8.2.	Analisis Segi Metode.....	146
BAB VII	KESIMPULAN DAN SARAN	151
7.1.	Kesimpulan.....	151
7.2.	Saran.....	153
DAFTAR PUSTAKA	155
BIODATA PENULIS	159
LAMPIRAN A	161
LAMPIRAN B	167
LAMPIRAN C	181
LAMPIRAN D	185
LAMPIRAN E	188

Halaman ini sengaja dikosongkan

DAFTAR GAMBAR

Gambar 1.1 Relevansi Tugas Akhir dengan Roadmap Lab RDIB 6	
Gambar 2.1 Proses Windowing	15
Gambar 2.2 Ilustrasi Gradient Boosting	16
Gambar 3.1 Alur Kerja Penelitian	26
Gambar 4.1 Pola Data Suhu Udara	43
Gambar 4.2 Pola Data Kelembapan Udara.....	44
Gambar 4.3 Pola Data Curah Hujan.....	44
Gambar 4.4 Pola Data Kecepatan Angin.....	45
Gambar 4.5 Pola Data Jumlah Penduduk Kec. Kepanjen	45
Gambar 4.6 Pola Data Angka Bebas Jentik.....	46
Gambar 5.1 Menu Pada Minitab	65
Gambar 5.2 Menu Pada Excel.....	65
Gambar 6.1 Penanganan <i>Missing Value</i> pada Variabel Suhu Udara Kec. Kepanjen	71
Gambar 6.2 Penanganan <i>Missing Value</i> pada Variabel ABJ Kec. Wajak	71
Gambar 6.3 Persebaran RMSE berdasarkan Skenario Pembagian Data (Dataran Rendah).....	83
Gambar 6.4 Persebaran RMSE berdasarkan Skenario Pemilihan Fitur (Dataran Rendah).....	83
Gambar 6.5 Perbandingan Data Aktual dan Hasil Peramalan Kec. Kepanjen.....	85
Gambar 6.6 Perbandingan Data Aktual dan Hasil Peramalan Kec. Turen	88
Gambar 6.7 Perbandingan Data Aktual dan Hasil Peramalan Kec. Gondanglegi	89
Gambar 6.8 Perbandingan Data Aktual dan Hasil Peramalan Kec. Donomulyo.....	89
Gambar 6.9 Perbandingan Data Aktual dan Hasil Peramalan Kec. Bululawang.....	90
Gambar 6.10 <i>Confusion Matrix</i> Kec. Kepanjen	93
Gambar 6.11 <i>Confusion Matrix</i> Kec. Turen.....	94
Gambar 6.12 <i>Confusion Matrix</i> Kec. Gondanglegi.....	94
Gambar 6.13 <i>Confusion Matrix</i> Kec. Donomulyo	95

Gambar 6.14 <i>Confusion Matrix</i> Kec. Bululawang	96
Gambar 6.15 Perbandingan Data Aktual dan Hasil Peramalan di Dataran Rendah (Kec. Kepanjen).....	98
Gambar 6.16 Persebaran RMSE berdasarkan Skenario Pembagian Data (Dataran Sedang)	102
Gambar 6.17 Persebaran RMSE berdasarkan Skenario Pemilihan Fitur (Dataran Sedang)	103
Gambar 6.18 Perbandingan Data Aktual dan Hasil Peramalan Kec. Pakisaji	103
Gambar 6.19 Perbandingan Data Aktual dan Hasil Peramalan Kec. Dampit	106
Gambar 6.20 Perbandingan Data Aktual dan Hasil Peramalan Kec. Singosari.....	107
Gambar 6.21 Perbandingan Data Aktual dan Hasil Peramalan Kec. Tumpang.....	107
Gambar 6.22 Perbandingan Data Aktual dan Hasil Peramalan Kec. Lawang	108
Gambar 6.23 Perbandingan Data Aktual dan Hasil Peramalan Kec. Sumbermanjing.....	108
Gambar 6.24 Perbandingan Data Aktual dan Hasil Peramalan Kec. Wajak.....	109
Gambar 6.25 Perbandingan Data Aktual dan Hasil Peramalan Kec. Karangploso.....	109
Gambar 6.26 <i>Confusion Matrix</i> Kec. Pakisaji.....	112
Gambar 6.27 <i>Confusion Matrix</i> Kec. Dampit.....	113
Gambar 6.28 <i>Confusion Matrix</i> Kec. Singosari	114
Gambar 6.29 <i>Confusion Matrix</i> Kec. Tumpang	115
Gambar 6.30 <i>Confusion Matrix</i> Kec. Lawang.....	115
Gambar 6.31 <i>Confusion Matrix</i> Kec. Wajak	116
Gambar 6.32 <i>Confusion Matrix</i> Kec. Karangploso	117
Gambar 6.33 <i>Confusion Matrix</i> Kec. Sumbermanjing	117
Gambar 6.34 Perbandingan Data Aktual dan Hasil Peramalan di Dataran Sedang (Kec. Pakisaji).....	119
Gambar 6.35 Persebaran RMSE berdasarkan Skenario Pembagian Data (Dataran Tinggi).....	124

Gambar 6.36 Persebaran RMSE berdasarkan Skenario Pemilihan Fitur (Dataran Tinggi)	125
Gambar 6.37 Perbandingan Data Aktual dan Hasil Peramalan Kec. Ngajum	125
Gambar 6.38 Perbandingan Data Aktual dan Hasil Peramalan Kec. Poncokusumo	128
Gambar 6.39 Perbandingan Data Aktual dan Hasil Peramalan Kec. Jabung.....	129
Gambar 6.40 Confusion Matrix Kec. Ngajum	131
Gambar 6.41 <i>Confusion Matrix</i> Kec. Poncokusumo.....	132
Gambar 6.42 <i>Confusion Matrix</i> Kec. Jabung	133
Gambar 6.43 Perbandingan Data Aktual dan Hasil Peramalan di Dataran Tinggi (Kec. Ngajum).....	135
Gambar 6.44 Hasil Performa Peramalan (a) RMSE dan (b) SMAPE tiap Dataran	140
Gambar 6.45 Hasil Performa Klasifikasi tiap Dataran	142
Gambar 6.46 Persebaran Kejadian Demam Berdarah Berdasarkan Kecamatan (Tahun 2019-2020).....	145
Gambar 6.47 Tren Angka Kejadian Demam Berdarah	145
Gambar 6.48 Hasil Sensitivity Analysis pada Parameter XGBoost	147
Gambar 6.49 Dekomposisi Tren & Siklus pada Dataran (a) Rendah (b) Sedang (c) Tinggi	148
Gambar A.1 Peramalan Suhu Udara dengan Metode Dekomposisi Multiplikatif (Stasiun Karangates)	161
Gambar A.2 Peramalan Suhu Udara dengan Metode Dekomposisi Multiplikatif (Stasiun Karangploso).....	161
Gambar A.3 Peramalan Kelembapan dengan Metode Dekomposisi Multiplikatif (Stasiun Karangates)	162
Gambar A.4 Peramalan Kelembapan dengan Metode Dekomposisi Multiplikatif (Stasiun Karangploso).....	162
Gambar A.5 Peramalan Curah Hujan dengan Metode ARIMA (Stasiun Karangates).....	163
Gambar A.6 Peramalan Curah Hujan dengan Metode ARIMA (Stasiun Karangploso)	163

Gambar A.7 Peramalan Kecepatan Angin dengan Metode ARIMA (Stasiun Karangates).....	164
Gambar A.8 Peramalan Kecepatan Angin Metode ARIMA (Stasiun Karangploso)	164
Gambar A.9 Peramalan Jumlah Penduduk dengan Metode Double Exponential Smoothing	165
Gambar A.10 Peramalan Angka Bebas Jentik dengan Metode Exponential Smoothing	165

DAFTAR TABEL

Tabel 2.1 Studi Sebelumnya.....	7
Tabel 2.2 Confusion Matrix	22
Tabel 4.1 Skenario Pemilihan Fitur.....	34
Tabel 4.2 Skenario Windowing.....	35
Tabel 4.3 Skenario Pembagian Data	36
Tabel 4.4 Parameter pada Arsitektur XGBoost.....	37
Tabel 4.5 Skenario Pengujian Model	40
Tabel 5.1 Spesifikasi Perangkat Keras dan Lunak	47
Tabel 5.2 Lingkungan Uji Coba	48
Tabel 5.3 Penjelasan dari Kode 5.1	49
Tabel 5.4 Penjelasan dari Kode 5.2.....	50
Tabel 5.5 Penjelasan dari Kode 5.3.....	50
Tabel 5.6 Penjelasan dari Kode 5.4.....	51
Tabel 5.7 Penjelasan dari Kode 5.5.....	53
Tabel 5.8 Penjelasan dari Kode 5.7.....	54
Tabel 5.9 Penjelasan dari Kode 5.9.....	56
Tabel 5.10 Penjelasan dari Kode 5.10.....	58
Tabel 5.11 Penjelasan dari Kode 5.11	60
Tabel 5.12 Penjelasan dari Kode 5.13.....	63
Tabel 5.13 Penjelasan dari Kode 5.14.....	64
Tabel 5.14 Penjelasan dari Kode 5.15.....	67
Tabel 6.1 Hasil Pra Proses Variabel Iklim bulan Januari 2010 ...	69
Tabel 6.2 Data Harian Iklim Bulan Januari 2010.....	70
Tabel 6.3 Hasil Korelasi Antarkecamatan di Dataran Rendah....	72
Tabel 6.4 Perbandingan Variabel Suhu Udara Sebelum dan Setelah Normalisasi pada Kec. Kepanjen	73
Tabel 6.5 Uji Korelasi Variabel Independen (Dataran Rendah) .	74
Tabel 6.6 Variabel Independen Tiap Skenario (Dataran Rendah)	74
Tabel 6.7 Uji Korelasi Variabel Independen (Dataran Sedang) ..	75
Tabel 6.8 Variabel Independen Tiap Skenario (Dataran Sedang)	76
Tabel 6.9 Uji Korelasi Variabel Independen (Dataran Tinggi) ...	76
Tabel 6.10 Variabel Independen Tiap Skenario (Dataran Tinggi)	77

Tabel 6.11 Hasil Windowing Pada Kec. Pakisaji degan 3 Lag ...	79
Tabel 6.12 Hasil RMSE untuk <i>test-set</i> di Kec. Kepanjen.....	80
Tabel 6.13 Hasil <i>Tuning</i> Parameter di Dataran Rendah	81
Tabel 6.14 Hasil Uji Proporsional Data pada Kec. Kepanjen	85
Tabel 6.15 Perbandingan RMSE Uji Kecamatan Lain pada Dataran Rendah.....	86
Tabel 6.16 Nilai SMAPE Kecamatan Lain (Dataran Rendah)	87
Tabel 6.17 Evaluasi Klasifikasi pada <i>Test-set</i> (Dataran Rendah)	91
Tabel 6.18 Evalausi Klasifikasi pada Uji Proporsi dan Uji Kecamatan Lain (Dataran Rendah)	92
Tabel 6.19 Hasil Peramalan 2 Tahun Mendatang pada Kec. Kepanjen.....	97
Tabel 6.20 Hasil RMSE untuk <i>test-set</i> di Kec. Pakisaji	100
Tabel 6.21 Hasil <i>Tuning</i> Parameter di Dataran Sedang.....	101
Tabel 6.22 Hasil Uji Proporsional Data pada Kec. Pakisaji	104
Tabel 6.23 Perbandingan Hasil RMSE pada Dataran Sedang ...	105
Tabel 6.24 Hasil Nilai SMAPE Kecamatan Lain (Dataran Sedang)	106
Tabel 6.25 Evaluasi Klasifikasi pada <i>Test-set</i> (Dataran Sedang)	110
Tabel 6.26 Evalausi Klasifikasi pada Uji Proporsi dan Uji Kecamatan Lain (Dataran Sedang).....	111
Tabel 6.27 Hasil Peramalan 2 Tahun Mendatang pada Kec. Pakisaji	118
Tabel 6.28 Hasil RMSE untuk <i>test-set</i> di Kec. Ngajum.....	121
Tabel 6.29 Hasil <i>Tuning</i> Parameter di Dataran Tinggi.....	122
Tabel 6.30 Hasil Uji Proporsional Data pada Kec. Ngajum.....	126
Tabel 6.31 Perbandingan Hasil RMSE pada Dataran Tinggi	127
Tabel 6.32 Hasil Nilai SMAPE Kecamatan Lain (Dataran Tinggi)	128
Tabel 6.33 Evaluasi Klasifikasi pada <i>Test-set</i> (Dataran Tinggi)	130
Tabel 6.34 Evalausi Klasifikasi pada Uji Proporsi dan Uji Kecamatan Lain (Dataran Rendah)	130

Tabel 6.35 Hasil Peramalan 2 Tahun Mendatang pada Kec. Ngajum.....	134
Tabel 6.36 Hasil Pencarian Satu Model Terbaik.....	136
Tabel 6.37 Perbandingan RMSE Terbaik dengan Algoritma Lain	137
Tabel 6.38 Perbandingan Performa Klasifikasi dengan Metode Lain.....	138
Tabel 6.39 Parameter pada Model terbaik Tiap Dataran.....	139
Tabel 6.40 Model Final pada Tiap Dataran.....	140
Tabel 6.41 Tingkat Robustness Model Final.....	141
Tabel 6.42 Angka Kasus Demam Berdarah Per Kecamatan pada Tahun 2019 - 2020	144
Tabel B.1 Hasil Peramalan menggunakan XGBoost pada Dataran Rendah.....	167
Tabel B.2 Hasil Peramalan menggunakan XGBoost pada Dataran Sedang	171
Tabel B.3 Hasil Peramalan menggunakan XGBoost pada Dataran Tinggi	177
Tabel C.1 Evaluasi Klasifikasi pada Uji Proporsi Data di Dataran Rendah.....	181
Tabel C.2 Evaluasi Klasifikasi pada Uji Kecamatan Lain di Dataran Rendah	181
Tabel C.3 Evaluasi Klasifikasi pada Uji Proporsi Data di Dataran Sedang	182
Tabel C.4 Evaluasi Klasifikasi pada Uji Kecamatan Lain di Dataran Sedang	183
Tabel C.5 Evaluasi Klasifikasi pada Uji Proporsi Data di Dataran Tinggi	184
Tabel C.6 Evaluasi Klasifikasi pada Uji Kecamatan Lain di Dataran Tinggi.....	184
Tabel D.1 Uji Korelasi Antarkecamatan pada Dataran Rendah	185
Tabel D.2 Uji Korelasi pada Dataran Sedang	186
Tabel D.3 Uji Korelasi pada Dataran Tinggi.....	187
Tabel E.1 Hasil Uji pada <i>Test-set</i> (Dataran Rendah)	189
Tabel E.2 Hasil Uji Proporsional Data (Dataran Rendah).....	191

Tabel E.3 Hasil Uji Kecamatan Lain (Dataran Rendah)	192
Tabel E.4 Hasil Uji pada <i>Test-set</i> (Dataran Sedang)	194
Tabel E.5 Hasil Uji Proporsional Data (Dataran Rendah).....	196
Tabel E.6 Hasil Uji Kecamatan Lain (Dataran Rendah)	197
Tabel E.7 Hasil Uji pada <i>Test-set</i> (Dataran Tinggi)	199
Tabel E.8 Hasil Uji Proporsional Data (Dataran Tinggi)	201
Tabel E.9 Hasil Uji Kecamatan Lain (Dataran Tinggi)	202

DAFTAR KODE

Kode 5.1 Interpolasi pada Variabel Angka Bebas Jentik	49
Kode 5.2 Ekstrapolasi dan Regresi pada Variabel Angka Bebas Jentik (ABJ).....	50
Kode 5.3 Perhitungan Korelasi Antarkecamatan	50
Kode 5.4 Normalisasi Min Max	51
Kode 5.5 Uji Korelasi Variabel Independen	52
Kode 5.6 Proses Windowing	52
Kode 5.7 Pembagian Data Pelatihan dan Datas Pengujian	53
Kode 5.8. Penentuan Parameter	55
Kode 5.9 Pembentukan Model XGBoost	56
Kode 5.10 Pengujian pada <i>test-set</i>	58
Kode 5.11 Uji Model Peramalan pada Proporsi Data Lain	59
Kode 5.12 Uji Model Peramalan pada Kecamatan Lain	61
Kode 5.13 Penentuan Status KLB dan non-KLB	62
Kode 5.14 Evaluasi Klasifikasi	64
Kode 5.15 Peramalan 24 Periode Kedepan	66

Halaman ini sengaja dikosongkan

BAB I

PENDAHULUAN

Pada bab pendahuluan ini diuraikan proses identifikasi masalah penelitian yang meliputi latar belakang masalah, perumusan masalah, batasan masalah, tujuan tugas akhir, manfaat kegiatan tugas akhir, manfaat kegiatan tugas akhir dan relevansi terhadap pengerjaan tugas akhir. Berdasarkan uraian pada bab ini, harapannya gambaran umum permasalahan dan pemecahan masalah pada tugas akhir dapat dipahami.

1.1. Latar Belakang

Penyakit Demam Berdarah Dengue (DBD) merupakan penyakit musiman dan menular yang disebabkan oleh virus Dengue. Penyakit ini dapat ditularkan melalui nyamuk spesies *Aedes aegypti* dan *Aedes albopictus* sebagai vektor (penular) primer dan nyamuk *Aedes polynesiensis*, *Aedes scutellaris* serta *Ae (Finlaya) niveus* sebagai vektor (penular) sekunder. Penyakit ini sangat rentan terjadi di Indonesia karena iklim tropis sangat cocok untuk menjadi tempat berkembangnya organisme penyebar penyakit tersebut [1].

Hingga saat ini, penyakit Demam Berdarah Dengue (DBD) masih menjadi salah satu masalah kesehatan utama di Indonesia karena jumlah kasusnya yang cenderung semakin meningkat dari tahun ke tahun. Secara nasional, jumlah kasus penyakit DBD terhitung pada bulan Februari 2019 adalah 16.692 kasus sedangkan sebelumnya pada bulan Januari 2019 sebanyak 13.683 kasus [2]. Di samping itu, DBD pernah menjadi Kejadian Luar Biasa hampir di seluruh wilayah Indonesia pada berbagai kurun waktu [3]. KLB demam berdarah merupakan timbulnya suatu penyakit menular tertentu sebagaimana dimaksud dalam Pasal 4 Permenkes No. 1501/2010, yang sebelumnya tidak ada atau tidak dikenal pada suatu daerah dan atau jumlah penderita baru dalam periode waktu satu bulan menunjukkan kenaikan dua kali atau lebih dibandingkan dengan angka rata-rata per bulan dalam tahun sebelumnya dan atau angka kematian kasus suatu

penyakit dalam satu kurun waktu tertentu menunjukkan kenaikan 50% atau lebih dibandingkan dengan angka kematian kasus suatu penyakit periode sebelumnya dalam kurun waktu yang sama [4].

Di Indonesia, KLB demam berdarah yang dulunya terjadi setiap 5 tahun di Indonesia, kini makin sering terjadi dan menghasilkan angka kematian yang semakin meningkat [5]. Jawa Timur merupakan provinsi yang menyumbang angka kasus demam berdarah tertinggi di Indonesia pada tahun 2019 [2]. Sedangkan dari 1.081 penderita demam berdarah dengue (DBD) di Jawa Timur, 60% penderita di antaranya adalah warga Kabupaten Malang. Oleh sebab itu, pemerintah Jawa Timur khususnya Kabupaten Malang masih gencar dalam melakukan upaya untuk menekan jumlah penderita demam berdarah. Beberapa upaya pemerintah yang dilakukan untuk pencegahan penyakit ini adalah melakukan pengasapan (*fogging*), pemberian *abate* (obat pembunuh jentik nyamuk) kepada masyarakat, serta program Jumat Bersih yang rutin digelar tiap minggunya [6]. Sebagai pendukung upaya pemerintah tersebut, peramalan KLB penyakit demam berdarah dapat dilakukan untuk menghasilkan upaya persiapan yang lebih baik dan dilakukan pada waktu yang tepat sehingga mengurangi angka kejadian dan menyelamatkan lebih banyak jiwa.

Peramalan *outbreak* penyakit demam berdarah pernah dilakukan sebelumnya di Malaysia dan China. Di Malaysia prediksi *outbreak* penyakit demam berdarah pernah dilakukan menggunakan pendekatan Least Squares Support Vector Machines dan menghasilkan nilai akurasi sebesar 86.84% [7]. Pada tahun 2018, dilakukan penelitian lain mengenai prediksi *outbreak* demam berdarah dengan meramalkan jumlah kasus demam berdarah di China menggunakan beberapa metode *machine learning* dan melibatkan variabel kondisi iklim seperti suhu, curah hujan, kelembapan serta kecepatan angin. Penelitian tersebut menyimpulkan bahwa Extreme Gradient Boosting (XGBoost) memiliki nilai error paling kecil dibandingkan

algoritma pembandingan lainnya yaitu Random Forest, Lasso, ADABoost, SVM, Linear Regression, GBoost dan Rideg [8]. Selain meramalkan kasus penyakit demam berdarah, metode Extreme Gradient Boosting (XGBoost) juga pernah digunakan untuk meramalkan jumlah kasus penyakit lain seperti kolera dan *foodborne disease* serta menghasilkan nilai akurasi yang cukup baik [9] [10]. Maka dari itu, penelitian ini mengusulkan penggunaan metode Extreme Gradient Boosting untuk meramalkan jumlah kasus demam berdarah pada Kabupaten Malang dengan mempertimbangkan bahwa metode tersebut menghasilkan performa yang baik untuk digunakan dalam meramalkan data angka kejadian suatu penyakit, salah satunya adalah penyakit demam berdarah. Selain itu, data yang digunakan adalah data *time series* dengan adanya unsur tren dan musiman, sama seperti studi kasus pada kali ini. XGBoost diharapkan dapat memberikan skalabilitas serta efisiensi yang lebih baik dari pendahulunya, yaitu *Gradient Boosting* dengan formalisasi model yang lebih teratur untuk mengendalikan *overfitting*, sehingga dapat memberikan kinerja yang lebih baik [11]. Dari hasil peramalan jumlah kasus yang didapat dengan metode XGBoost tersebut, selanjutnya dilakukan klasifikasi untuk menentukan status KLB dan non-KLB berdasarkan peraturan pemerintah yang berlaku menggunakan Klasifikasi Berbasis Aturan. Hasil dari peramalan tersebut diharapkan akan membantu Dinas Kesehatan Kabupaten Malang dalam menentukan langkah-langkah persiapan penanganan penyakit demam berdarah yang lebih baik dan tepat sasaran di masa yang akan datang.

1.2. Rumusan Masalah

Berdasarkan uraian latar belakang, maka rumusan masalah yang menjadi fokus dan diselesaikan dalam tugas akhir ini antara lain:

1. Bagaimana model peramalan Kejadian Luar Biasa demam berdarah menggunakan metode XGBoost dan Klasifikasi Berbasis Aturan.

2. Bagaimana tingkat akurasi peramalan Kejadian Luar Biasa demam berdarah menggunakan metode XGBoost dan Klasifikasi Berbasis Aturan.
3. Apa saja *insights* yang dapat dihasilkan dari segi metode maupun dari segi manajerial.

1.3. Batasan Permasalahan

Berdasarkan permasalahan yang disebutkan di atas, maka batasan masalah pada penelitian tugas akhir ini adalah sebagai berikut:

1. Penelitian ini menggunakan variabel bebas yaitu suhu, kelembapan, curah hujan, kecepatan angin, jumlah penduduk dan Angka Bebas Jentik (ABJ).
2. Data yang digunakan dalam penelitian ini adalah data sekunder terkait banyaknya kasus demam berdarah, kepadatan penduduk dan Angka Bebas Jentik (ABJ) yang didapatkan dari Dinas Kesehatan Kabupaten Malang. Sedangkan data iklim seperti suhu udara, curah hujan, kecepatan angin dan kelembapan udara didapatkan dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Karangploso dan Karangtretes.
3. Data akan dikategorikan menjadi tiga kategori berdasarkan letak geografis ketinggian tiap kecamatan yaitu dataran tinggi, dataran sedang dan dataran rendah.
4. Penelitian ini dilakukan untuk meramalkan Kejadian Luar Biasa demam berdarah pada Januari 2019 hingga Desember 2020.
5. Peramalan angka kejadian demam berdarah menggunakan metode Extreme Gradient Boosting sedangkan penentuan status KLB dan non-KLB menggunakan Klasifikasi Berbasis Aturan.
6. Penerapan klasifikasi dalam tugas akhir ini bukan klasifikasi sebagaimana di *data mining* atau *machine learning*. Dalam tugas akhir ini klasifikasi dilakukan

dengan menggunakan aturan-aturan yang telah ditentukan sebelumnya (*predefined rules*) mengacu pada peraturan pemerintah, bukan membangun aturan-aturan baru dari model.

1.4. Tujuan Penelitian

Tujuan dari pengerjaan penelitian tugas akhir ini adalah sebagai berikut:

1. Menemukan model peramalan Kejadian Luar Biasa demam berdarah menggunakan metode XGBoost dan Klasifikasi Berbasis Aturan.
2. Mengetahui tingkat akurasi peramalan Kejadian Luar Biasa demam berdarah menggunakan metode XGBoost dan Klasifikasi Berbasis Aturan.
3. Mengetahui *insights* yang dihasilkan dari percobaan baik dari segi metode maupun segi manajerial.

1.5. Manfaat Penelitian

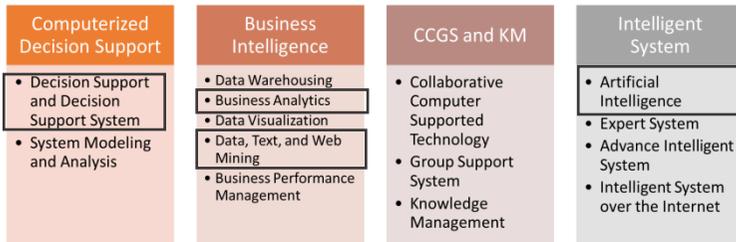
Manfaat yang diharapkan dapat diperoleh dari tugas akhir ini adalah:

1. Membantu Dinas Kesehatan Kabupaten Malang dalam mengatasi wabah penyakit demam berdarah berdasarkan hasil ramalan Kejadian Luar Biasa yang didapat. Sehingga, pihak terkait dapat mempersiapkan upaya pencegahan dengan lebih baik dan tepat sasaran.
2. Menambah sumber pengetahuan bagi peneliti lain mengenai penerapan metode Extreme Gradient Boosting (XGBoost) untuk peramalan deret berkala serta penerapan Klasifikasi Berbasis Aturan untuk melakukan klasifikasi Kejadian Luar Biasa suatu penyakit.

1.6. Relevansi

Hingga saat ini, penelitian dengan topik peramalan sangat dibutuhkan di berbagai sektor organisasi termasuk organisasi yang bergerak di bidang kesehatan. Salah satu penerapan topik

tersebut adalah peramalan jumlah kasus penyakit demam berdarah. Pada tugas akhir ini, peramalan kasus demam berdarah dilakukan dengan metode Extreme Gradient Boosting. Topik ini berkaitan dengan beberapa mata kuliah yaitu Statistika, Analitika Bisnis, Teknik Peramalan, Sistem Pendukung Keputusan dan Penggalian Data. Topik tugas akhir ini sesuai dengan bidang keilmuan Business Analytics, Decision Support dan Artificial Intelligence pada Laboratorium Rekayasa Data dan Intelegensi Bisnis (RDIB) di Departemen Sistem Informasi Institut Teknologi Sepuluh Nopember seperti yang tertera pada Gambar 1.1.



Gambar 1.1 Relevansi Tugas Akhir dengan Roadmap Lab RDIB

BAB II TINJAUAN PUSTAKA

Pada bab ini membahas penelitian sebelumnya yang berhubungan dengan tugas akhir dan teori-teori yang berkaitan dengan permasalahan tugas akhir.

2.1. Penelitian Sebelumnya

Tabel 2.1 menampilkan daftar penelitian sebelumnya yang mendasari tugas akhir ini.

Tabel 2.1 Studi Sebelumnya

1. Using Climate Factors to Predict the Outbreak of Dengue Fever [8]
Penulis/Tahun/Sumber : Nan Jia, Xianyi Liao, Jie Chen, Xiangping Chen, Jing Chen, Guanghui Dong, Kangkang Liu, Gang Hu; 2018
Metode : <i>- Extreme Gradient Boosting</i>
Kesimpulan : Penelitian ini berisi mengenai peramalan wabah demam berdarah di Guangzhou, Cina, dengan menggunakan algoritma XGBoost. Peramalan tersebut dilakukan dengan mempertimbangkan beberapa faktor iklim seperti suhu, kelembapan dan curah hujan. Hasil penelitian tersebut menyatakan bahwa nilai akurasi dengan menggunakan xgboost adalah yang terbaik dibandingkan algoritma lain seperti beberapa algoritma <i>machine learning</i> yang lain seperti SVM, Lasso, Linear Regression, Random Forest, AdaBoot, GBoost dan Rideg pada kasus peramalan wabah

demam berdarah tersebut, karena menghasilkan nilai RMSE dan MAE yang paling kecil.

Keterkaitan dengan Tugas Akhir :

Penelitian ini menjadi referensi penerapan metode XGBoost untuk diterapkan pada peramalan jumlah kasus demam berdarah serta dapat menjadi acuan bahwa faktor iklim dapat mempengaruhi jumlah kasus demam berdarah.

2. Forecasting the Yemen Cholera Outbreak with Machine Learning [10]

Penulis/Tahun/Sumber :

Rohil Badkundri, Victor Valbuena, Srikusmanjali Pinnamareddy, Brittney Cantrell, Janet Standeven; 2019

Metode :

- *Extreme Gradient Boosting*

Kesimpulan :

Penelitian ini membahas tentang peramalan wabah penyakit kolera di Yaman dengan menggunakan data tahun 2017 dan 2018. Data-data yang digunakan untuk menunjang prediksi adalah curah hujan, jumlah kasus kolera sebelumnya, angka kematian penyakit kolera, tingkat kematian perang saudara di Yaman dan data interaksi antar gubernur. Selain itu, penelitian ini juga membandingkan model yang dibangun dengan XGBoost dan Regresi Linier. Dari hasil perhitungan nilai akurasi yang dilakukan, XGBoost menjadi algoritma yang menghasilkan nilai *error* yang paling kecil.

Keterkaitan dengan Tugas Akhir:

Penggunaan metode XGBoost untuk melakukan peramalan *time series* pada sebuah wabah penyakit dengan melibatkan variabel-variabel independen tertentu.

3. A Regularization-Based eXtreme Gradient Boosting Approach in Foodborne Disease Trend Forecasting [9]

Penulis/Tahun/Sumber :

Shanen Chena, Jian Xub, Lili Chenb, Xi Zhanga, Li Zhangb, Jinfeng Lib; 2019

Metode :

- *Extreme Gradient Boosting*

Kesimpulan :

Penelitian ini membahas tentang peramalan penyakit yang diakibatkan oleh makan (*foodborne disease*) menggunakan algoritma XGBoost dengan melibatkan beberapa variabel seperti cahaya matahari, kelembapan relatif, suhu tanah, tekanan, curah hujan, suhu udara, kecepatan angin, kandungan PM_{2.5}, PM₁₀, SO₂, CO, NO₂, O₃ dan AQI. Penelitian ini juga membandingkan performa algoritma lain seperti SARIMA, LR, SVR, treebag dan gbm. Dari penelitian yang dilakukan, dapat diketahui bahwa XGBoost menghasilkan nilai error yang paling kecil dan lebih stabil dibandingkan lima algoritma lainnya.

Keterkaitan dengan Tugas Akhir:

Penelitian ini menjadi referensi penggunaan metode XGBoost untuk melakukan peramalan *time series* sebuah

wabah penyakit dengan melibatkan variabel-variabel independen terkait.

4. Peramalan KLB Campak Menggunakan Gabungan Metode JST Backpropagation dan CART [12]

Penulis/Tahun/Sumber :

Sulistyowati, Edi Winarko; 2014

Metode :

- JST dan CART

Kesimpulan :

Penelitian ini membahas tentang peramalan KLB penyakit campak menggunakan gabungan metode JST Backpropagation dan CART. Metode JST backpropagation digunakan untuk meramalkan jumlah kejadian campak, kemudian metode CART digunakan untuk melakukan penentuan KLB atau non-KLB di suatu daerah pada kurun waktu tertentu. Hasil dari penelitian ini menghasilkan nilai akurasi 90.01% untuk proses peramalan dengan JST, sedangkan CART mampu menentukan daerah KLB atau non-KLB dengan tingkat akurasi sebesar 83.33%.

Keterkaitan dengan Tugas Akhir :

Dalam penelitian tersebut, pendekatan yang dilakukan adalah melakukan peramalan jumlah kasus penyakit terlebih dahulu kemudian diklasifikasi menjadi KLB dan non-KLB. Maka dari itu, penelitian ini menjadi referensi pendekatan yang dapat digunakan dalam meramalkan KLB untuk suatu penyakit.

2.2. Landasan Teori

Landasan teori berisi teori-teori yang digunakan dalam pengerjaan penelitian tugas akhir. Dalam landasan teori, acuan yang digunakan adalah berdasarkan penelitian dan buku.

2.2.1. Kejadian Luar Biasa

Kejadian luar biasa atau yang biasa disingkat KLB (dalam bahasa Inggris : *outbreak*) merupakan suatu status yang diterapkan di Indonesia untuk pengklasifikasian penyakit yang sifatnya merebak dan dapat berkembang lebih jauh menjadi sebuah wabah penyakit. Sedangkan menurut Peraturan Menteri Kesehatan Republik Indonesia No. 1501/MENKES/PER/X/2010, KLB adalah munculnya atau meningkatnya kasus penyakit atau kematian secara epidemiologis pada suatu daerah dalam kurun waktu tertentu. Di Indonesia, istilah KLB dan wabah sering dianggap sama, namun keduanya memiliki makna yang agak berbeda. Wabah penyakit merujuk kepada kondisi dimana penyakit telah meluas dan parah terjadi. Sedangkan istilah KLB lebih mengarah ke peringatan dini sebelum wabah itu terjadi. Di Indonesia, suatu penyakit akan dikatakan berstatus KLB jika penyakit tersebut memenuhi beberapa kriteria yang telah diatur pada Pasal 6 Permenkes No. 1501/2010 dan dijelaskan lebih lanjut dalam Pedoman Penyelidikan dan Penanggulangan KLB Penyakit Menular dan Keracunan Pangan. KLB DBD ditetapkan bila ditemukan satu atau lebih kondisi berikut [4]:

- a. Timbulnya suatu penyakit menular tertentu sebagaimana dimaksud dalam Pasal 4 Permenkes No. 1501/2010, yang sebelumnya tidak ada atau tidak dikenal pada suatu daerah.
- b. Jumlah penderita baru dalam periode waktu 1 (satu) bulan menunjukkan kenaikan dua kali atau lebih dibandingkan dengan angka rata-rata per bulan dalam tahun sebelumnya.
- c. Angka kematian kasus suatu penyakit (Case Fatality Rate) dalam 1 (satu) kurun waktu tertentu menunjukkan

kenaikan 50% (lima puluh persen) atau lebih dibandingkan dengan angka kematian kasus suatu penyakit periode sebelumnya dalam kurun waktu yang sama.

2.2.2. Demam Berdarah Dengue

Demam berdarah dengue (DBD) atau lebih sering disebut dengan demam berdarah adalah penyakit yang disebabkan oleh virus Dengue. Penyakit ini pertama kali masuk ke Indonesia di Surabaya pada tahun 1968 lalu menyebar ke kota-kota lain seperti Jakarta, Bandung dan Yogyakarta. Pada tahun 1994, DBD telah menyebar ke 27 provinsi di Indonesia.

Penyakit ini dapat ditularkan melalui gigitan nyamuk spesies *Aedes Aegypti*. Beberapa gejala umum dari penyakit ini adalah demam tinggi, nyeri pada kepala, otot, dan sendi serta ruam pada kulit. Meskipun penyakit ini sering menjangkit anak-anak berusia kurang dari 15 tahun, tidak menutup kemungkinan bahwa orang dewasa juga dapat terjangkit penyakit tersebut. Penyakit ini dapat dikatakan berbahaya karena dapat menyebabkan kematian jika tidak dilakukan penanganan yang tepat [13].

2.2.3. Peramalan Time Series

Teknik peramalan digunakan dalam berbagai penelitian untuk mengukur atau memperkirakan suatu kondisi di masa yang akan datang. Sedangkan, *time series* atau deret waktu merupakan kumpulan pengamatan atau data statistik yang diurutkan secara sesuai dengan waktu. Waktu atau periode peramalan biasa disebut dengan lead time dan bernilai beragam sesuai dengan persoalan yang ingin dipecahkan. Sehingga, peramalan *time series* atau deret waktu merupakan metode peramalan yang menggunakan data yang berdasarkan deret waktu untuk mengetahui kondisi di masa depan [14].

Peramalan deret waktu terbagi menjadi dua jenis yakni *univariate* dan *multivariate*. Data deret waktu *univariate* mengacu pada data yang terdiri dari pengamatan tunggal yang

direkam secara berurutan selama kenaikan waktu yang sama, dimana dalam proses peramalannya variabel independen yang digunakan adalah nilai *lag* variabel itu sendiri. Sedangkan model deret waktu *multivariate* merupakan pengembangan dari model univariat yang melibatkan dua atau lebih variabel input lain. Proses peramalan *multivariate* biasa dilakukan untuk mempelajari keterkaitan antara variabel deret waktu tertentu [15].

2.2.4. Normalisasi Data

Normalisasi data dilakukan pada tahap pra-proses dengan mengubah skala data untuk menghilangkan sifat dominan hanya pada satu atau lebih variabel sehingga data dapat diproses dengan lebih baik. Normalisasi adalah proses transformasi atribut numerik menjadi *range* yang lebih kecil atau memiliki persebaran yang relatif normal [16]. Dalam melakukan normalisasi, terdapat beberapa metode atau teknik yang dapat dilakukan, seperti Normalisasi Z-Score, Min-max, dan Normalisasi dengan skala desimal.

Normalisasi Min-max adalah normalisasi yang dapat menjaga hubungan antar nilai dari data yang asli [17]. Normalisasi ini memetakan nilai v_i dari atribut A menjadi \hat{v}_i kedalam *range* baru $[new_min_A, new_max_A]$ seperti pada persamaan 2.1.

$$\hat{v}_i = \frac{v_i - min_A}{max_A - min_A} (new_max_A - new_min_A) + new_min_A \quad (2.1)$$

Dimana :

- v_i' = nilai data setelah normalisasi
- v_i = nilai data sebelum normalisasi
- min_A = nilai minimum pada data sebelum normalisasi
- max_A = nilai maksimum pada data sebelum normalisasi
- new_min_A = nilai minimum pada data setelah normalisasi
- new_max_A = nilai maksimum pada data setelah normalisasi

2.2.5. Uji Korelasi

Uji korelasi dilakukan untuk mengetahui seberapa kuat suatu atribut menyiratkan atribut lainnya berdasarkan pada data yang tersedia. Pengujian korelasi pada data yang sifatnya nominal dapat menggunakan uji χ^2 (chi-square) sedangkan untuk data numerik dapat menggunakan koefisien korelasi dan kovarians. Koefisien korelasi atau yang biasa disebut juga sebagai *Pearson Product Moment* yang bernilai mulai dari -1 sampai 1. Jika nilai koefisien korelasi lebih dari 0, maka atribut A dan B memiliki korelasi yang positif, dimana jika nilai A meningkat maka nilai B juga akan meningkat. Namun, jika nilai koefisien korelasi kurang dari 0, maka atribut A dan B memiliki korelasi yang negatif. Selain itu, nilai koefisien korelasi sama dengan 0 menunjukkan bahwa kedua atribut tersebut tidak saling berkaitan atau independen [17]. Koefisien korelasi dirumuskan pada persamaan 2.2.

$$r_{A,B} = \frac{\sum_{i=1}^n (a_i - \bar{A})(b_i - \bar{B})}{n\sigma_A\sigma_B} = \frac{\sum_{i=1}^n (a_i - b_i) - n\bar{A}\bar{B}}{n\sigma_A\sigma_B} \quad (2.2)$$

Dimana :

- $r_{A,B}$ = koefisien korelasi atribut A dan B
- n = jumlah data
- a_i = nilai dari atribut A pada data ke- i
- b_i = nilai dari atribut B pada data ke- i
- \bar{A} , \bar{B} = rata-rata nilai atribut A dan atribut B
- σ_A , σ_B = standar deviasi atribut A dan atribut B

2.2.6. Angka Bebas Jentik (ABJ)

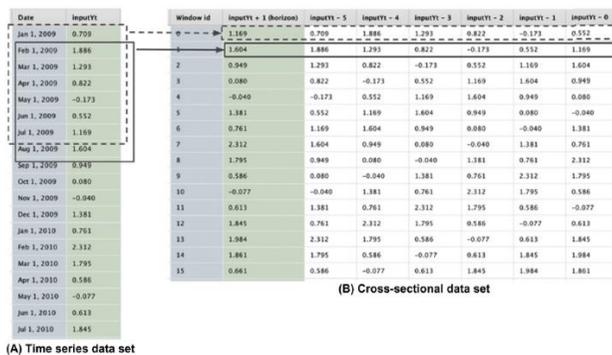
Angka Bebas Jentik merupakan ukuran yang digunakan untuk mengetahui kepadatan jentik. ABJ dapat diketahui dengan menghitung jumlah atau bangunan yang tidak dijumpai jentik dibagi dengan seluruh jumlah rumah atau bangun pada suatu daerah. Sedangkan keadaan bebas jentik merupakan suatu keadaan dimana ABJ lebih atau sama dengan 95%. ABJ merupakan salah satu indikator yang digunakan dalam upaya pengendalian penyakit DBD, karena keadaan bebas jentik dapat

merepresentasikan bahwa jumlah telur, jentik dan kempompong nyamuk penular DBD (*Aedes aegypti*) berkurang atau tidak ada. Rumus yang digunakan dalam perhitungan ABJ tertera pada persamaan 2.3 [1].

$$ABJ = \frac{\text{Jumlah rumah tidak ditemukan jentik}}{\text{Jumlah rumah diperiksa}} \times 100\% \quad (2.3)$$

2.2.7. Windowing

Windowing adalah teknik yang digunakan untuk mentransformasi data deret berkala menjadi dataset input mesin pembelajaran yang bersifat *cross-sectional*. Proses *windowing* tercantum pada Gambar 2.1 dimana data deret berkala (A) yang semula dijabarkan per baris sesuai dengan periode tertentu, kemudian dilakukan ekstraksi data menjadi data *cross-sectional* (B) dengan menambahkan beberapa kolom sesuai dengan jumlah *lag* yang ditentukan. Karakteristik *windows* dan ekstraksi data *cross-sectional* ditentukan oleh beberapa parameter yaitu *window size*, *step size*, dan *horizon width* [14].

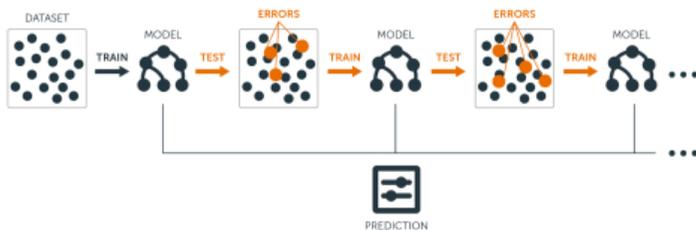


Gambar 2.1 Proses Windowing

2.2.8. Extreme Gradient Boosting

Extreme Gradient Boosting atau yang biasa disingkat dengan XGBoost merupakan salah satu algoritma metode *ensemble*, yakni metode dalam *machine learning* yang menggabungkan

beberapa model dasar untuk menghasilkan satu model prediksi yang optimal [18]. Seperti yang tercantum dalam namanya, Extreme Gradient Boosting mengimplementasikan kerangka berpikir *gradient boosting*, dimana model pohon keputusan dibuat secara berurutan untuk melengkapi model dasar yang sebelumnya telah dibangun dengan prinsip penurunan gradien seperti yang diilustrasikan pada Gambar 2.2.



Gambar 2.2 Ilustrasi Gradient Boosting

Hal mendasar yang membedakan Extreme Gradient Boosting dengan algoritma Gradient Boosting standar yaitu XGBoost menggunakan formalisasi model dengan reguralisasi yang lebih teratur untuk mengontrol *over-fitting*, yang dapat memberikan performa model yang lebih baik [19]. XGBoost dapat digunakan untuk menyelesaikan permasalahan klasifikasi maupun regresi. Dalam penelitian ini, XGBoost akan digunakan untuk menyelesaikan masalah regresi.

Pada permasalahan regresi, hasil prediksi nilai \hat{y}_i didapatkan dari penjumlahan skor yang diprediksi dari semua *tree* berjumlah K yang dirumuskan pada Persamaan 2.4.

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (2.4)$$

Dimana x_i adalah sampel pelatihan yakni variabel independen ke i , dan $f_k(x_i)$ adalah skor dihasilkan dari pohon ke- k . Sedangkan F merupakan himpunan dari fungsi yang berisi semua pohon regresi yang terbentuk.

Fungsi objektif model regresi XGBoost ditunjukkan pada Persamaan 2.5, dimana $\sum_{j=1}^{T_k} L(f_k)$ merupakan *loss function* dan $\sum_{j=1}^{T_k} \Omega(f_k)$ merupakan *regulation term*.

$$Obj(\theta) = \sum_{j=1}^{T_k} L(f_k) + \sum_{j=1}^{T_k} \Omega(f_k) \quad (2.5)$$

Loss function berperan untuk mengukur kesalahan model dalam memprediksi suatu data, sehingga dapat diketahui apakah model dapat memprediksi dengan baik. Penjabaran rumus *loss function* terdapat pada Persamaan 2.6. Dimana G_{jk} merupakan statistik gradien urutan pertama dari *loss function* (*sum of gradient*), sedangkan H_{jk} statistik gradien urutan kedua dari *loss function* (*sum of hessian*), dan w adalah bobot.

$$L(f_k) = \sum_{j=1}^{T_k} [G_{jk} w_{jk} + \frac{1}{2} H_{jk} w_{jk}^2] \quad (2.6)$$

Regulation term berperan untuk mengontrol seberapa rumit model yang dibentuk. *Regulation term* menggunakan mekanisme pemberian nilai penalti untuk kompleksitas model, sehingga model dapat terhindar dari *overfitting*. Detail rumus ditampilkan pada Persamaan 2.7, dimana T adalah total jumlah *leaves* pada pohon yang terbentuk, w adalah bobot setiap *leaves*, sedangkan γ , λ , dan α adalah koefisien konstan.

$$\Omega(f_k) = \gamma T_k + \frac{1}{2} \lambda \sum_{j=1}^{T_k} w^2_{jk} + \alpha \sum_{j=1}^{T_k} |w_{jk}| \quad (2.7)$$

XGBoost menggunakan metode pelatihan aditif dan iteratif dalam melakukan proses pembelajarannya. Pembelajaran dimulai dari prediksi konstan dan menambahkan fungsi baru setiap kali iterasi. Fungsi objektif secara utuh ditampilkan pada Persamaan 2.8.

$$Obj(\theta) = \sum_{j=1}^{T_k} [G_{jk} w_{jk} + \frac{1}{2} H_{jk} w_{jk}^2] + \gamma T_k + \frac{1}{2} \lambda \sum_{j=1}^{T_k} w_{jk}^2 + \alpha \sum_{j=1}^{T_k} |w_{jk}| \quad (2.8)$$

Implementasi dari XGBoost untuk suatu *dataset* dapat dilakukan dengan beberapa langkah [19] :

1. Menghitung prediksi awal dengan merata-rata nilai variabel dependen.
2. Menghitung nilai residual yakni selisih prediksi awal tersebut dengan data aktual.
3. Membuat pohon keputusan untuk memprediksi nilai residual tersebut, dimana pohon keputusan tersebut terdiri dari *nodes* yang berisi variabel independen sebagai *parent node* dan variabel dependen pada *leaf node*, dengan cara :
 - a. Menghitung nilai *similarity score* untuk *root node* dan *child node* dengan rumus seperti pada persamaan 2.9

$$Similarity\ Score = \frac{Sum\ of\ Residuals^2}{Number\ of\ Residuals + \lambda} \quad (2.9)$$

- b. Menghitung nilai Gain untuk menentukan bagaimana pembagian(*split*) data yang terbaik. Perhitungan membutuhkan nilai dari *similarity score* pada *child node* yang terdiri dari node kiri (Left) dan node kanan (Right) dan *similarity score* pada *root node*. Perhitungan Gain dapat dilakukan dengan menggunakan formula pada persamaan 2.10.

$$Gain = Left_{similarity} + Right_{similarity} - Root_{similarity} \quad (2.10)$$

Dari nilai Gain yang telah dihitung, dapat diketahui skenario *split* yang paling optimal yakni merujuk pada nilai Gain yang paling besar. Maka model akan menggunakan skenario *split* data yang paling optimal tersebut.

- c. Melakukan pemangkasan(*pruning*) pohon keputusan yang telah dibangun dengan mendefinisikan nilai γ terlebih dahulu yang merupakan parameter untuk mengukur kompleksitas pohon keputusan. Selanjutnya dilakukan perhitungan antara nilai Gain dan γ seperti yang tertera pada persamaan 2.11.

$$Gain - \gamma = \begin{cases} \text{if positive, then do not prune} \\ \text{if negative, then prune} \end{cases} \quad (2.11)$$

- d. Menghitung *output value* untuk tiap *leaf nodes* pada pohon keputusan yang sudah dilakukan *pruning* dengan persamaan pada 2.12.

$$Output Value = \frac{Sum\ of\ Residuals}{Number\ of\ Residuals + \lambda} \quad (2.12)$$

- e. Melakukan perhitungan prediksi dengan menjumlahkan nilai prediksi awal (langkah 1) ditambah hasil prediksi dari pohon keputusan yang terbentuk yang telah dikalikan dengan *learning rate* (nilai penyusutan).
- f. Hitung kembali nilai residual dari hasil prediksi yang terbaru, lalu buat kembali pohon keputusan lainnya.
- g. Ulangi langkah a-f hingga mencapai titik akhir ronde *boosting* yang ditandai dengan nilai residual yang kecil, atau telah mencapai jumlah maksimum pohon keputusan yang terbentuk.
4. Hitung hasil prediksi akhir dengan menggabungkan hasil dari semua pohon keputusan yang terbentuk.

2.2.9. Klasifikasi Berbasis Aturan

Klasifikasi merupakan mekanisme pemetaan objek-objek ke dalam salah satu dari beberapa kategori yang telah ditentukan sebelumnya. Tujuan klasifikasi adalah untuk memprediksi secara akurat kelas target untuk setiap kasus dalam data. Klasifikasi bersifat diskrit dan tidak mempertimbangkan urutan [20]. Untuk melakukan proses klasifikasi banyak metode yang dapat dilakukan, salah satunya adalah Klasifikasi Berbasis

Aturan atau *rule-based classification*. Klasifikasi Berbasis Aturan dapat digunakan untuk merujuk pada skema klasifikasi apapun yang menggunakan aturan jika-maka untuk prediksi kelas [21]. Dalam penelitian ini, konteks klasifikasi berbasis aturan bukan dalam sudut pandang *machine learning* yang melakukan ekstraksi aturan dari data terlebih dahulu, melainkan hanya menerapkan aturan-aturan yang telah ditentukan sebelumnya (*predefined rules*).

2.2.10. Pengukuran Performa Peramalan

Suatu peramalan dapat dievaluasi dengan melihat nilai akurasi peramalan. Nilai akurasi peramalan didapatkan dengan membandingkan nilai aktual dan nilai hasil peramalan yang telah dilakukan. Sedangkan, nilai error atau kesalahan dalam suatu peramalan merepresentasikan selisih dari nilai aktual dan hasil peramalan untuk periode yang sama [22]. Sejalan dengan hal tersebut, ada beberapa nilai ukuran yang bisa digunakan untuk mengukur nilai akurasi sebuah peramalan.

a. RMSE

Salah satu pengukuran yang dapat digunakan adalah *Root Mean Square Error* (RMSE) [23]. RMSE merupakan hasil akar dari rata-rata dari keseluruhan nilai error yang sudah dikuadratkan seperti pada persamaan 2.13. Hasil ramalan yang baik akan menghasilkan nilai RMSE yang kecil, sedangkan hasil ramalan yang buruk akan menghasilkan nilai RMSE yang besar.

$$RMSE = \sqrt{\frac{\sum_{t=1}^N (A_t - F_t)^2}{n}} \quad (2.13)$$

Dimana :

- n = banyaknya data
- A_t = nilai aktual indeks pada periode ke t
- F_t = nilai *forecasting* data pada periode ke t

b. SMAPE

Symmetric Mean Absolute Percentage Error (SMAPE) merupakan pengembangan dari konsep *Mean Absolute Percentage Error* (MAPE), dimana keduanya merupakan ukuran akurasi berdasarkan persentase kesalahan [24]. Kedua pengukuran tersebut memiliki rumus yang berbeda, dimana rumus MAPE ditunjukkan pada Persamaan 2.14 sedangkan SMAPE ditunjukkan pada Persamaan 2.15. Seperti yang dapat dilihat pada Persamaan 2.14, data aktual dan ramalan yang bernilai 0 akan menghasilkan nilai MAPE yang tidak *valid* karena memiliki formula dengan pembagi nilai ramalan. Sehingga, SMAPE dapat digunakan sebagai alternatif dari MAPE untuk menanggulangi keterbatasan tersebut karena mampu menghitung akurasi. SMAPE dapat digunakan untuk mengukur tingkat kesalahan dari model pada data lain dengan pola tren sama atau berbeda meskipun data aktual dan ramalan bernilai 0. Semakin rendah nilai SMAPE menunjukkan model peramalan memiliki kemampuan yang baik.

$$MAPE = \sum_{t=1}^n \left| \frac{A_t - F_t}{F_t} \right| \times 100\% \quad (2.14)$$

$$SMAPE = \frac{100\%}{n} \sum_{t=1}^n \frac{|F_t - A_t|}{|A_t| + |F_t|} \quad (2.15)$$

Dimana

- n = banyaknya data
- A_t = nilai aktual indeks pada periode ke t
- F_t = nilai forecasting data pada periode ke t

2.2.11. Pengukuran Performa Klasifikasi

Pengukuran performa klasifikasi dapat dilakukan dengan menggunakan *confusion matrix*. *Confusion matrix* merupakan matriks yang menunjukkan data aktual dan hasil klasifikasi yang diprediksi. *Confusion matrix* berukuran $L \times L$, di mana L adalah jumlah nilai label kelas. Beberapa istilah pada *confusion matrix* untuk kelas biner seperti yang tertera pada Tabel 2.2 yaitu True

Positive (TP), True Negative (TN), False Positiv (FP) dan False Negative (FN). True Negative (TN) adalah nilai yang merepresentasikan data negatif yang diklasifikasikan dengan benar, sedangkan False Positive (FP) merepresentasikan data negatif yang diklasifikasikan sebagai data positif. Sementara itu, True Positive (TP) merupakan data positif yang diklasifikasikan dengan benar dan False Negative (FN) merepresentasikan data negatif namun diklasifikasikan sebagai data positif [25].

Tabel 2.2 Confusion Matrix

		Hasil Prediksi	
		Positif	Negatif
Data Aktual	Positif	True Positive (TP)	False Negative (FN)
	Negatif	False Positive (FP)	True Negative (TN)

a. Akurasi

Nilai akurasi merepresentasikan efektivitas keseluruhan dari sebuah metode klasifikasi. Untuk klasifikasi kelas biner, akurasi dapat dihitung dengan menjumlahkan TP dan TN kemudian dibagi dengan keseluruhan nilai pada matriks. Rumus akurasi ditunjukkan pada persamaan 2.16.

$$Akurasi = \frac{TP + TN}{TP + FN + FP + TN} \quad (2.16)$$

b. Spesifisitas

Spesifisitas merepresentasikan seberapa efektif classifier mengidentifikasi label negatif. Pada klasifikasi kelas biner spesifisitas dapat dihitung dengan membagi nilai TN dengan penjumlahan FP dan TN. Rumus spesifisitas ditunjukkan pada persamaan 2.17.

$$Specificity = \frac{TN}{FP + TN} \quad (2.17)$$

c. Sensitivitas

Sensitivitas merepresentasikan seberapa efektif classifier mengidentifikasi label positif. Pada klasifikasi kelas biner

sensitivitas dapat dihitung dengan membagi nilai TP dengan penjumlahan TP dengan FN. Rumus sensitivitas ditunjukkan pada persamaan 2.18.

$$Sensitivity = \frac{TP}{TP + FN} \quad (2.18)$$

Halaman ini sengaja dikosongkan

BAB III

METODOLOGI PENELITIAN

Bab Metodologi Penelitian menjelaskan tentang alur penelitian yang dilakukan beserta penjelasan untuk masing-masing tahapan.

3.1. Metodologi

Sub-bab ini menjelaskan tentang alur penelitian yang meliputi input, proses, serta output yang akan dihasilkan seperti tercantum pada Gambar 3.1.

3.2. Uraian Metodologi

Berikut adalah uraian dari metodologi yang digunakan sebagaimana yang digambarkan pada bagian sebelumnya.

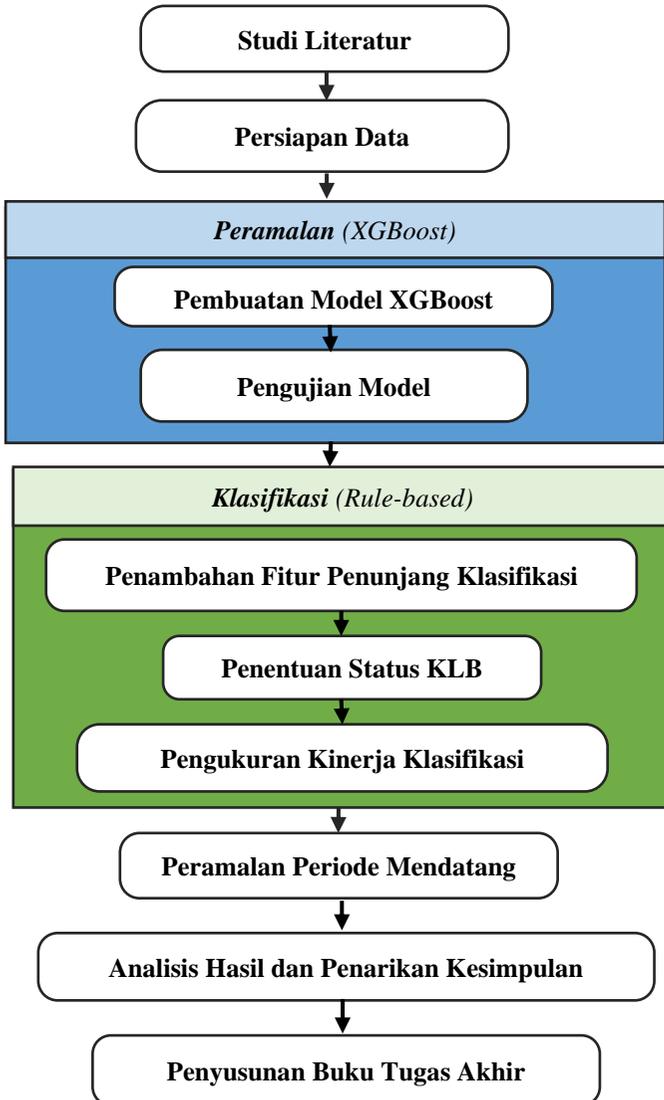
Metodologi yang digunakan pada penelitian ini mengadaptasi metode *Extreme Gradient Boosting* yang disesuaikan dengan kebutuhan penelitian dan pengembangan.

3.2.1. Studi Literatur

Studi literatur merupakan tahap awal dari penelitian ini, dimana pada tahap ini berfokus untuk mengumpulkan dan memahami berbagai referensi seperti artikel, jurnal dan buku yang berkaitan dengan objek penelitian kali ini yaitu KLB demam berdarah, lingkup permasalahan yakni peramalan *time series* dan metode yang digunakan yaitu pra-proses data, XGBoost dan Klasifikasi Berbasis Aturan. Tujuan dari dilakukannya tahap ini yaitu untuk lebih memahami ilmu dan teori yang mendasari penelitian yang hendak dilakukan.

3.2.2. Persiapan Data

Tahap ini berfokus untuk mempersiapkan data yang hendak digunakan dalam penelitian ini. Persiapan data yang dilakukan adalah pengumpulan data dan melakukan pra-proses data.



Gambar 3.1 Alur Kerja Penelitian

a. Pengumpulan Data

Pada tahap ini akan dilakukan pengambilan data kasus demam berdarah, kepadatan penduduk, Angka Bebas Jentik (ABJ) pada beberapa kecamatan di Kabupaten Malang dari Dinas Kesehatan Kabupaten Malang. Berkaitan dengan itu, data penunjang lain seperti data iklim yaitu suhu, kelembaban, curah hujan, dan kecepatan angin diambil dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Karangploso dan Karangates. Data yang dikumpulkan adalah pada periode Januari 2010 - Desember 2018 untuk beberapa kecamatan di Kabupaten Malang dengan dikelompokkan menjadi tiga kategori yaitu dataran tinggi, dataran sedang dan dataran rendah. Dataran tinggi terdiri dari kecamatan-kecamatan dengan ketinggian 500 – 700 mdpl, dataran sedang pada ketinggian 200 – 500 mdpl, dan dataran rendah pada ketinggian 0 – 200 mdpl.

b. Pra-Proses Data

Setelah semua data dikumpulkan, langkah selanjutnya adalah melakukan pra-proses data seperti normalisasi data, windowing, pemilihan fitur, dan pembagian data.

i. Normalisasi Data

Pada tahap ini proses normalisasi dilakukan terhadap data yang sebelumnya telah dihimpun. Normalisasi ini perlu dilakukan karena skala nilai yang ada pada data masih tidak seimbang, mengingat data iklim dan Angka Bebas Jentik (ABJ) berskala satuan serta puluhan, namun kepadatan penduduk berskala ratusan ribu. Metode normalisasi data yang akan digunakan pada penelitian ini adalah metode *min-max* sesuai dengan penjelasan pada subbab 2.2.4.

ii. Windowing

Tahap kedua pra-proses data yang akan dilakukan adalah *windowing*. Pendekatan ini dilakukan untuk mentransformasi data deret berkala menjadi data *cross-sectional* agar selanjutnya dapat menjadi input dari algoritma pembelajaran mesin yang

digunakan untuk peramalan. *Windowing* dapat dilakukan dengan menambahkan fitur *lag* pada data seperti yang telah tercantum pada subbab 2.2.6.

iii. Pemilihan Fitur

Pemilihan fitur yang digunakan untuk proses peramalan selanjutnya dapat dilakukan dengan melakukan uji korelasi. Uji korelasi dilakukan untuk mengetahui seberapa besar pengaruh variabel independen yang digunakan terhadap data yang akan diramalkan. Pada tahapan ini akan dilakukan pengujian korelasi variabel iklim (suhu, kelembapan, curah hujan, dan kecepatan angin), jumlah penduduk dan angka bebas jentik terhadap data jumlah kasus demam berdarah berdasarkan tiap kecamatan di Kabupaten Malang. Dari hasil uji korelasi yang dilakukan, maka dapat dipilih variabel independen yang akan digunakan untuk melakukan peramalan selanjutnya.

iv. Pembagian Data

Pada tahap ini data dibagi menjadi dua bagian yaitu data pelatihan dan data pengujian. Data pelatihan digunakan untuk membangun model peramalan, sedangkan data pengujian digunakan untuk memvalidasi seberapa baik model yang telah dibangun untuk melakukan peramalan.

3.2.3. Tahap Peramalan

Tahap peramalan berfokus dalam melakukan peramalan dengan model XGBoost yang optimal. Tahap ini terdiri dari tiga sub-tahap yaitu pembuatan model XGBoost, pengujian model, serta peramalan jumlah kasus demam berdarah.

a. Pemodelan

Tahap ini berfokus dalam melakukan pelatihan dengan membangun model XGBoost dari data pelatihan (*train set*) yang dimiliki. Model dibangun dengan menggunakan data dari beberapa skenario pembagian data agar model yang dihasilkan memiliki *error* serendah mungkin. Model awal dibentuk

menggunakan parameter *default* sebelum nantinya dilakukan optimalisasi parameter pada tahap berikutnya.

b. Mengoptimalkan Parameter

Pengoptimalan parameter perlu dilakukan untuk menghasilkan model yang lebih baik. Model XGBoost memiliki beberapa parameter yang bisa dioptimalkan sesuai dengan penggunaannya yang dapat mengatur seberapa kompleks model akan dibangun.

c. Pengujian Model Peramalan

Model yang sebelumnya telah dibangun kemudian digunakan untuk memprediksi jumlah kasus demam berdarah pada data pengujian (*test set*). Kemudian hasil dari prediksi tersebut dibandingkan dengan nilai aktual sehingga menghasilkan nilai RMSE. Model yang nantinya akan dipilih adalah model dengan nilai RMSE paling kecil diantara semua model yang terbentuk.

3.2.4. Tahap Klasifikasi

Tahap klasifikasi berfokus untuk mengklasifikasi data hasil peramalan pada tahap sebelumnya menjadi dua kelas yaitu KLB dan non-KLB. Tahap ini terdiri dari tiga sub-tahap yaitu penambahan fitur penunjang klasifikasi, penentuan status KLB dan non-KLB serta menghitung nilai akurasi hasil klasifikasi.

a. Penambahan Fitur Penunjang Klasifikasi

Sebelum melakukan penentuan status KLB dan non-KLB, perlu dilakukan penambahan fitur penunjang baru. Fitur baru ini akan digunakan dalam pendefinisian aturan-aturan (*rules*) untuk mempermudah jalannya proses klasifikasi. Fitur yang dimaksud dalam tahap ini adalah tingkat kenaikan suatu kasus demam berdarah dibandingkan dengan angka rata-rata per bulan dalam tahun sebelumnya.

b. Penentuan Status KLB

Tahap ini bertujuan untuk mengklasifikasikan data ke dalam dua kelas yaitu KLB dan non-KLB dengan mekanisme *rule-based*.

Penelitian ini melibatkan dua aturan yang digunakan dalam proses klasifikasi, yakni persyaratan (a) dan (b) untuk pemberian status KLB demam berdarah pada suatu daerah dalam kurun waktu tertentu yang tercantum pada subbab 2.2.1.

c. Pengukuran Kinerja Klasifikasi

Setelah melakukan pemberian label KLB dan non-KLB pada data, maka langkah selanjutnya yaitu menghitung nilai akurasi dari hasil klasifikasi atau pelabelan tersebut. Evaluasi hasil klasifikasi dilakukan dengan membangun *confusion matrix* dari hasil prediksi data pengujian (*test set*), diikuti dengan perhitungan nilai *akurasi*, spesifisitas dan sensitivitas.

3.2.5. Peramalan Periode Mendatang

Peramalan dilakukan dengan menggunakan model terbaik yang telah dipilih pada tahap sebelumnya. Keluaran dari tahap ini adalah hasil ramalan jumlah kasus demam berdarah yang telah berlabel KLB dan non-KLB pada tahun 2019-2020.

3.2.6. Analisa Hasil dan Penarikan Kesimpulan

Setelah melakukan klasifikasi, langkah selanjutnya adalah menganalisa hasil keseluruhan penelitian yang sudah dilakukan dan penarikan kesimpulan akhir. Analisa dilakukan berdasarkan hasil evaluasi peramalan dan klasifikasi yang sudah dilakukan sebelumnya, sehingga dari hasil analisa tersebut dapat ditarik kesimpulan apakah metode Extreme Gradient Boosting dan Klasifikasi Berbasis Aturan cocok digunakan untuk meramalkan KLB demam berdarah pada Kabupaten Malang.

3.2.7. Penyusunan Buku Tugas Akhir

Penyusunan buku tugas akhir dilakukan setelah mendapatkan hasil akhir penelitian. Tujuan penyusunan buku tugas akhir ini agar hasil penelitian dapat didokumentasikan dengan baik sehingga dapat menjadi rujukan atau referensi penelitian selanjutnya.

BAB IV PERANCANGAN

Bab ini berisikan penjelasan tentang perancangan model peramalan dengan menggunakan XGBoost dan Klasifikasi Berbasis Aturan. Model yang akan dibentuk disesuaikan dengan langkah-langkah yang telah dijelaskan pada bab Metodologi. Rancangan yang dibutuhkan yaitu meliputi proses persiapan data, pembuatan model XGBoost, pengujian model, peramalan jumlah kasus demam berdarah, penambahan fitur penunjang klasifikasi, penentuan status KLB dan non-KLB dan pengukuran kinerja klasifikasi.

4.1. Persiapan data

Dalam penelitian kali ini tahap pertama yang perlu dilakukan adalah mempersiapkan data. Persiapan data meliputi pengumpulan dan pra-proses data.

4.1.1. Pengumpulan Data

Data yang digunakan pada tugas akhir ini adalah data jumlah kasus demam berdarah pada 16 kecamatan di Kabupaten Malang yang telah dikelompokkan menjadi tiga kategori dataran, yaitu dataran rendah, dataran sedang, dan dataran tinggi. Data tersebut didapatkan dari Dinas Kabupaten Malang pada kurun waktu Januari 2010 hingga Desember 2018, sehingga diperoleh data jumlah kasus demam berdarah sebanyak 108 data per kecamatan.

Selain itu, data lain yang diperlukan adalah data variabel independen untuk proses peramalan, antara lain data iklim (suhu, curah hujan, kecepatan angin, dan kelembapan udara), jumlah penduduk dan Angka Bebas Jentik (ABJ) untuk tiap kecamatan. Data iklim diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) melalui situs BMKG. Data iklim yang didapatkan adalah data harian mulai tanggal 1 Januari 2010 hingga 31 Desember 2018 sehingga terdapat 3650 data untuk suhu, curah hujan, kecepatan angin, dan kelembapan

udara untuk tiap daerah. Selain itu, data Angka Bebas Jentik diperoleh dari Dinas Kabupaten Malang dalam bentuk periode bulanan. Sedangkan, data sekunder untuk jumlah penduduk yang didapatkan adalah data bulanan periode 2009 hingga 2018 per kecamatan.

4.1.2. *Pra-Processing Data*

Tahap pra-proses data perlu dilakukan agar proses pembentukan model dapat berjalan dengan baik dan sesuai prosedur. Terdapat beberapa langkah yang dilakukan pada proses ini, yaitu mengubah data variabel iklim, penanganan *missing value*, pemilihan kecamatan pembentuk model, normalisasi data, *windowing*, pemilihan fitur dan pembagian data.

a. Mengubah Data Harian Variabel Iklim

Data iklim (suhu, curah hujan, kecepatan angin dan kelembapan udara) yang didapatkan dari situs Badan Meteorologi, Klimatologi dan Geofisika (BMKG) masih dalam rentang waktu harian, sehingga perlu dilakukan pengubahan rentang waktu menjadi bulanan agar dapat diproses lebih lanjut. Pengubahan data tersebut dapat dilakukan dengan merata-rata data per bulan agar membentuk data dengan periode bulanan seperti pada persamaan 4.1.

$$\text{Rata-rata perbulan} = \frac{1}{n} \sum_{1}^n a \quad (4.1)$$

Dimana :

- n = jumlah hari pada bulan tertentu
- a = nilai variabel iklim

b. Penanganan *Missing Value*

Data yang didapatkan memiliki beberapa *missing value* khususnya pada variabel iklim dan Angka Bebas Jentik. Pada penelitian kali ini, data iklim yang kosong akan diisi nilainya dengan mekanisme perataan yaitu menggunakan nilai rata-rata perbulan. Hal ini dilakukan dengan pertimbangan data iklim yang bersifat musiman. Sedangkan, pada variabel Angka Bebas

Jentik terdapat dua macam pengisian *missing value*, yaitu interpolasi-ekstrapolasi biasa dan regresi linier. Angka Bebas Jentik yang kosong kurang dari 24 periode berturut-turut pada awal/akhir data akan diisi dengan nilai yang dihasilkan dari metode interpolasi dan ekstrapolasi biasa. Sedangkan, pada kasus variabel Angka Bebas Jentik yang memiliki nilai kosong lebih dari 24 periode berturut-turut pada awal/akhir data akan menggunakan metode regresi linier.

c. Pemilihan Kecamatan Pembentuk Model

Data yang akan digunakan untuk pembentukan model pada penelitian kali ini hanya data satu kecamatan terpilih di tiap dataran. Kemudian dari hasil model tersebut dilakukan uji coba untuk diterapkan pada kecamatan lain pada dataran yang sama. Hal tersebut dilakukan untuk menguji apakah model yang terbentuk bersifat *robust* yang berarti model tersebut dapat digunakan untuk data kecamatan yang lain.

Pemilihan kecamatan yang akan digunakan untuk proses pembentukan model dilakukan berdasarkan hasil perhitungan korelasi antarkecamatan di dataran yang sama. Kecamatan dengan nilai korelasi yang paling tinggi akan dipilih sebagai kecamatan pembentuk model.

d. Normalisasi Data

Data yang digunakan dalam penelitian kali ini memiliki skala nilai yang berbeda-beda, sehingga data tersebut perlu dilakukan normalisasi data. Data iklim dan Angka Bebas Jentik (ABJ) memiliki skala satuan dan puluhan, sedangkan jumlah penduduk berskala ribuan.

Metode normalisasi data yang digunakan pada penelitian ini adalah metode normalisasi min-max, dengan data hasil normalisasi yang dihasilkan bernilai 0 hingga 1. Variabel akan semakin mendekati 0 jika nilai mendekati angka minimum dan mendekati 1 untuk nilai yang mendekati angka maksimum. Nilai variabel yang baru didapatkan dari perhitungan dengan rumus yang telah tertera pada subbab 2.2.4.

e. Pemilihan Fitur dengan Uji Korelasi

Terdapat enam variabel independen yang digunakan pada tugas akhir ini, yaitu variabel suhu, curah hujan, kecepatan angin, kelembapan udara, Angka Bebas Jentik (ABJ), dan jumlah penduduk. Keenam variabel tersebut kemudian dibagi menjadi lima kategori skenario pemilihan fitur pada data seperti pada Tabel 4.1. Dari keempat skenario pemilihan fitur tersebut, skenario TOP3-VI dan TOP1-VI membutuhkan uji korelasi terlebih dahulu untuk menentukan variabel-variabel yang berpengaruh tiap kecamatan pembangun model.

Tabel 4.1 Skenario Pemilihan Fitur

No	Skenario	Deskripsi
1	ALL-VI	Fitur yang digunakan untuk membangun model adalah keseluruhan variabel independen yaitu Suhu, Curah Hujan, Kelembapan Udara, Kecepatan Angin, Angka Bebas Jentik (ABJ), dan Jumlah Penduduk
2	NO-VI	Fitur yang digunakan untuk membangun model tidak melibatkan variabel independen utama, melainkan hanya menggunakan fitur <i>lag</i> kejadian demam berdarah
3	TOP3-VI	Fitur yang digunakan untuk membangun model hanya tiga variabel independen yang paling berpengaruh terhadap jumlah kasus demam berdarah
3	TOP2-VI	Fitur yang digunakan untuk membangun model hanya dua variabel independen yang paling berpengaruh terhadap jumlah kasus demam berdarah
4	TOP1-VI	Fitur yang digunakan untuk membangun model hanya satu variabel independen yang paling berpengaruh terhadap jumlah kasus demam berdarah

Uji korelasi pada penelitian kali ini melibatkan variabel independen sebagai variabel yang akan diuji hubungannya terhadap angka kejadian demam berdarah. Uji korelasi tersebut

menggunakan koefisien korelasi *pearson* dengan rentang nilai 0 hingga 1. Nilai mendekati 0 menunjukkan variabel kurang berpengaruh secara signifikan terhadap variabel angka kejadian demam berdarah dan nilai mendekati 1 menunjukkan variabel berpengaruh terhadap variabel angka kejadian demam berdarah.

f. Windowing

Data deret berkala yang telah dinormalisasi perlu diubah menjadi data *cross-sectional* dengan menggunakan teknik *windowing* untuk menjadi input algoritma XGBoost. *Windowing* dilakukan dengan menambahkan fitur *lag* pada data dengan mengacu pada variabel yang akan diramal, yaitu kejadian demam berdarah.

Pada penelitian kali ini, terdapat 12 skenario *windowing* yang dilakukan dan ditentukan sesuai dengan jumlah *lag* yang akan ditambahkan yaitu mulai penambahan hanya 1 fitur *lag* hingga penambahan 12 fitur *lag* seperti yang tercantum pada Tabel 4.2. Di samping itu, percobaan terhadap skenario dengan *input* peramalan tanpa tambahan fitur *lag* juga akan dilakukan. Percobaan skenario *windowing* ini dapat membantu untuk menemukan jumlah *lag* yang paling optimal pada penelitian yang dilakukan sehingga menghasilkan model dengan nilai kesalahan terkecil.

Tabel 4.2 Skenario Windowing

Skenario	Deskripsi
Lag-1	Penambahan 1 fitur <i>lag</i> kejadian demam berdarah
Lag-2	Penambahan 2 fitur <i>lag</i> kejadian demam berdarah
Lag-3	Penambahan 3 fitur <i>lag</i> kejadian demam berdarah
Lag-4	Penambahan 4 fitur <i>lag</i> kejadian demam berdarah
Lag-5	Penambahan 5 fitur <i>lag</i> kejadian demam berdarah
Lag-6	Penambahan 6 fitur <i>lag</i> kejadian demam berdarah
Lag-7	Penambahan 7 fitur <i>lag</i> kejadian demam berdarah

Skenario	Deskripsi
Lag-8	Penambahan 8 fitur <i>lag</i> kejadian demam berdarah
Lag-9	Penambahan 9 fitur <i>lag</i> kejadian demam berdarah
Lag-10	Penambahan 10 fitur <i>lag</i> kejadian demam berdarah
Lag-11	Penambahan 11 fitur <i>lag</i> kejadian demam berdarah
Lag-12	Penambahan 12 fitur <i>lag</i> kejadian demam berdarah

g. Membagi Data Pelatihan dan Pengujian

Setiap kecamatan terpilih akan dilakukan pembagian data pelatihan dan pengujian terlebih dahulu sebelum membentuk suatu model. Data pelatihan akan digunakan untuk melatih atau membentuk model XGBoost, sedangkan data pengujian akan digunakan untuk menguji seberapa baik model yang telah dibentuk. Dalam penelitian kali ini, terdapat empat skenario pembagian data yang akan dilakukan untuk memastikan model yang dihasilkan bersifat *robust* untuk pembagian data yang berbeda-beda seperti pada Tabel 4.3.

Tabel 4.3 Skenario Pembagian Data

No	Skenario	Proporsi Data	
		Data Pelatihan (<i>Train Set</i>)	Data Pengujian (<i>Test Set</i>)
1	80-20	80%	20%
2	70-30	70%	30%
3	60-40	60%	40%
4	50-50	50%	50%

4.2. Tahap Peramalan

Tahap peramalan merupakan tahap yang berfokus melakukan peramalan dengan model XGBoost yang paling optimal. Pada tahap ini, ada empat proses yang perlu dilakukan yaitu penentuan parameter pada Model XGBoost, pembentukan model, dan pengujian model.

4.2.1. Penentuan Parameter pada Model XGBoost

Dalam membangun model XGBoost, diperlukan inisiasi parameter yang akan dijadikan sebagai nilai *input* pembentuk model. Parameter yang akan digunakan pada penelitian kali ini untuk diuji coba nilainya adalah tingkat pembelajaran, rasio sample dan fitur yang digunakan, fungsi regularisasi, kedalaman maksimum dan jumlah pohon keputusan yang terbentuk seperti yang tertera pada Tabel 4.4.

Tabel 4.4 Parameter pada Arsitektur XGBoost

Parameter	Nilai
learning_rate	0.1, 0.3, 0.5
subsample	0.6, 0.8, 1
colsample_bytree	0.5, 0.75, 1
min_child_weight	1, 2, 4
max_depth	2, 4, 8
n_estimators	50, 100
gamma	0, 0.01
reg_alpha	0, 0.01

a. Tingkat Pembelajaran (*learning_rate*)

Parameter tingkat pembelajaran pada model XGBoost dapat didefinisikan sebagai tingkat penyusutan ukuran langkah yang digunakan dalam pembaruan model untuk mencegah adanya *overfitting*. Perhitungan bobot fitur baru akan dilakukan setiap langkah *boosting* yang dilakukan dan parameter ini akan menyusutkan bobot fitur tersebut untuk membuat proses *boosting* lebih konservatif [26]. Tingkat pembelajaran yang dilakukan uji coba dalam tugas akhir ini yaitu 0.1, 0.3, dan 0.5 seperti yang tercantum pada Tabel 4.4.

b. Rasio Sampel (*subsample*)

Parameter *subsample* digunakan untuk menentukan rasio sampel baris yang digunakan saat membangun pohon keputusan setiap iterasinya. Parameter ini juga berperan untuk mencegah adanya *overfitting* pada model XGBoost yang terbentuk karena *subsampling* akan dilakukan secara acak setiap iterasi sesuai

dengan besaran parameter yang ditentukan [26]. Rasio sampel yang dilakukan uji coba dalam tugas akhir ini ditunjukkan pada Tabel 4.4.

c. Rasio Fitur (*colsample_bytree*)

Fitur *colsample_bytree* akan mengatur seberapa banyak variabel yang digunakan untuk membangun sebuah pohon keputusan pada tiap iterasinya. Penentuan jumlah tersebut tidak didasarkan pada kuantitas yang pasti, melainkan menggunakan rasio dari jumlah fitur secara keseluruhan. Pada penelitian ini, rasio yang digunakan adalah 0.5, 0.75 dan 1.

d. Fungsi Reguralisasi (*gamma*, *reg_alpha*, dan *min_child_weight*)

Dalam penerapan algoritma XGBoost, terdapat beberapa fungsi reguralisasi yaitu *gamma*, *alpha*, dan *min_child_weight*. Nilai *gamma* akan menentukan minimum *loss* yang dibutuhkan untuk membuat partisi *node* pada tree. Nilai *alpha* berperan dalam menentukan pembobotan regularisasi L1 dalam model XGBoost. Sedangkan *min_child_weight* akan berperan untuk menentukan jumlah minimum berat (hessian) yang diperlukan pada *child node* [26]. Ketiga parameter tersebut dilakukan uji coba dalam tugas akhir seperti yang tertera pada Tabel 4.4.

e. Kedalaman maksimum (*max_depth*)

Parameter ini berperan untuk mengatur kedalaman pohon keputusan yang akan dibentuk dengan menggunakan algoritma XGBoost. Nilai *max_depth* yang semakin tinggi akan meningkatkan kompleksitas model dan akan cenderung untuk menghasilkan model yang *overfit*. Selain itu, nilai *max_depth* yang terlalu besar juga akan menghasilkan pohon keputusan yang dalam sehingga dapat secara agresif mengonsumsi memori saat proses pelatihan berlangsung [26]. Maka dari itu, nilai *max_depth* pada penelitian kali ini dibatasi hanya sampai 8 sesuai yang tertera pada Tabel 4.4.

f. Jumlah estimator (*n_estimators*)

Jumlah estimator pada XGBoost merupakan jumlah dari pohon keputusan yang terbentuk dengan menerapkan mekanisme *gradient boosting* pada tiap proses pembentukannya. Parameter ini dapat juga diartikan sebagai jumlah putaran *boosting* yang dilakukan. Jumlah estimator yang semakin tinggi akan berpengaruh kepada kecepatan komputasi yang akan semakin lama pula [27]. Pada penelitian kali ini, percobaan dibatasi pada dua nilai $n_estimator$ yaitu 50 dan 100 seperti yang tertera pada Tabel 4.4.

4.2.2. Pembentukan Model Peramalan XGBoost

Pembentukan model yang dilakukan terdiri dari beberapa skenario yaitu kombinasi dari skenario pemilihan fitur (Tabel 4.1), *windowing* (Tabel 4.2), dan pembagian data (Tabel 4.3). Dari kombinasi tersebut, pada tiap dataran model dibangun dari satu kecamatan pembentuk model dengan melibatkan 192 skenario model. Sehingga secara keseluruhan skenario model yang dilakukan yakni sebanyak 576, dimana model dibangun dari tiga kecamatan pembentuk model yang telah dipilih untuk ketiga dataran. Pengolahan data dibagi menjadi tiga dataran pada tugas akhir ini mengacu penelitian sebelumnya, dimana hal tersebut dilakukan dengan mempertimbangkan perbedaan karakteristik data iklim di tiap dataran dan jumlah kejadian demam berdarah yang dapat dipengaruhi oleh data iklim tersebut [28] [29] [8]. Sejalan dengan hal itu, diketahui pula bahwa terdapat perbedaan nilai rata-rata angka kejadian DBD di Kabupaten Malang pada tahun 2010-2018 untuk tiap dataran, dimana di dataran rendah memiliki rata-rata kasus yang paling tinggi (4,5 kasus), diikuti dataran sedang (3,8 kasus) dan dataran tinggi memiliki rata-rata kasus paling kecil (1,9 kasus).

Selanjutnya untuk masing-masing skenario model diterapkan mekanisme *Grid Search Cross Validation* untuk menemukan parameter XGBoost yang menghasilkan model dengan *error* paling kecil. Inisiasi parameter yang digunakan ditampilkan pada Tabel 4.4.

4.2.3. Menguji Model Peramalan XGBoost

Pada penelitian ini pengukuran yang akan digunakan untuk menguji model peramalan XGBoost adalah RMSE. Model-model yang telah dibangun akan diuji pada data pengujian kecamatan pembangun model (*test-set*) lalu diuji pada proporsi data yang berbeda dan kecamatan lain pada dataran yang sama sesuai pada tabel 4.5. Dari hasil semua skenario yang dilakukan, dapat diketahui model yang menghasilkan nilai RMSE terkecil. Kemudian tiga model yang memiliki RMSE paling rendah akan terpilih untuk digunakan pada tahap klasifikasi.

Tabel 4.5 Skenario Pengujian Model

Dataran	Skenario Pembangun Model	Skenario Pengujian	
		Uji Proporsi Data	Uji Kecamatan Lain
Rendah	80%-20%	70%-30%, 60%-40%, 50%-50%	Turen, Bululawang, Gondanglegi, Donomulyo
	70%-30%	80%-20%, 60%-40%, 50%-50%	Turen, Bululawang, Gondanglegi, Donomulyo
	60%-40%,	80%-20%, 70%-30%, 50%-50%	Turen, Bululawang, Gondanglegi, Donomulyo
	50%-50%	70%-30%, 60%-40%, 50%-50%	Turen, Bululawang, Gondanglegi, Donomulyo
Sedang	80%-20%	70%-30%, 60%-40%, 50%-50%	Dampit, Singosari, Tumpang, Lawang, Sumbermanjing, Wajak, Karangploso
	70%-30%	80%-20%, 60%-40%, 50%-50%	Dampit, Singosari, Tumpang, Lawang, Sumbermanjing, Wajak, Karangploso
	60%-40%,	80%-20%, 70%-30%, 50%-50%	Dampit, Singosari, Tumpang, Lawang, Sumbermanjing, Wajak, Karangploso
	50%-50%	70%-30%, 60%-40%, 50%-50%	Dampit, Singosari, Tumpang, Lawang, Sumbermanjing, Wajak, Karangploso
Tinggi	80%-20%	70%-30%,	Poncokusumo, Jabung

Dataran	Skenario Pembangunan Model	Skenario Pengujian	
		Uji Proporsi Data	Uji Kecamatan Lain
		60%-40%, 50%-50%	
	70%-30%	80%-20%, 60%-40%, 50%-50%	Poncokusumo, Jabung
	60%-40%,	80%-20%, 70%-30%, 50%-50%	Poncokusumo, Jabung
	50%-50%	70%-30%, 60%-40%, 50%-50%	Poncokusumo, Jabung

4.3. Tahap Klasifikasi

Tahap klasifikasi merupakan tahap dimana data hasil peramalan akan dikategorikan menjadi dua kelas, yaitu KLB dan non-KLB dengan menggunakan metode Klasifikasi Berbasis Aturan. Pada tahap ini, ada empat proses yang perlu dilewati yaitu penambahan fitur penunjang klasifikasi, penentuan status KLB dan non-KLB untuk data hasil peramalan, dan pengukuran kinerja klasifikasi.

4.3.1. Penambahan Fitur Penunjang Klasifikasi

Sebelum data dilakukan proses klasifikasi berbasis aturan, perlu dilakukan penambahan fitur penunjang baru yaitu rata-rata kejadian demam berdarah pada tahun sebelumnya. Fitur ini akan membantu proses klasifikasi. Hal ini mengacu pada aturan kedua pada peraturan pemerintah seperti yang disebutkan pada subbab 2.2.1. Sedangkan untuk aturan pertama, tidak diperlukan penambahan fitur apapun. Hal ini dikarenakan pada tahun sebelumnya (2009) semua kecamatan yang menjadi objek penelitian telah memiliki setidaknya satu orang yang menjadi pasien demam berdarah.

4.3.2. Pelabelan KLB pada Data Hasil Peramalan

Penentuan status KLB dan non-KLB merupakan inti dari tahap klasifikasi. Tiga model terbaik yang telah terpilih pada tahap peramalan kemudian diklasifikasikan ke dalam dua kategori yaitu positif KLB dengan label pada data sebagai 'KLB' dan negatif KLB atau yang diberi label 'non-KLB'. Sehingga, keluaran pada tahap ini adalah data hasil peramalan model terbaik yang sudah berlabel KLB dan non-KLB pada kurun waktu Januari 2010 hingga Desember 2018.

4.3.3. Pengukuran Performa Klasifikasi

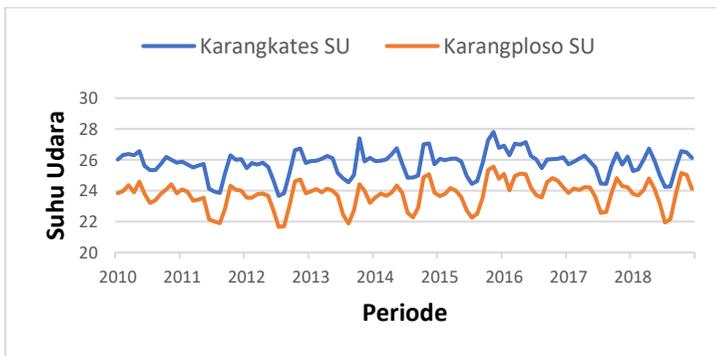
Pengukuran kinerja klasifikasi biner yang akan digunakan pada penelitian kali ini yaitu tiga macam pengukuran, yaitu nilai akurasi, spesifisitas, dan sensitivitas. Ketiga pengukuran tersebut memiliki fungsi yang berbeda-beda. Akurasi akan menggambarkan seberapa baik klasifikasi yang telah dilakukan secara keseluruhan. *Spesifisitas* menggambarkan seberapa baik klasifikasi dilakukan untuk kelas positif (KLB), sedangkan sensitivitas untuk kelas negatif (non-KLB). Ketiga pengukuran tersebut akan dihitung berdasarkan confusion matrix yang dibuat dari data hasil prediksi dan data aktual. Pengukuran kinerja klasifikasi tidak hanya dilakukan untuk kecamatan pembangun model, melainkan semua kecamatan lain pada kategori dataran yang sama. Sehingga, dapat dipilih model mana yang paling terbaik untuk tiap kategori dataran.

4.4. Peramalan Data 24 Periode Selanjutnya

Model yang terbaik untuk tiap dataran akan digunakan untuk meramalkan dua tahun (24 bulan) kedepan yaitu pada kurun waktu Januari 2019 hingga Desember 2020. Periode tersebut dipilih untuk dilakukan peramalan dengan mempertimbangkan tahun dibuatnya penelitian ini. Tahap peramalan 24 periode selanjutnya dimulai dengan meramalkan data variabel independen periode Januari 2019 hingga Desember 2020, lalu meramalkan angka kejadian demam berdarah untuk masing-masing kecamatan, kemudian dilanjutkan dengan memberi label KLB atau non-KLB pada data tersebut. Sehingga, akan

didapatkan data berlabel KLB maupun non-KLB untuk periode Januari 2019 hingga Desember 2020 pada 16 kecamatan di Kabupaten Malang.

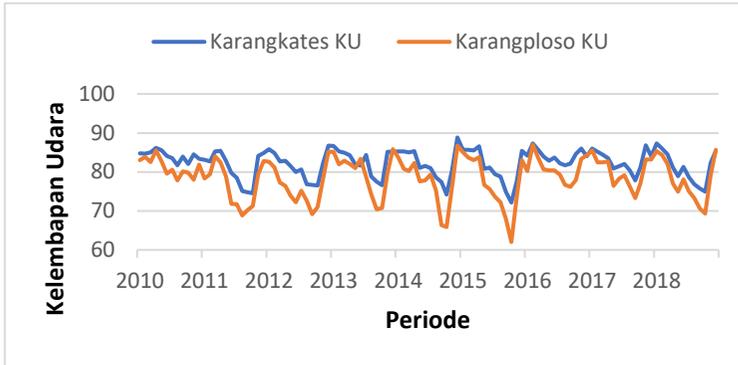
Pada proses peramalan variabel independen, metode yang digunakan untuk berbeda untuk masing-masing variabel. Pemilihan metode tersebut didasarkan pada pola data tiap variabel. Pada Gambar 4.1 dapat dilihat bahwa variabel suhu udara memiliki pola data yang bersifat siklus, yakni berulang dalam dalam jangka waktu yang panjang. Pola data yang serupa juga dapat dilihat pada variabel kelembapan udara yang tertera pada Gambar 4.2. Maka dari itu, kedua variabel tersebut diramalkan dengan menggunakan metode dekomposisi multiplikatif.



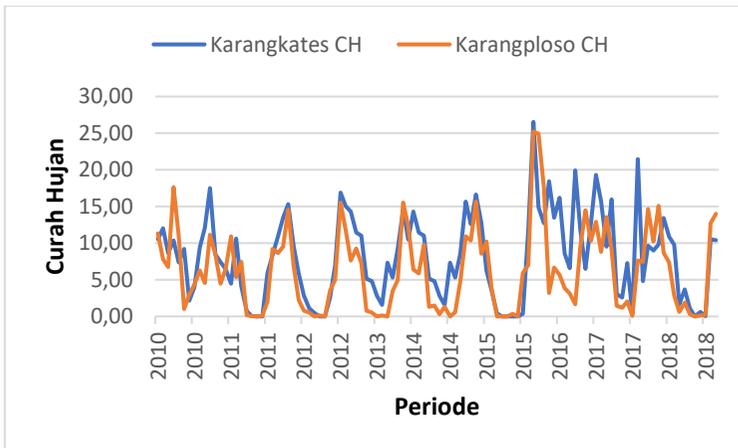
Gambar 4.1 Pola Data Suhu Udara

Pada variabel curah hujan, pola data yang terbentuk adalah pola non-stasioner seperti pada Gambar 4.3. Maka dari itu, metode peramalan yang digunakan untuk variabel curah hujan adalah metode ARIMA (Autoregressive Integrated Moving Average) karena ketepatannya baik untuk data yang bersifat *time series* dan non stasioner. Namun, metode ARIMA juga dapat diterapkan untuk data yang sifatnya stasioner seperti variabel kecepatan angin. Untuk meramalkan data yang sifatnya stasioner dengan metode ARIMA, perlu dilakukan transformasi data terlebih dahulu. Sehingga variabel curah hujan

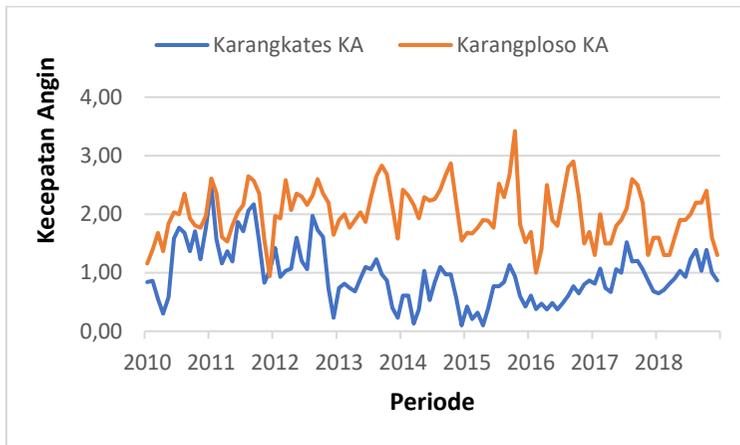
menggunakan metode ARIMA dengan melewati serangkaian transformasi data (*box-cox* dan *differencing*) sebelum dilakukan peramalan.



Gambar 4.2 Pola Data Kelembapan Udara

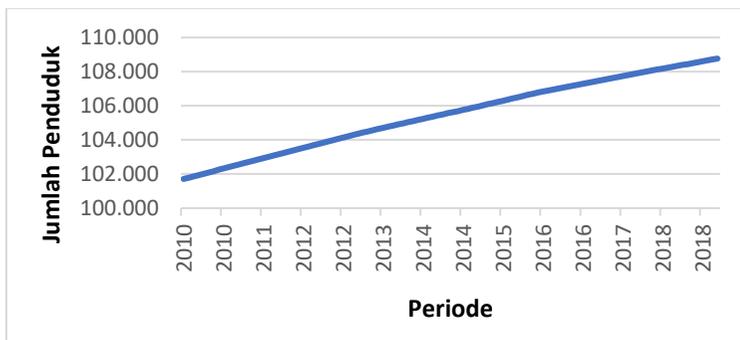


Gambar 4.3 Pola Data Curah Hujan



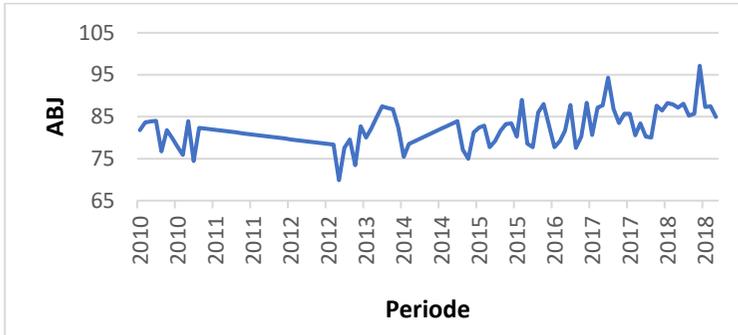
Gambar 4.4 Pola Data Kecepatan Angin

Pada variabel jumlah penduduk, data hanya memiliki pola tren yang menaik seperti pada Gambar 4.5. Maka dari itu, metode peramalan yang digunakan pada data jumlah penduduk adalah Double Exponential Smoothing.



Gambar 4.5 Pola Data Jumlah Penduduk Kec. Kepanjen

Pola data variabel angka bebas jentik dapat dilihat pada Gambar 4.6. Metode peramalan yang digunakan pada data Angka Bebas Jentik adalah Exponential Smoothing. Keseluruhan hasil proses peramalan variabel independen yang dilakukan terlampir pada Lampiran A.



Gambar 4.6 Pola Data Angka Bebas Jentik

4.5. Perbandingan dengan Metode/Algoritma Lain

Hasil peramalan yang didapatkan perlu dibandingkan dengan hasil dari metode atau algoritma lainnya untuk mengetahui apakah model yang dibangun sudah cukup baik. Terdapat dua jenis perbandingan yang dilakukan, yaitu perbandingan performa peramalan dan perbandingan performa klasifikasi. Perbandingan performa peramalan akan dilakukan dengan membandingkan nilai RMSE yang dihasilkan algoritma XGBoost dalam meramalkan kejadian demam berdarah dengan algoritma lain seperti KNN Regression, Random Forest, Linear Regression, Decision Tree, dan RBFNN. Sedangkan untuk perbandingan performa klasifikasi dilakukan dengan membandingkan nilai akurasi, spesifisitas, dan sensitivitas yang dihasilkan dengan metode klasifikasi yang dilakukan secara langsung pada data yang telah berlabel KLB (tanpa proses peramalan kejadian demam berdarah terlebih dahulu).

BAB V IMPLEMENTASI

Pada bab ini, dijelaskan mengenai implemenatasi metode XGBoost dan Klasifikasi Berbasis Aturan ke dalam bahasa Python dengan menggunakan perangkat lunak Spyder dan Jupyter Notebook, serta dengan bantuan pengolah data lain yaitu Microsoft Excel. Tahapan yang akan dilakukan mengacu pada Bab Metodologi yang telah disesuaikan pada rancangan pada Bab Perancangan.

5.1. Persiapan Implementasi

Beberapa spesifikasi perangkat keras maupun lunak diperlukan untuk menunjang berjalannya proses penelitian. Pada penelitian kali ini, perangkat keras dan lunak yang digunakan dapat dilihat pada Tabel 5.1.

Tabel 5.1 Spesifikasi Perangkat Keras dan Lunak

Perangkat Keras dan Lunak	Spesifikasi
Laptop	Asus A55V series
<i>Processor</i>	Intel(R) Core(TM) i5-3230M CPU @ 2.60GHz
<i>Memory</i>	8,00 GB RAM
Sistem Operasi	Windows 10 Pro 64-bit

Sedangkan untuk teknologi yang digunakan dalam proses penelitian kali ini seperti *code editor*, bahasa pemrograman, dan *library* disajikan dalam Tabel 5.2.

5.2. Persiapan Data

Tahap persiapan data terdiri dari tahap pengumpulan dan pra-proses data. Langkah-langkah yang dilakukan pada tahap ini adalah persiapan data iklim yang dilakukan dengan bantuan Microsoft Excel, serta penanganan *missing value*, pemilihan kecamatan pembentuk model, *windowing*, uji korelasi, dan

pembagian data menjadi data pelatihan dan pengujian menggunakan bantuan Jupyter Notebook dan Spyder.

Tabel 5.2 Lingkungan Uji Coba

Teknologi	Spesifikasi
Bahasa Pemrograman	Python
<i>Code Editor (IDE)</i>	Spyder, Jupyter Notebook
<i>Library</i>	<ul style="list-style-type: none"> • pandas • numpy • xgboost • sklearn.preprocessing • sklearn.metrics • sklearn.model_selection • matplotlib

5.2.1. Mengubah Data Harian Variabel Iklim

Data iklim yaitu variabel suhu, curah hujan, kecepatan angin, dan kelembapan udara yang semula memiliki rentang waktu harian perlu diubah menjadi rentang waktu bulanan seperti yang telah disebutkan pada subbab 4.1.2. Pengubahan data harian tersebut dapat dilakukan dengan menggunakan fungsi AVERAGE pada Microsoft Excel.

5.2.2. Penanganan Missing Value

Data variabel independen yang telah dikumpulkan masih terdapat beberapa nilai *null*, khususnya pada variabel Angka Bebas Jentik dan variabel iklim. Maka dari itu, diperlukan penanganan *missing value* pada variabel-variabel tersebut. Pada penelitian kali ini, data *null* pada variabel iklim diisi dengan nilai rata-rata musiman pada bulan tersebut menggunakan bantuan Microsoft Excel dengan fungsi AVERAGE. Sedangkan pengisian nilai *null* pada variabel Angka Bebas Jentik dibagi menjadi dua jenis, yaitu interpolasi data dengan fungsi `interpolate()` serta regresi linier. Interpolasi pada Python dapat dilakukan dengan menggunakan *syntax* yang tertera pada Kode 5.1.

```
#interpolasi
> raw_df = raw_df.interpolate(limit_direction='both')
```

Kode 5.1 Interpolasi pada Variabel Angka Bebas Jentik

Keterangan dari kode 5.1 tertulis pada Tabel 5.3.

Tabel 5.3 Penjelasan dari Kode 5.1

Atribut	Penjelasan
raw_df	Data keseluruhan yang mencakup variabel independen dan dependen
interpolate()	Fungsi interpolasi untuk mengisi nilai <i>null</i> pada data
limit_direction = 'both'	Atribut pada fungsi <i>interpolate</i> yang mengatur pembatasan arah interpolasi. Jika nilai atribut 'both', maka interpolasi dapat dilakukan ke depan maupun ke belakang.

Khusus di kecamatan Wajak, pengisian *missing value* menggunakan regresi linier perlu dilakukan karena data bernilai *null* terdapat pada tiga tahun pertama. Hal ini dapat dilakukan dengan menggunakan kode yang tertulis pada Kode 5.2.

Regresi linier dapat dilakukan dengan memisah terlebih dahulu data X sebagai variabel independen dan data Y sebagai variabel dependen. Dalam hal ini semua variabel lain yaitu suhu, curah hujan, kecepatan angin, kelembapan udara, jumlah penduduk, dan angka kejadian demam berdarah akan berperan sebagai variabel independen untuk menentukan nilai variabel dependen yakni Angka Bebas Jentik (ABJ).

```
#Regresi linier
X_wajak_train = raw_wajak.iloc[50:,raw_wajak.columns
!= 'ABJ']
X_wajak_predict = raw_wajak.iloc[:50,raw_wajak.colum
ns != 'ABJ']
y_wajak_train = raw_wajak.iloc[50:,raw_wajak.columns
== 'ABJ']

reg = LinearRegression().fit(X_wajak_train, y_wajak)
predict = reg.predict(X_wajak_predict)
```

Kode 5.2 Ekstrapolasi dan Regresi pada Variabel Angka Bebas Jentik (ABJ)

Keterangan atribut-atribut dari kode 5.2 tertulis pada Tabel 5.3.

Tabel 5.4 Penjelasan dari Kode 5.2

Atribut	Penjelasan
fit()	Fungsi untuk melakukan pelatihan pada pasangan data X dan Y yang sudah ditentukan
predict()	Fungsi untuk melakukan prediksi pada data X agar menghasilkan nilai Y

5.2.3. Pemilihan Kecamatan Pembentuk Model

Pemilihan kecamatan pembentuk model didasarkan pada nilai korelasi tertinggi antarkecamatan di tiap datarannya seperti yang telah dijelaskna pada subbab 4.1.2. Perhitungan korelasi pada penelitian kali ini menggunakan bantuan aplikasi Jupyter Notebook dengan menggunakan kode seperti yang tertera pada Kode 5.3.

```
#korelasi antarkecamatan
corr_dataran = dataran_df.corr(method='pearson')
```

Kode 5.3 Perhitungan Korelasi Antarkecamatan

Kode 5.3 akan menghasilkan matriks korelasi yang berisikan angka 0 hingga 1. Pemilihan Kecamatan didasarkan pada kecamatan yang memiliki nilai korelasi mendekati 1 dengan kecamatan-kecamatan lainnya. Keterangan atribut-atribut dari kode 5.2 tertulis pada Tabel 5.5.

Tabel 5.5 Penjelasan dari Kode 5.3

Atribut	Penjelasan
dataran_df	Data yang berisikan variabel kejadian demam bedarah pada masing-masing kecamatan di dataran yang sama
corr()	Fungsi untuk perhitungan nilai korelasi
method = 'pearson'	Atribut pada fungsi corr() untuk pendefinisian metode yang diguankan, dalam hal ini korelasi pearson

5.2.4. Normalisasi Data

Normalisasi data yang digunakan pada tugas akhir ini adalah normalisasi min-max dengan rentang nilai hasil normalisasi 0 hingga 1. Normalisasi dapat dilakukan dengan menggunakan kode pada Kode 5.4.

```
#normalisasi
y_raw = raw_df.iloc[:,raw_df.columns == 'KDB']

#normalisasi x
scalerx = MinMaxScaler(feature_range=(0,1))
X_scaled = pd.DataFrame(scalerx.fit_transform(raw_df
), columns = ['SU', 'KU', 'CH','KA', 'JP', 'ABJ', 'K
DB'])

#normalisasi y
scalery = MinMaxScaler(feature_range=(0,1))
Y_scaled = scalery.fit_transform(y_raw)
```

Kode 5.4 Normalisasi Min Max

Seperti yang tertera pada Kode 5.4, pemisahan *scaler* yang digunakan pada data variabel independen (suhu, kelembapan udara, curah hujan, kecepatan angin, jumlah penduduk, dan ABJ) dengan variabel dependen (angka kejadian demam berdarah) dimaksudkan agar memudahkan proses normalisasi-denormalisasi. Keterangan dari kode 5.4 tertulis pada Tabel 5.6.

Tabel 5.6 Penjelasan dari Kode 5.4

Atribut	Penjelasan
MinMaxScaler()	Fungsi yang menginisiasi <i>scaler</i> untuk normalisasi data dengan metode min-max
feature_range = (0,1)	Atribut pada fungsi MinMaxScaler() untuk mendefinisikan rentang nilai hasil normalisasi
fit_transform()	Fungsi untuk melakukan <i>fitting</i> dan mentransformasi nilai sesuai dengan <i>scaler</i> yang digunakan

5.2.5. Pemilihan Fitur dengan Uji Korelasi

Seperti yang telah dijelaskan pada subbab 4.1.2, uji korelasi dalam tugas akhir ini menggunakan perhitungan korelasi *pearson*. Proses yang dilakukan serupa dengan tahap 5.2.4 karena menggunakan fungsi kode yang sama yaitu *corr* seperti dituliskan pada Kode 5.6.

```
#uji korelasi variabel independen
corr_vi = raw_df.corr(method='pearson')
```

Kode 5.5 Uji Korelasi Variabel Independen

Kode 5.6 akan menghasilkan matriks korelasi yang berisi angka 0 hingga 1. Dari matriks tersebut, dapat dilihat variabel-variabel yang berpengaruh terhadap angka kejadian demam berdarah. Kemudian variabel-variabel tersebut akan dipilih sebagai masukan dalam membangun model. Mengacu pada Tabel 4.2, variabel yang kurang berpengaruh akan dihilangkan dengan menggunakan method ‘drop()’ untuk skenario yang hanya melibatkan 1, 2, dan 3 variabel independen utama yang paling berpengaruh saja (skenario TOP3-VI, TOP2-VI, dan TOP1-VI).

5.2.6. Windowing

Windowing dilakukan agar data deret waktu berubah menjadi *cross-sectional* sehingga data tersebut dapat diproses lebih lanjut dengan algoritma *supervised learning* yaitu XGBoost seperti yang telah dijelaskan pada 4.1.2. Pada penelitian ini, *windowing* dilakukan dengan 12 macam skenario yang dapat tertuang menjadi baris kode pada Kode 5.4.

```
#Windowing
for x in range(case) :
    X_scaled['kdblag-' + str(x+1)] = X_scaled['K
    DB'].shift(x+1)
```

Kode 5.6 Proses Windowing

Proses ini mengacu pada skenario *windowing* yang telah tercantum dalam Tabel 4.1, sehingga jumlah iterasi penambahan fitur *lag* disesuaikan dengan skenario tersebut.

Seperti yang tertera pada Kode 5.5, jumlah iterasi akan ditentukan dengan variabel 'case'. Sebagai contoh, skenario Lag-1 akan memiliki variabel 'case' bernilai 1, skenario Lag-2 akan memiliki variabel 'case' bernilai 2, dst. Keterangan atribut yang ada pada kode 5.5 tertulis pada Tabel 5.7.

Tabel 5.7 Penjelasan dari Kode 5.5

Atribut	Penjelasan
shift(x+1)	Fungsi untuk memindahkan indeks dengan jumlah periode yang telah ditentukan. Pada penelitian ini fitur akan bertambah satu untuk tiap iterasinya.

5.2.7. Pembagian Data Pelatihan dan Data Pengujian

Data pelatihan dan data pengujian yang digunakan dalam tugas akhir ini menggunakan perbandingan 50:50, 60:40, 70:30, dan 80:20 yang mengacu pada skenario dalam Tabel 4.3. Kode program yang disusun untuk melakukan pembagian data pelatihan dan pengujian ditunjukkan pada Kode 5.7.

Seperti yang tercantum pada Kode 5.7, sebelum pembagian data dilakukan maka perlu pendefinisian variabel masukan dan variabel keluaran yang digunakan untuk membangun model terlebih dahulu. Variabel masukan merupakan variabel independen yang digunakan dalam membangun model dan variabel keluaran didefinisikan sebagai variabel yang akan diramal yaitu angka kejadian demam berdarah. Pada penelitian ini, variabel masukan yang digunakan akan mengacu pada skenario pemilihan fitur yang sedang dijalankan sesuai dengan Tabel 4.2.

```
#pembagian train-set dan test-set
X_scaled = X_scaled.iloc[:,X_scaled.columns != 'KDB']
y_scaled = y_scaled[case[:,]]

X_train, X_test, y_train, y_test = train_test_split(X_scaled, y_scaled, test_size = ukuran_tes, shuffle=False)
```

Kode 5.7 Pembagian Data Pelatihan dan Data Pengujian

Ukuran proporsi yang berbeda tiap skenario pembagian data didefinisikan dengan variabel ‘ukuran_tes’ seperti yang tertera pada Kode 5.7. Variabel ini berisikan nilai sebesar 0.2 untuk pembagian data dengan proporsi 80:20, 0.3 untuk proporsi 70:30, 0.4 untuk proporsi 60:40 dan 0.5 untuk pembagian dengan proporsi 50:50. Keterangan atribut yang ada pada kode 5.7 tertulis pada Tabel 5.8.

Tabel 5.8 Penjelasan dari Kode 5.7

Atribut	Penjelasan
X_scaled, y_scaled	Data masukan (X) dan keluaran (Y) yang telah dinormalisasi
train_test_split()	Fungsi untuk melakukan pembagian data pelatihan dan data pengujian
shuffle=False	Atribut dalam fungsi train_test_split() yang digunakan untuk mengontrol data dipisah dengan sesuai urutan

5.3. Tahap Peramalan

Tahap ini berfokus untuk melakukan peramalan angka kejadian demam berdarah. Langkah-langkah yang dilakukan adalah penentuan parameter, pembentukan model, dan pengujian.

5.3.7. Penentuan Parameter pada Model XGBoost

Parameter yang terdapat pada model XGBoost akan di-*tuning* pada tugas akhir ini berjumlah delapan, yaitu inisiasi tingkat pembelajaran, rasio sampel serta fitur yang digunakan, tiga fungsi regularisasi, kedalaman maksimum dan jumlah pohon keputusan yang terbentuk. Inisiasi parameter dibangun dengan kode yang tercantum pada Kode 5.8. Penjelasan untuk masing-masing parameter telah dituliskan pada subbab 4.2.1.

```

#penentuan parameter
gbm_param_grid = {
    'learning_rate' : [0.1, 0.3, 0.5],
    'subsample' : [0.6, 0.8, 1],
    'colsample_bytree' : [0.5, 0.75, 1],
    'min_child_weight' : [1, 2, 4],
    'n_estimators' : [50, 100],
    'max_depth' : [2, 4, 8],
    'gamma' : [0, 0.01],
    'reg_alpha' : [0, 0.01]
}

```

Kode 5.8. Penentuan Parameter

5.3.8. Pembentukan Model XGBoost

Seperti yang telah dijelaskan pada subbab 4.2.2, model peramalan XGBoost pada tugas akhir ini dibangun dengan menggunakan metode *grid-search* untuk *tuning* parameter. Hal ini dilakukan dengan tujuan untuk menghasilkan model yang dengan nilai galat yang kecil. Sedangkan untuk jenis *cross validation* yang digunakan dalam proses *grid-search* penelitian kali ini adalah *time series cross validation* karena data yang digunakan bersifat *time series*. Kode yang digunakan tertulis dalam Kode 5.9.

Pembuatan model diawali dengan inisiasi model *default* kemudian ditambah dengan algoritma *grid-search* untuk melakukan *tuning* parameter yang telah ditentukan pada tahap 5.3.7. Hal ini dilakukan dengan menggunakan fungsi 'GridSearchCV' pada python. Model lalu dibangun dengan memanggil fungsi 'fit' dengan pasangan data pelatihan (X_{train} dan y_{train}). Peramalan untuk data *test-set* selanjutnya dilakukan dengan menggunakan fungsi 'predict'. Karena data yang digunakan untuk membangun model menggunakan data normalisasi, maka perlu adanya mekanisme denormalisasi data setelah proses peramalan selesai dilakukan. Denormalisasi ini sangat penting dilakukan agar peramalan dapat menghasilkan angka kejadian demam berdarah dalam skala aslinya. Pada bahasa Python, denormalisasi data dapat dilakukan dengan

menggunakan fungsi *'inverse_transform'*. Data peramalan yang telah didenormalisasi kemudian diubah menjadi bilangan bulat agar menghasilkan nilai representatif untuk jumlah orang yang menderita demam berdarah.

```
#pembuatan model
gbm = xgb.XGBRegressor()

#grid-search
tscv = TimeSeriesSplit(n_splits=5)
grid_mse = GridSearchCV(estimator=gbm,
param_grid=gbm_param_grid, n_jobs=15,
scoring='neg_root_mean_squared_error', cv=tscv,
verbose=1)

#fit & predict
grid_mse.fit(X_train, y_train)
pred_test =
grid_mse.best_estimator_.predict(X_test).reshape(-1, 1)

#denormalisasi
denorm_pred_test = scaler.inverse_transform(pred_test)
denorm_pred_test_round =
np.round(scaler.inverse_transform(pred_test))
```

Kode 5.9 Pembentukan Model XGBoost

Penjelasan atribut dari Kode 5.9 tertulis pada Tabel 5.9.

Tabel 5.9 Penjelasan dari Kode 5.9

Atribut	Penjelasan
XGBRegressor	Fungsi yang digunakan untuk menginisiasi model regresi XGBoost yang dipanggil dari <i>library xgboost</i>
TimeSeriesSplit	Fungsi yang digunakan untuk melakukan split pada data yang sifatnya <i>time series</i>
n_splits	Atribut pada TimeSeriesSplit yang mengatur jumlah split yang dilakukan pada data
GridSearchCV	Fungsi yang digunakan untuk melakukan <i>grid search</i> cross validation pada model
estimator	Model yang akan menerapkan <i>grid search</i>
param_grid	Parameter yang akan dilakukan <i>tuning</i>

Atribut	Penjelasan
n_jobs	Jumlah CPU yang digunakan untuk melakukan perhitungan
scoring	Pengukuran yang digunakan dalam perhitungan <i>grid search</i>
cv	Menentukan strategi <i>split cross-validation</i> yang dilakukan pada data
verbose	Atribut yang mengatur informasi yang ditampilkan pada hasil peramalan
fit	Fungsi untuk melakukan pelatihan pada data
predict	Fungsi untuk melakukan peramalan
inverse_transform	Fungsi yang digunakan untuk denormalisasi

5.3.9. Pengujian Model Peramalan

Pada penelitian kali ini, pengujian dilakukan dengan tiga tahap, yaitu pengujian pada *test-set* kecamatan pembangun model, pengujian pada proporsi data yang lainnya di kecamatan pembangun model, dan pengujian pada kecamatan lain dalam kategori dataran yang sama seperti yang telah dirancang pada Tabel 4.5. Pengukuran yang digunakan saat pengujian yaitu MSE dan SMAPE dengan rumus yang telah dijabarkan sebelumnya pada subbab 2.2.10.

a. Pengujian pada *test-set*

Data hasil peramalan *test-set* perlu dilakukan pengujian agar diketahui seberapa baik model yang telah dibangun jika diterapkan pada kecamatan yang sama. Pengujian ini dapat dilakukan dengan terlebih dahulu mendefinisikan rumus SMAPE, kemudian memanggil fungsi ‘*mean_squared_error*’ dalam rumus akar sehingga menghasilkan nilai RMSE. Hal ini dapat dilakukan dengan kode yang tertulis pada Kode 5.10.

```
#pengujian pada test set
def smape(A, F):
    return (100/len(A) * np.sum(np.abs(F-
A)/(np.abs(A)+np.abs(F))))
rmse_test = np.sqrt(mean_squared_error(y_test,
denorm_pred_test))
smape_test = smape(y_test, denorm_pred_test)
```

Kode 5.10 Pengujian pada *test-set*

Penjelasan Kode 5.10 tertera pada Tabel 5.10.

Tabel 5.10 Penjelasan dari Kode 5.10

Atribut	Penjelasan
np.abs	Fungsi pada library <i>numpy</i> untuk melakukan perhitungan nilai mutlak pada data
np.len	Fungsi pada library <i>numpy</i> untuk melakukan perhitungan jumlah item dalam suatu objek
np.sum	Fungsi pada library <i>numpy</i> untuk melakukan penjumlahan pada data
np.sqrt	Fungsi pada library <i>numpy</i> untuk melakukan perhitungan akar pada data
mean_squared_error	Fungsi untuk menghitung nilai MSE pada data

b. Pengujian pada Proporsi Data Lain

Untuk mengetahui tingkat *robustness* model yang telah dibangun, salah satu caranya yaitu melakukan pengujian pada proporsi data yang lain di kecamatan pembangun model seperti telah dijelaskan pada bab 4.2.3. Pengujian dilakukan berdasarkan skenario yang digunakan dalam membangun sebuah model. Jika model dibangun dengan pembagian data 80:20, maka model akan diuji pada pembagian data 70:30, 60:40, dan 50:50. Hal ini berlaku untuk semua skenario uji proporsi data dengan mengacu pada skenario dalam Tabel 4.5. Pengujian ini dilakukan dengan menggunakan kode pada Kode 5.11.

```

#pengujian pada proporsi data lain
validasi_proporsi = []
if ukuran_tes==0.2:
    validasi_proporsi = [0.3,0.4,0.5]
elif ukuran_tes==0.3:
    validasi_proporsi = [0.2,0.4,0.5]
elif ukuran_tes==0.4:
    validasi_proporsi = [0.2,0.3,0.5]
elif ukuran_tes==0.5:
    validasi_proporsi = [0.2,0.3,0.4]
count=0
dict_var_proporsi = {}

for i in validasi_proporsi:
    X_train_val, X_test_val, y_train_val, y_test_val
= train_test_split(X_scaled, y_raw, test_size = i,
shuffle=False)

    pred_val =
grid_mse.best_estimator_.predict(X_test_val).reshape(-1,
1)
    denorm_pred_val =
scalery.inverse_transform(pred_val)

    dict_var_proporsi[count] =
np.sqrt(mean_squared_error(y_test_val,
denorm_pred_test_val))
    count=count+1

    dict_var_proporsi[count] = smape(y_test_val,
denorm_pred_test_val)
    count=count+1

```

Kode 5.11 Uji Model Peramalan pada Proporsi Data Lain

Pengujian proposi data dilakukan dengan melakukan pembagian data pelatihan dan data pengujian sesuai dengan skenario uji, memprediksi angka kejadian demam berdarah dengan fungsi ‘predict’, melakukan denormalisasi pada data, dan mengevaluasi hasil peramalan dengan RMSE serta SMAPE. Penjelasan atribut pada kode 5.11 terdapat pada Tabel 5.11.

Tabel 5.11 Penjelasan dari Kode 5.11

Atribut	Penjelasan
validasi_proporsi	<i>Array</i> yang digunakan untuk menentukan uji proporsi data yang dilakukan
dict_var_proporsi	<i>Dictionary</i> yang digunakan untuk menyimpan hasil RMSE dan SMAPE

c. Pengujian pada Kecamatan Lain

Tahap terakhir pengujian model peramalan pada penelitian kali ini adalah melakukan uji di kecamatan lain pada kategori dataran yang sama. Hal ini dilakukan untuk mengetahui seberapa baik model yang telah dibangun dalam meramalkan data lain. Kode yang digunakan untuk melakukan uji model pada kecamatan lain tercantum pada Kode 5.12.

Seperti yang tertera pada Kode 5.12, penanganan data yang dilakukan pada uji ini hampir sama dengan penanganan data untuk proses pembentukan model, yaitu mulai dari normalisasi data hingga prediksi serta perhitungan RMSE dan SMAPE di kecamatan tertentu. Namun dalam melakukan uji model peramalan pada kecamatan lain, proses pelatihan data tidak dilakukan melainkan model yang digunakan adalah model yang telah dilatih sebelumnya menggunakan *train-set* pada proses pembentukan model. Skenario pengujian pada kecamatan lain mengacu pada Tabel 4.5.

5.4. Tahap Klasifikasi

Tahap ini berfokus untuk menggolongkan data hasil peramalan menjadi dua kategori yaitu KLB dan non-KLB. Langkah-langkah yang perlu dilakukan meliputi penambahan fitur penunjang klasifikasi, penentuan status KLB dan non-KLB pada data hasil peramalan, dan pengukuran kinerja klasifikasi yang telah dilakukan.

```

#pengujian pada kecamatan lain
y_raw_kec_df = raw_kec_df.iloc[:,raw_kec_df.columns ==
'KDB']

#normalisasi
scalerx2 = MinMaxScaler(feature_range=(0,1))
raw_kec_scaled =
pd.DataFrame(scalerx2.fit_transform(raw_kec_df), columns
= ['SU', 'KU', 'CH','KA', 'JP', 'ABJ', 'KDB'])
raw_kec_scaled = raw_kec_scaled.drop(['KU', 'KA',
'ABJ','SU'], axis=1)

scalery2 = MinMaxScaler(feature_range=(0,1))
scalery2 = scalery2.fit(y_kec_all)

#windowing
for x in range(case) :
    raw_kec_scaled['kdblag-'+ str(x+1)] =
raw_kec_scaled['KDB'].shift(x+1)
X_kec = lagged_kec.iloc[case:,:lagged_kec.columns !=
'KDB']

#prediksi & normalisasi
pred_kec =
grid_mse.best_estimator_.predict(X_kec).reshape(-1, 1)
denorm_kec = scalery2.inverse_transform(pred_kec)
denorm_kec_round = np.round(denorm_kec)

#perhitungan RMSE & smape
rmse_kec = np.sqrt(mean_squared_error(y_kec,
denorm_kec))
smape_kec = smape(y_kec, denorm_kec)

```

Kode 5.12 Uji Model Peramalan pada Kecamatan Lain

5.4.1. Penambahan Fitur Penunjang Klasifikasi

Fitur yang perlu ditambahkan untuk melakukan Klasifikasi Berbasis Aturan pada data yaitu angka rata-rata kejadian demam berdarah pada tahun sebelumnya. Penambahan fitur ini dilakukan dengan fungsi AVERAGE pada Microsoft Excel untuk merata-rata angka kejadian demam berdarah 12 periode

pada tahun sebelumnya dan memasukkan hasil tersebut ke kolom baru bernama 'avg_recentyear_KDB'.

5.4.2. Penentuan Status KLB untuk Data Hasil Peramalan

Seperti yang telah dijelaskan pada subbab 4.3.2, proses penentuan status KLB dan non-KLB ini merupakan inti dari tahap klasifikasi. Pada tahap ini data hasil peramalan dari model peramalan terbaik diberi label KLB dan non-KLB berdasarkan aturan yang telah ditentukan. Pelabelan ini dapat dilakukan dengan menggunakan kode yang tercantum pada Kode 5.13.

```
#pelabelan KLB dan non-KLB
raw_df_klb = raw_df_klb.iloc[case:,7:]
klb_df = pd.concat([for_df, raw_df_klb], axis=1,
sort=False)

def rulebased(data):
    data.loc[data['df_forecast'] >=
(2*data['avg_recentyear_KDB']), 'KLB_pred'] = 'KLB'
    data.loc[data['df_forecast'] <
(2*data['avg_recentyear_KDB']), 'KLB_pred'] = 'non-KLB'

rulebased(klb_df)
```

Kode 5.13 Penentuan Status KLB dan non-KLB

Klasifikasi Berbasis Aturan dilakukan dengan terlebih dahulu menyiapkan *dataframe* yang terdiri dari nilai KLB aktual, angka rata-rata kejadian demam berdarah tahun sebelumnya, dan angka kejadian demam berdarah hasil peramalan. Data tersebut kemudian diterapkan aturan klasifikasi yakni suatu daerah akan berstatus KLB jika angka kejadian demam berdarahnya lebih besar atau sama dengan rata-rata angka kejadian pada tahun sebelumnya. Penjelasan kode 5.13 terdapat pada Tabel 5.13.

Tabel 5.12 Penjelasan dari Kode 5.13

Atribut	Penjelasan
raw_df_klb	Data yang berisikan status KLB aktual dan angka rata-rata kejadian demam berdarah satu tahun sebelumnya
for_df	Data yang berisikan hasil peramalan yang telah dilakukan pada tahap 5.3
concat	Fungsi yang digunakan untuk menggabungkan <i>dataframe</i>
axis=1	Atribut pada fungsi concat untuk menandakan penggabungan yang dilakukan adalah secara horizontal (penambahan kolom)
sort=False	Atribut pada fungsi concat untuk mengontrol pengurutan pada index, jika bernilai 'False' maka tidak dilakukan pengurutan index (index disamakan dengan <i>dataframe</i> yang digabungkan).

5.4.3. Pengukuran Performa Klasifikasi

Hasil klasifikasi perlu dievaluasi dengan beberapa pengukuran agar dapat diketahui seberapa baik klasifikasi yang telah dilakukan. Pengukuran yang akan digunakan dalam penelitian kali ini adalah akurasi, sensitivitas, dan spesifisitas seperti yang telah dijelaskan pada subbab 4.3.3. Evaluasi hasil klasifikasi dapat dilakukan menggunakan Kode 5.14.

Seperti halnya pengujian pada tahap peramalan (subbab 5.3.9), evaluasi hasil klasifikasi dilakukan dengan tiga macam data, yaitu data *test-set*, data dengan proporsi yang berbeda, dan data kecamatan lain pada dataran yang sama. Nilai akurasi, spesifisitas, dan sensitivitas didapatkan dengan membuat *confusion matrix* terlebih dahulu. Kemudian ketiga nilai itu didapatkan dari rumus yang sudah dijelaskan pada subbab 2.2.10. Dari nilai-nilai yang dihasilkan kemudian dibandingkan dan menghasilkan model XGBoost dan Klasifikasi Berbasis Aturan yang paling baik.

```
#pengukuran kinerja klasifikasi
labels = ['KLB', 'non-KLB']
cm = confusion_matrix(klb_df.iloc[:,klb_df.columns ==
'KLB_act'], klb_df.iloc[:,klb_df.columns == 'KLB_pred'],
labels)
sensitivitas = cm[0,0]/(cm[0,0]+cm[0,1])
spesifisitas = cm[1,1]/(cm[1,0]+cm[1,1])
akurasi = akurasi_score(klb_df.iloc[:,klb_df.columns ==
'KLB_act'], klb_df.iloc[:,klb_df.columns == 'KLB_pred'])
```

Kode 5.14 Evaluasi Klasifikasi

Penjelasan Kode 5.14 tertang pada Tabel 5.14.

Tabel 5.13 Penjelasan dari Kode 5.14

Atribut	Penjelasan
confusion_matrix	Fungsi untuk membuat <i>confusion matrix</i> dari data hasil klasifikasi
akurasi_score	Fungsi yang digunakan untuk memanggil nilai akurasi klasifikasi

5.5. Peramalan Data 24 Periode Selanjutnya

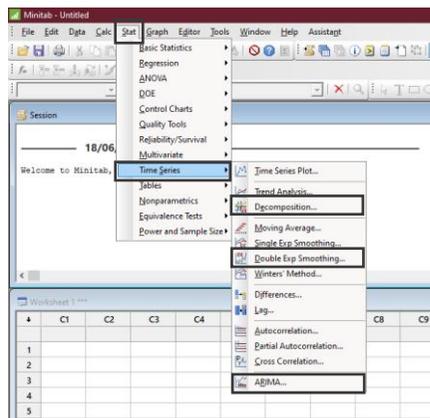
Peramalan data untuk periode Januari 2019 hingga Desember 2020 dapat dilakukan dengan tiga tahap, yaitu peramalan variabel independen, peramalan angka kejadian demam berdarah, dan pelabelan KLB dan non-KLB untuk data tersebut.

5.5.1. Peramalan Variabel Independen 24 Periode Selanjutnya

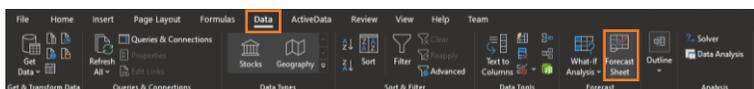
Seperti yang telah dijelaskan pada sub-bab 4.4, peramalan variabel independen didasari oleh pola data yang terbentuk pada tiap variabel. Diperoleh empat metode peramalan yang sesuai, yakni dekomposisi multiplikatif, ARIMA, Exponential Smoothing dan Double Exponential Smoothing. Penggunaan *tools* Minitab digunakan untuk mempermudah proses peramalan variabel iklim dan jumlah penduduk dengan metode dekomposisi multiplikatif, ARIMA, dan Double Exponential Smoothing. Sedangkan Microsoft Excel digunakan untuk

memproses data Angka Bebas Jentik dengan metode Exponential Smoothing.

Proses peramalan pada Minitab dilakukan dengan menggunakan pilihan pada menu Stat > Time Series, lalu memilih metode yang digunakan seperti yang ditunjukkan pada Gambar 5.1. Sedangkan proses peramalan pada Excel dapat dilakukan dengan memilih data yang hendak diramal, kemudian pergi ke menu Data > Forecast Sheet seperti yang tertera pada Gambar 5.2.



Gambar 5.1 Menu Pada Minitab



Gambar 5.2 Menu Pada Excel

5.5.2. Peramalan Kejadian Demam Berdarah untuk 24 Periode Selanjutnya

Setelah seluruh variabel masukan yang dibutuhkan pada kurun waktu Januari 2019 hingga 2020 telah didapatkan, maka peramalan kejadian demam berdarah 24 periode mendatang dapat dilakukan. Peramalan dilakukan setiap satu periode dengan 24 kali pengulangan agar menghasilkan angka kejadian

demam berdarah untuk 24 periode seperti yang tertera pada Kode 5.15.

```
def predict_kd(cls, ram_loc, loc, grid_mse, shift_total,
scalery):
    output = 24

    vi = ram_loc.iloc[-output:, :]
    raw_vi = ram_loc.iloc[:-output, :]
    scaler_vi = MinMaxScaler()
    scaler_vi.fit(raw_vi)
    vi = scaler_vi.transform(vi)
    vi = pd.DataFrame(vi, columns=ram_loc.columns[0:6])

    kdb24 = loc.iloc[-output:, loc.columns == 'KDB']
    kdb24 = scalery.transform(kdb24)
    kdb24 = pd.DataFrame(kdb24, columns=['KDB'])

    for i in range(output):
        kdb24 = kdb24.append(pd.Series(),
ignore_index=True)
        kdb24 = cls.Lag(kdb24, shift_total)

        if shift_total != 0:
            input = pd.concat([vi.iloc[i:i+1,:],
                                kdb24.iloc[-
1:,:1:]], axis=1)
        else:
            input = vi.iloc[i:i+1,:]

        kdb24 = kdb24.iloc[-output:, kdb24.columns ==
'KDB']
        kdb24 = kdb24.fillna(cls.prediction(input,
grid_mse))

        kdb24 = np.round(scalery.inverse_transform(kdb24))
    return kdb24
```

Kode 5.15 Peramalan 24 Periode Kedepan

Pada Kode 5.15, peramalan angka kejadian demam berdarah 24 periode selanjutnya diawali dengan *generate* data variabel independen serta melakukan normalisasi pada data tersebut. Lalu iterasi sebanyak 24 kali dilakukan setelah *dataframe* untuk menyimpan hasil peramalan telah dibuat. Dalam iterasi ini, data angka kejadian demam berdarah dinormalisi agar dapat menjadi

input peramalan sebagai fitur *lag*. Selanjutnya data diramalkan dengan menggunakan model terbaik yang telah dipilih dan dilakukan denormalisasi. Sedangkan untuk skenario peramalan yang tidak melibatkan variabel independen utama (suhu, curah hujan, kelembapan udara, jumlah penduduk dan angka bebas jentik) sebagai *input*, maka data diramalkan dengan hanya menggunakan fitur *lag* hasil proses *windowing* sehingga kode yang berhubungan dengan variabel independen dapat dihilangkan untuk skenario ini. Penjelasan atribut yang ada pada Kode 5.15 tertera pada Tabel 5.14.

Tabel 5.14 Penjelasan dari Kode 5.15

Atribut	Penjelasan
ram_loc	Data yang berisi hasil ramalan variabel independen pada tahap 5.5.1.
loc	Data yang berisi angka aktual kejadian demam berdarah
prediction	<i>Method</i> yang dibuat untuk memanggil model yang terbaik dengan menggunakan fungsi <i>best_estimator</i>
append, concat	Fungsi yang digunakan untuk menggabungkan data pada <i>dataframe</i> ataupun <i>series</i> .
scalery	<i>Scaler</i> untuk variabel target

5.5.3. Pelabelan KLB untuk 24 Periode Selanjutnya

Tahap terakhir untuk peramalan data 24 periode mendatang adalah memberikan label KLB dan non-KLB untuk data kurun waktu Januari 2019 hingga Desember 2020. Tahap ini dilakukan menggunakan Microsoft Excel dengan menggunakan aturan yang sama dengan tahap 5.4.8. Data yang telah diberi label kemudian dapat dilakukan analisa lebih lanjut.

Halaman ini sengaja dikosongkan

BAB VI HASIL DAN PEMBAHASAN

Pada bab ini dijelaskan mengenai hasil dari proses uji coba dan pembahasan serta analisis terhadap hasil yang diperoleh dari proses pelatihan dan pengujian model XGBoost dan Klasifikasi Berbasis Aturan.

6.1. Hasil Pra-Proses Data

Hasil dari implementasi pra-proses data berkaitan dengan perubahan data iklim, perubahan data penduduk, penanganan *missing value*, pemilihan kecamatan, normalisasi, *windowing* dan hasil uji korelasi variabel bebas.

6.1.1. Hasil Pembentukan Data Iklim

Salah satu contoh data variabel iklim harian yang diproses pada tahap ini yaitu data iklim pada bulan Januari tahun 2010 yang tertera pada Tabel 6.1. Pada tabel Tabel 6.1, atribut 'Tavg' merepresentasikan suhu rata-rata, 'RH_avg' merepresentasikan rata-rata kelembapan udara, 'RR' merepresentasikan rata-rata curah hujan, dan 'ff_avg' merepresentasikan rata-rata kecepatan angin. Masing-masing variabel tersebut dirata-rata dan didapatkan hasil variabel iklim untuk bulan Januari 2010 seperti yang tertera pada Tabel 6.2.

Tabel 6.1 Hasil Pra Proses Variabel Iklim bulan Januari 2010

Periode	Suhu	Kelembapan Udara	Curah Hujan	Kecepatan Angin
Januari 2010	26.02	84.74	10.52	0.84

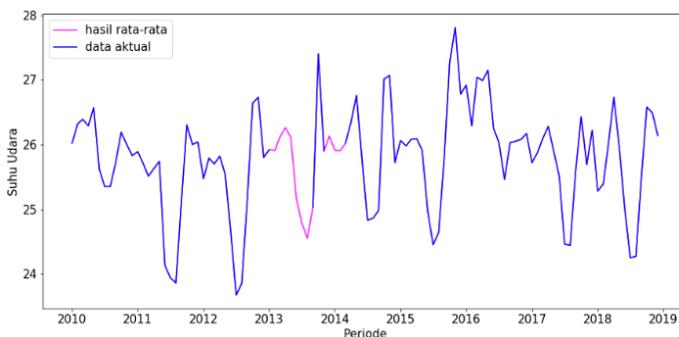
6.1.2. Hasil Penanganan *Missing Value*

Pada subbab ini menjelaskan penanganan *missing value* yang dilakukan pada variabel iklim (suhu, curah hujan, kelembapan udara, dan kecepatan angin) serta Angka Bebas Jentik (ABJ). Salah satu contoh penanganan *missing value* pada variabel iklim

tertera pada Gambar 6.1, dimana variabel suhu udara yang bernilai *null* diisi dengan melakukan rata-rata angka suhu udara pada bulan yang sama.

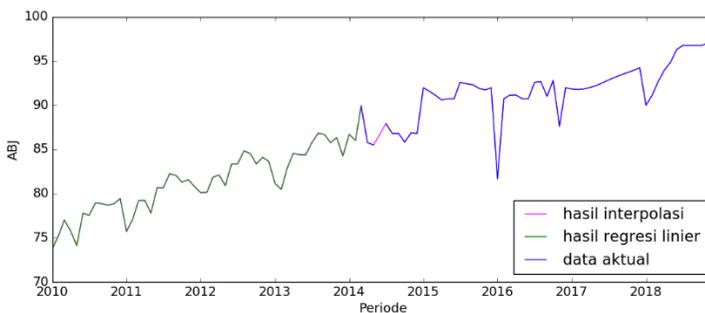
Tabel 6.2 Data Harian Iklim Bulan Januari 2010

Tanggal	Tavg	RH_avg	RR	ff_avg
01-01-2010	26,2	86	1	1
02-01-2010	26,9	85	10	0
03-01-2010	27,3	84	6	1
04-01-2010	27,2	84	10	1
05-01-2010	26,3	83	1	0
06-01-2010	26,8	83	0	0
07-01-2010	26,4	86	0	0
08-01-2010	25,7	88	15	0
09-01-2010	24,2	92	12	0
10-01-2010	25,6	88	7	0
11-01-2010	25,9	81	2	2
12-01-2010	26,2	85	20	2
13-01-2010	25,7	83	44	2
14-01-2010	25,7	83	1	1
15-01-2010	26,3	76	6	1
16-01-2010	26	94	2	1
17-01-2010	26	80	8	1
18-01-2010	27,3	75	0	1
19-01-2010	26,6	84	5	1
20-01-2010	25,7	86	2	1
21-01-2010	24,8	89	52	1
22-01-2010	25,3	88	4	0
23-01-2010	26,9	82	0	0
24-01-2010	26,2	85	0	0
25-01-2010	25,6	84	19	2
26-01-2010	25,4	84	90	1
27-01-2010	25,3	86	1	2
28-01-2010	25,1	88	2	1
29-01-2010	26,2	85	5	1
30-01-2010	25,8	86	0	1
31-01-2010	26,1	84	1	1



Gambar 6.1 Penanganan *Missing Value* pada Variabel Suhu Udara Kec. Kepanjen

Salah satu contoh penanganan *missing value* pada variabel Angka Bebas Jentik tertera pada Gambar 6.2, dimana variabel ABJ yang bernilai *null* lebih dari 20% data yakni pada tahun 2010 hingga 2014 diisi dengan menggunakan regresi linier, hal ini dilakukan untuk menghindari pengisian data yang bersifat garis lurus linier. Sedangkan nilai *null* pada tahun 2014 diisi dengan mekanisme inperpolasi seperti yang telah dijelaskan sebelumnya pada subbab 5.2.3.



Gambar 6.2 Penanganan *Missing Value* pada Variabel ABJ Kec. Wajak

6.1.3. Hasil Pemilihan Kecamatan Pembentuk Model

Salah satu contoh hasil perhitungan korelasi antarkecamatan ditunjukkan pada Tabel 6.3, yakni perhitungan korelasi pada

dataran rendah. Dalam matriks tersebut, setiap atribut nama kecamatan merepresentasikan jumlah kasus demam berdarah pada masing-masing kecamatan dengan periode bulanan. Nilai di dalam matriks tersebut merepresentasikan seberapa kuat hubungan antarvariabel, dalam hal ini jumlah kasus kejadian demam berdarah pada suatu kecamatan dengan kecamatan lainnya.

Tabel 6.3 Hasil Korelasi Antarkecamatan di Dataran Rendah

	Turen	Kepanjen	Gondanglegi	Bululawang	Donomulyo
Turen	1	0.813	0.559	0.341	0.538
Kepanjen	0.813	1	0.639	0.575	0.544
Gondanglegi	0.559	0.639	1	0.413	0.445
Bululawang	0.34	0.575	0.413	1	0.390
Donomulyo	0.538	0.544	0.445	0.390	1

Dari hasil korelasi tersebut dapat diketahui bahwa kecamatan Kepanjen memiliki korelasi yang paling kuat terhadap kecamatan lain. Kecamatan Kepanjen memiliki nilai korelasi yang tinggi yaitu diatas 0.5 terhadap Kecamatan Turen, Gondanglegi, Bululawang dan Domoluyo. Kecamatan tersebut memiliki nilai rata-rata korelasi sebesar 0,6475. Maka dari itu, kecamatan Kepanjen akan menjadi kecamatan pembentuk model pada dataran rendah.

Mekanisme pemilihan kecamatan pembentuk model yang sama juga dilakukan pada dataran sedang dan tinggi. Hasil uji korelasi antarkecamatan yang dilakukan dapat dilihat pada Lampiran E. Pada dataran sedang, didapatkan kecamatan pembentuk model yaitu Kecamatan Pakisaji. Sedangkan untuk dataran tinggi kecamatan yang memiliki nilai korelasi yang paling tinggi terdapat pada kecamatan Ngajum sehingga kecamatan tersebut terpilih menjadi kecamatan pembentuk model pada dataran tinggi.

6.1.4. Hasil Normalisasi

Pada penelitian ini, semua data variabel independen yang digunakan sebagai *input* dalam membangun model perlu melewati proses normalisasi *min-max*. Salah satu contoh variabel yang dinormalisasi adalah variabel curah hujan. Perbandingan nilai variabel curah hujan sebelum dan setelah normalisasi pada periode Januari 2010 hingga Maret 2010 tertera pada Tabel 6.4.

Tabel 6.4 Perbandingan Variabel Suhu Udara Sebelum dan Setelah Normalisasi pada Kec. Kepanjen

Periode	Variabel Curah Hujan	
	Sebelum Normalisasi	Sesudah Normalisasi
Jan-2010	26.02	0.567633
Feb-2010	26.32	0.640097
Mar-2010	26.39	0.657005

Dapat dilihat dari Tabel 6.4 bahwa data yang semula berskala puluhan menjadi data dengan rentang nilai 0 sampai 1 setelah dilakukan normalisasi *min-max*. Perhitungan nilai normalisasi yang dilakukan mengacu pada rumus Persamaan 2.1.

6.1.5. Hasil Pemilihan Fitur dengan Uji Korelasi

Pada subbab ini dijelaskan hasil uji korelasi variabel independen yang dilakukan untuk tiap dataran. Uji korelasi ini akan digunakan untuk memilih variabel independen yang digunakan dalam membangun model untuk skenario TOP1-VI, TOP2-VI, dan TOP3-VI seperti yang ditunjukkan pada Tabel 4.1.

a. Dataran Rendah

Data yang digunakan untuk melakukan uji korelasi pada dataran rendah adalah data kecamatan Kepanjen sebagai kecamatan pembentuk model. Hasil uji korelasi variabel independen di dataran rendah ditunjukkan pada Tabel 6.5, dimana atribut 'CH' merupakan variabel curah hujan, 'JP' merupakan variabel jumlah penduduk, 'SU' merupakan variabel suhu udara, 'KU' merupakan variabel kelembapan udara, 'ABJ' merupakan

variabel Angka Bebas Jentik dan 'KA' merupakan variabel kecepatan angin.

Tabel 6.5 Uji Korelasi Variabel Independen (Dataran Rendah)

	Variabel Independen					
	CH	JP	SU	KU	ABJ	KA
Korelasi Terhadap KDB	0,295	0,279	0,106	0,087	-0,087	-0,091

Dari Tabel 6.5, dapat dilihat bahwa di dataran rendah tiga nilai korelasi yang paling kuat terhadap kasus demam berdarah dimiliki oleh variabel curah hujan, jumlah penduduk dan suhu udara. Ketiga variabel tersebut memiliki nilai korelasi positif yang berarti semakin sering terjadi hujan, semakin tinggi jumlah penduduk, dan semakin tinggi suhu udara akan menyebabkan meningkatnya jumlah kasus demam berdarah. Dari hasil tersebut maka didapatkan variabel terpilih untuk digunakan pada skenario pembentukan model seperti yang tertera pada Tabel 6.6.

Tabel 6.6 Variabel Independen Tiap Skenario (Dataran Rendah)

Skenario	Variabel Independen
ALL-VI	Curah Hujan, Jumlah Penduduk, Suhu Udara, Kelembapan Udara, Angka Bebas Jentik, Kecepatan Angin
TOP3-VI	Curah Hujan, Jumlah Penduduk, Suhu Udara
TOP2-VI	Curah Hujan, Jumlah Penduduk
TOP1-VI	Curah Hujan
NO-VI	-

Khusus untuk skenario 'NO-VI', variabel independen utama (curah hujan, jumlah penduduk, suhu udara, kelembapan udara, angka bebas jentik, dan kecepatan angin) tidak digunakan untuk membangun model. Pada skenario ini, masukan untuk

pembangun model hanya didapat dari hasil proses *windowing*, yakni fitur *lag*, seperti yang dijelaskan pada subbab 5.2.7.

b. Dataran Sedang

Data yang digunakan untuk melakukan uji korelasi pada dataran sedang adalah data kecamatan Pakisaji sebagai kecamatan pembentuk model. Hasil uji korelasi variabel independen di dataran rendah ditunjukkan pada Tabel 6.7, dimana atribut 'CH' merupakan variabel curah hujan, 'JP' merupakan variabel jumlah penduduk, 'SU' merupakan variabel suhu udara, 'KU' merupakan variabel kelembapan udara, 'ABJ' merupakan variabel Angka Bebas Jentik dan 'KA' merupakan variabel kecepatan angin.

Tabel 6.7 Uji Korelasi Variabel Independen (Dataran Sedang)

	Variabel Independen					
	ABJ	CH	KU	JP	KA	SU
Korelasi Terhadap KDB	-0,379	0,205	0,191	-0,133	-0,045	0,038

Dari Tabel 6.7, dapat dilihat bahwa tiga variabel dengan koefisien korelasi tertinggi terhadap jumlah kasus demam berdarah pada dataran sedang adalah Angka Bebas Jentik, curah hujan dan kelembapan udara. Nilai korelasi paling kuat pada variabel ABJ yakni sebesar -0,379 menunjukkan bahwa ada hubungan yang bertolak belakang antara variabel kejadian demam berdarah dengan ABJ berarti semakin tinggi angka ABJ, maka semakin rendah angka kejadian demam berdarah. Sedangkan, variabel curah hujan dan kelembapan udara memiliki nilai korelasi positif dimana seringnya terjadi hujan dan tingkat kelembapan yang tinggi pada suatu daerah dapat menyebabkan kenaikan pada jumlah kasus demam berdarah. Dari hasil tersebut maka didapatkan variabel terpilih untuk digunakan pada skenario pembentukan model seperti yang tertera pada Tabel 6.8.

Khusus untuk skenario ‘NO-VI’, variabel independen utama (curah hujan, jumlah penduduk, suhu udara, kelembapan udara, angka bebas jentik, dan kecepatan angin) tidak digunakan satupun untuk membangun model. Pada skenario ini, masukan untuk pembangun model hanya didapat dari hasil proses *windowing* pada tahap 5.2.7.

Tabel 6.8 Variabel Independen Tiap Skenario (Dataran Sedang)

Skenario	Variabel Independen
ALL-VI	Angka Bebas Jentik, Curah Hujan, Kelembapan Udara, Jumlah Penduduk, Kecepatan Angin, Suhu Udara
TOP3-VI	Angka Bebas Jentik, Curah Hujan, Kelembapan Udara
TOP2-VI	Curah Hujan, Kelembapan Udara
TOP1-VI	Kelembapan Udara
NO-VI	-

c. Dataran Tinggi

Data yang digunakan untuk melakukan uji korelasi pada dataran tinggi adalah data Kecamatan Ngajum sebagai kecamatan pembentuk model. Hasil uji korelasi variabel independen di dataran rendah ditunjukkan pada Tabel 6.9, dimana atribut ‘CH’ merupakan variabel curah hujan, ‘JP’ merupakan variabel jumlah penduduk, ‘SU’ merupakan variabel suhu udara, ‘KU’ merupakan variabel kelembapan udara, ‘ABJ’ merupakan variabel Angka Bebas Jentik dan ‘KA’ merupakan variabel kecepatan angin.

Tabel 6.9 Uji Korelasi Variabel Independen (Dataran Tinggi)

	Variabel Independen					
	JP	KU	CH	KA	ABJ	SU
Korelasi Terhadap KDB	-0,276	0,236	0,209	0,125	-0,108	0,051

Dari Tabel 6.9, dapat dilihat bahwa tiga variabel dengan koefisien korelasi tertinggi terhadap jumlah kasus demam berdarah pada dataran tinggi adalah jumlah penduduk, kelembapan udara, dan curah hujan. Nilai korelasi paling kuat terdapat pada variabel jumlah penduduk yakni sebesar $-0,276$ yang menunjukkan bahwa ada hubungan yang bertolak belakang antara variabel jumlah penduduk dengan kasus demam berdarah. Sedangkan, variabel kelembapan udara dan curah hujan memiliki nilai korelasi positif dimana tingkat kelembapan yang tinggi dan seringnya terjadi hujan pada suatu daerah dapat menyebabkan kenaikan pada jumlah kasus demam berdarah. Dari hasil tersebut maka didapatkan variabel terpilih untuk digunakan pada skenario pembentukan model seperti yang tertera pada Tabel 6.10.

Tabel 6.10 Variabel Independen Tiap Skenario (Dataran Tinggi)

Skenario	Variabel Independen
ALL-VI	Jumlah Penduduk, Kelembapan Udara, Curah Hujan, Kecepatan Angin, Angka Bebas Jentik, Suhu Udara
TOP3-VI	Jumlah Penduduk, Kelembapan Udara, Curah Hujan
TOP2-VI	Jumlah Penduduk, Kelembapan Udara
TOP1-VI	Jumlah Penduduk
NO-VI	-

Khusus untuk skenario ‘NO-VI’, variabel independen utama (curah hujan, jumlah penduduk, suhu udara, kelembapan udara, angka bebas jentik, dan kecepatan angin) tidak digunakan satupun untuk membangun model. Pada skenario ini, masukan untuk pembangun model hanya didapat dari hasil proses *windowing* pada tahap 5.2.7.

Dari keseluruhan perhitungan korelasi variabel independen terhadap kejadian demam berdarah yang didapatkan, dapat diketahui bahwa nilai-nilai korelasi yang dihasilkan tidak melebihi 0,5. Hal ini menunjukkan bahwa dalam studi kasus ini variabel-variabel independen yang digunakan tidak memiliki

korelasi yang cukup kuat dengan kejadian demam berdarah. Namun pada penelitian ini, variabel-variabel yang tersebut tetap digunakan berdasarkan urutan nilai korelasinya dari yang terbesar. Untuk menanggulangi masalah nilai korelasi yang kecil, dilakukan skenario pembentukan peramalan univariat (hanya dengan fitur lag) untuk diujicobakan beserta skenario lainnya yang telah dicantumkan pada Tabel 4.1. Hal ini dilakukan untuk mengetahui sensitivitas variabel-variabel yang digunakan tersebut terhadap kejadian demam berdarah sehingga dapat membuktikan dan mencari tahu dari sekian banyak skenario yang dilakukan, pemilihan fitur mana yang memiliki hasil performa peramalan yang paling baik.

6.1.6. Hasil Windowing

Salah satu hasil proses *windowing* yang dilakukan pada penelitian kali ini tertera pada Tabel 6.11, dimana terdapat 3 fitur *lag* tambahan untuk data periode Januari 2010 hingga Desember 2010. Penambahan fitur *lag* tersebut didapatkan dari angka kejadian demam berdarah yang telah dinormalisasi pada satu bulan sebelumnya untuk fitur *lag* pertama, 2 bulan sebelumnya untuk fitur *lag* kedua, dan 3 bulan sebelumnya untuk fitur *lag* ketiga.

Pada Tabel 6.11, atribut 'KDB' merepresentasikan angka kejadian demam berdarah yang sudah dinormalisasi. Sedangkan 'kdblag-1', 'kdblag-2', dan 'kdblag-3' berturut-turut menunjukkan fitur lag tambahan pertama, kedua, dan ketiga. Kekosongan dua data periode Januari pada atribut 'kdblag-2' dan 'kdblag-3' serta satu data periode Februari pada atribut 'kdblag-3' disebabkan oleh tidak ada data sebelum Januari 2010 yang dapat digunakan untuk mengisi fitur *lag* tersebut. Maka dari itu, untuk periode yang terdapat nilai *null* pada fitur *lag* dihilangkan agar variabel input dalam pembentukan model tidak ada yang bernilai *null*.

Tabel 6.11 Hasil Windowing Pada Kec. Pakisaji degan 3 Lag

Periode	KDB	kdblag-1	kdblag-2	kdblag-3
Jan-10	0,32	NaN	NaN	NaN
Feb-10	0,55	0,32	NaN	NaN
Mar-10	0,19	0,55	0,32	NaN
Apr-10	0,13	0,19	0,55	0,32
Mei-10	0,23	0,13	0,19	0,55
Jun-10	0,06	0,23	0,13	0,19
Jul-10	0,06	0,06	0,23	0,13
Ags-10	0,23	0,06	0,06	0,23
Sep-10	0,03	0,23	0,06	0,06
Okt-10	0,00	0,03	0,23	0,06
Nov-10	0,06	0,00	0,03	0,23
Des-10	0,00	0,06	0,00	0,03

6.2. Hasil Peramalan dan Klasifikasi Dataran Rendah

Pada subbab ini dijelaskan hasil peramalan pada dataran rendah menggunakan metode XGBoost yang dilanjutkan dengan Klasifikasi Berbasis Aturan untuk menentukan status KLB dan non-KLB pada data. Sub-bab ini terdiri dari tiga bagian yaitu hasil peramalan, hasil klasifikasi, dan hasil perhitungan 24 periode mendatang.

6.2.1. Hasil Peramalan pada Dataran Rendah

Pada subbab ini pembahasan difokuskan pada hasil peramalan yang dilakukan dengan metode XGBoost di dataran rendah. Sub-bab ini terdiri dari dua bagian yaitu hasil uji coba skenario dan parameter, hasil peramalan pada kecamatan pembentuk model dan hasil peramalan pada kecamatan selain pembentuk model.

a. Hasil Uji Coba Skenario Peramalan

Pada proses pembentukan model peramalan di dataran rendah, terdapat beberapa skenario yang dilakukan dan masing-masing

skenario menghasilkan satu model. Dari keseluruhan model yang telah dibangun tersebut kemudian dilakukan pengujian dengan membandingkan nilai RMSE tiap model.

Tabel 6.12 Hasil RMSE untuk *test-set* di Kec. Kepanjen

Skenario			RMSE
Pembagian Data	Pemilihan Fitur	Windowing	
80-20	ALL-VI	6	2,5533
	TOP3-VI	3	2,7680
	TOP2-VI	7	2,7916
	TOP1-VI	1	2,7223
	NO-VI	1	2,3992
70-30	ALL-VI	6	2,7735
	TOP3-VI	10	3,0812
	TOP2-VI	11	2,5324
	TOP1-VI	2	2,5863
	NO-VI	2	2,5340
60-40	ALL-VI	0	4,9510
	TOP3-VI	12	4,1306
	TOP2-VI	2	4,7536
	TOP1-VI	0	4,0464
	NO-VI	1	4,3761
50-50	ALL-VI	12	5,2783
	TOP3-VI	0	4,4679
	TOP2-VI	0	4,7536
	TOP1-VI	0	4,2348
	NO-VI	1	3,9266

Data yang digunakan untuk pembentukan model pada dataran rendah yaitu data Kecamatan Kepanjen dengan menggunakan *input* sesuai dengan skenario yang telah dijelaskan pada subbab

4.1.2 bagian F. Hasil dari pengujian model pada *test-set* kecamatan Kepanjen ditunjukkan pada Tabel 6.16. Sedangkan hasil peramalan lengkap dengan informasi parameter yang digunakan dijelaskan pada Lampiran E.

Pada Tabel 6.12, dapat diketahui nilai RMSE untuk tiap kombinasi skenario pembagian proporsi data dan pemilihan fitur diikuti dengan jumlah *lag* optimal pada skenario tersebut. Dari model-model yang dibangun, model terbaik didapatkan dari urutan nilai RMSE yang terkecil. Terdapat tiga kandidat model terbaik dengan RMSE yang paling kecil seperti yang tertera pada Tabel 6.12, yakni yang pertama pada skenario pembagian data 80-20 dan menggunakan *input* berupa fitur *lag* berjumlah 1 dengan nilai RMSE sebesar 2,3991916. Kemudian kandidat model yang kedua yakni dengan pembagian data 70-30 yang melibatkan variabel *input* curah hujan, jumlah penduduk dan fitur *lag* berjumlah 2 dengan nilai RMSE sebesar 2,534007. Model terbaik ketiga yaitu model dengan skenario pembagian data 70-30 dan variabel *input* berupa fitur *lag* berjumlah 2 dengan nilai RMSE sebesar 2,534007. Ketiga model tersebut memiliki nilai SMAPE berturut-turut, 26.74%, 27.57%, dan 25.60%.

Tabel 6.13 Hasil *Tuning* Parameter di Dataran Rendah

Model	Skenario	Parameter XGBoost							
		col	lr	gm	md	mcw	n_est	alpha	sub
1	80-20	0.5	0.5	0	2	1	100	0.01	0.6
	NO-VI								
	Lag 1								
2	70-30	0.5	0.5	0.01	8	1	50	0	1
	TOP2								
	Lag 11								
3	70-30	0.5	0.1	0.01	8	1	50	0.01	0.6
	NOVI								
	Lag 2								

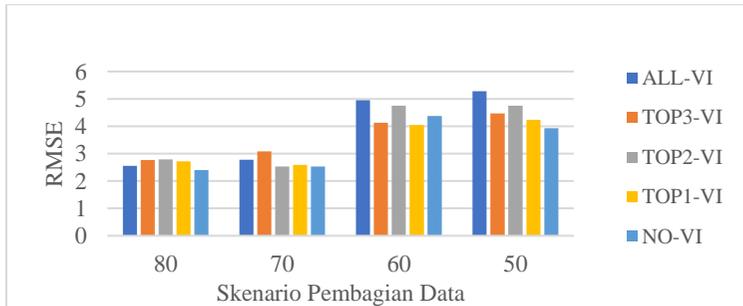
Pada Tabel 6.13, informasi mengenai parameter hasil *tuning* menggunakan *grid-search* ditampilkan untuk tiga model peramalan terbaik di dataran rendah. Tabel ini akan menjadi acuan informasi model yang akan digunakan pada untuk

menampilkan hasil tahap pengujian proporsi dan kecamatan lain. Pada tabel tersebut, atribut 'Model' menunjukkan urutan model terbaik berdasarkan nilai RMSE terkecil, 'Skenario' menunjukkan skenario pembentukan model yang digunakan, dan 'Parameter' menunjukkan parameter yang dihasilkan dari proses *tuning* yang dilakukan. Parameter yang tertera pada tabel tersebut antara lain, 'col' merepresentasikan *colsample_bytree*, 'lr' merepresentasikan *learning_rate*, 'gm' merepresentasikan *gamma*, 'md' merepresentasikan *max_depth*, 'mcw' merepresentasikan *min_child_weight*, 'n_est' merepresentasikan *n_estimator*, alpha merepresentasikan *reg_alpha*, dan 'sub' yang merepresentasikan parameter *subsample*.

Dari hasil *tuning* parameter, dapat diketahui bahwa tiga model dengan RMSE terkecil menggunakan parameter yang sama untuk rasio fitur yang digunakan dalam membentuk model pada parameter *colsample_bytree* sebesar 0.5, dan nilai regularisasi *min_child_weight* sebesar 1. Hal ini berarti model hanya menggunakan separuh dari keseluruhan fitur yang ada untuk membangun model tiap iterasinya dengan pembobotan untuk memecah *tree* sebesar 1. Sedangkan dihasilkan pula nilai parameter *learning rate* sebesar 0.5 pada model pertama dan kedua, sedangkan 0.1 untuk model ketiga. Model kedua dan ketiga memiliki kesamaan parameter pada nilai *gamma* sebesar 0.01, *max_depth* sebesar 8, jumlah estimator berjumlah 50. Pada model pertama, nilai parameter *gamma*, *max_depth* dan jumlah estimator berturut-turut sebesar 0, 2, dan 100. Untuk nilai *alpha* dan *subsample*, model pertama dan ketiga memiliki kesamaan yaitu sebesar 0.01 dan 0.6, sedangkan pada model kedua nilai kedua parameter tersebut berturut-turut adalah 0 dan 1. Informasi hasil *tuning* parameter lebih lengkap terdapat pada Lampiran E.

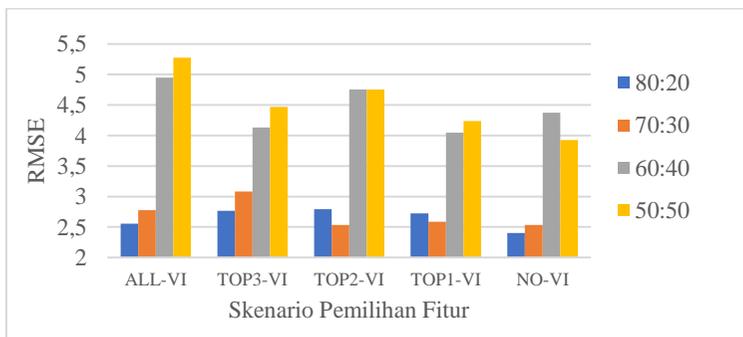
Dari hasil peramalan yang didapatkan pada Tabel 6.12, selain informasi mengenai model terbaik, dapat diketahui pula pola persebaran nilai RMSE yang dihasilkan. Persebaran RMSE

berdasarkan skenario pembagian data di dataran rendah ditunjukkan pada Gambar 6.3.



Gambar 6.3 Persebaran RMSE berdasarkan Skenario Pembagian Data (Dataran Rendah)

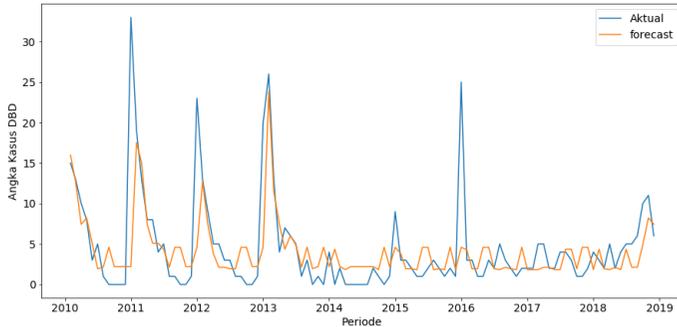
Dari grafik yang terbentuk pada Gambar 6.3, dapat diketahui bahwa RMSE yang dihasilkan untuk pembagian data 80-20 dan 70-30 relatif lebih rendah dibandingkan RMSE yang dihasilkan dari pembagian data 60-40 dan 50-50. Hal ini disebabkan oleh semakin sedikitnya proporsi data pelatihan, maka semakin sedikit pula jumlah baris data yang menjadi *input* dalam pembentukan model. Hal tersebut dapat menyebabkan model yang dihasilkan *underfit*, yakni model tidak bisa melihat logika di belakang data, sehingga tidak bisa melakukan peramalan secara tepat dan memiliki nilai *error* yang relatif tinggi.



Gambar 6.4 Persebaran RMSE berdasarkan Skenario Pemilihan Fitur (Dataran Rendah)

Dari grafik yang terbentuk pada Gambar 6.4, dapat diketahui bahwa RMSE yang dihasilkan pada masing-masing skenario pemilihan fitur relatif beragam dan fluktuatif. Hal ini dapat disebabkan oleh beragamnya pembobotan masing-masing fitur saat berjalannya proses pelatihan dan adanya perbedaan jumlah *lag* yang optimal pada tiap skenarionya. Untuk proporsi data 60-40 dan 50-50, skenario pemilihan fitur yang memiliki nilai RMSE paling tinggi terdapat pada skenario 'ALL-VI', yakni pembentukan model dilakukan dengan menggunakan keseluruhan variabel independen. Pada proporsi data tersebut, model akan menghasilkan nilai RMSE yang lebih rendah jika hanya melibatkan beberapa variabel independen utama atau tidak sama sekali (hanya menggunakan fitur *lag*). Sedangkan untuk proporsi data 70-30, RMSE tertinggi ada pada skenario TOP3-VI dimana pembentukan model hanya melibatkan tiga variabel yang paling berpengaruh saja. Pada proporsi data ini, model akan menghasilkan nilai akurasi yang lebih tinggi jika dilatih dengan data variabel independen utama berjumlah satu, dua, atau tidak sama sekali (hanya menggunakan fitur *lag*). RMSE tertinggi pada proporsi data 80-20 terdapat pada TOP2-VI, dan selain skenario tersebut model dapat menghasilkan nilai RMSE yang lebih rendah.

Perbandingan data aktual dan hasil peramalan yang dilakukan pada kecamatan pembentuk model yaitu kepanjen dengan menggunakan model peramalan dengan RMSE terkecil tertera pada Gambar 6.5. Dapat dilihat pada gambar tersebut bahwa model dapat menangkap pola data dengan baik. Hal ini juga dapat dilihat dari nilai *Root Mean Square Error* (RMSE) yang dihasilkan yaitu sebesar 2,5323573 dan *Symmetric Mean Absolute Percentage Error* (SMAPE) sebesar 27.57%.



Gambar 6.5 Perbandingan Data Aktual dan Hasil Peramalan Kec. Kepanjen

Selain nilai RMSE pengujian pada *test-set*, tingkat *robustness* model juga perlu diperhitungkan untuk mengetahui seberapa baik model diterapkan pada data lainnya. Salah satu cara mengetahui tingkat *robustness* model yaitu dengan menguji model tersebut pada proporsi data yang berbeda. Hasil uji proporsional data untuk tiga model terbaik tertera pada Tabel 6.14. Hasil lengkap perbandingan model dapat dilihat pada Lampiran E.

Tabel 6.14 Hasil Uji Proporsional Data pada Kec. Kapanjen

Model Terbaik ke-	RMSE Uji Proporsional Data			Mean	St.dev
	proporsi 1	proporsi 2	proporsi 3		
1	2,211	3,765	3,508	3,161	0,680
2	2,621	2,394	2,221	2,412	0,164
3	2,655	3,858	3,779	3,431	0,550

Seperti yang tertera pada Tabel 6.14, terdapat tiga uji proporsi yang dilakukan, skenario uji tersebut telah dijelaskan pada subbab 4.2.3. Dari hasil uji proporsi yang didapat kemudian dihitung nilai rata-rata dan standar deviasinya agar dapat diketahui mana model yang memiliki tingkat *robustness* paling tinggi. Mengacu pada Tabel 6.14, dua atribut digunakan untuk menunjang perhitungan yakni atribut ‘Mean’ yang

menunjukkan nilai rata-rata hasil uji proporsi yang telah dilakukan dan ‘St.dev’ yang menunjukkan nilai standar deviasi yang dihasilkan. Dapat diketahui dari hasil uji proporsi tersebut bahwa model terbaik kedua ternyata menghasilkan nilai RMSE yang paling kecil diantara dua model lainnya baik secara rata-rata maupun nilai standar deviasi RMSE dengan nilai rata-rata RMSE sebesar 2,412 dan nilai standar deviasi RMSE sebesar 0,164. Hal ini menunjukkan bahwa meskipun suatu model memiliki nilai RMSE paling kecil pada pengujian *test-set*, belum tentu model tersebut juga baik digunakan pada proporsi data yang lain. Dalam kasus ini, model terbaik (model dengan RMSE terkecil) ternyata memiliki tingkat *robustness* yang tidak terlalu baik untuk proporsi data yang lain, sedangkan model yang bukan terbaik justru menghasilkan nilai *robustness* yang lebih tinggi terhadap proporsi data yang lain.

b. Hasil Peramalan pada Kecamatan Selain Pembentuk Model

Selain pengujian yang dilakukan pada data Kecamatan Kepanjen, model juga perlu diuji terhadap data kecamatan lain pada dataran yang sama. Di dataran rendah, kecamatan uji terdiri dari Kecamatan Turen, Kecamatan Donomulyo, Kecamatan Bululawang, dan Kecamatan Gondanglegi. Hasil perbandingan nilai RMSE yang dihasilkan pada tiap kecamatan uji untuk tiga model dengan RMSE terkecil yang ditunjukkan pada Tabel 6.15. Hasil lengkap perbandingan RMSE uji kecamatan lain terlampir pada Lampiran E.

Tabel 6.15 Perbandingan RMSE Uji Kecamatan Lain pada Dataran Rendah

		Model Terbaik ke-		
		1	2	3
RMSE Uji Kecamatan Lain	Turen	16,855	15,479	16,794
	Donomulyo	2,642	2,329	2,630
	Bululawang	3,387	2,541	3,085
	Gondanglegi	2,401	2,531	2,384
Mean		6,321	5,720	6,223
St.dev		6,093	5,635	6,108

Dari hasil pengujian terhadap kecamatan lain, dapat diketahui bahwa model terbaik kedua menghasilkan nilai rata-rata RMSE dan standar deviasi RMSE terkecil. Model tersebut menjadi model yang memiliki nilai rata-rata RMSE dan standar deviasi paling kecil baik untuk uji proporsi data maupun uji model pada kecamatan lain di dataran rendah. Sehingga, dapat diketahui pula bahwa suatu model terbaik yang memiliki nilai RMSE terkecil pada pengujian *test-set*, belum tentu memiliki tingkat *robustness* yang terbaik pula ketika model diuji pada data kecamatan lainnya.

Selain perhitungan nilai RMSE, pengukuran lain yang digunakan untuk menunjang analisis permalan yakni nilai SMAPE atau rata-rata persentase kesalahan absolut simetris. Nilai SMAPE yang dihasilkan model terbaik dengan RMSE terkecil di dataran rendah pada pengujian kecamatan selain pembentuk model ditunjukkan pada Tabel 6.16.

Tabel 6.16 Nilai SMAPE Kecamatan Lain (Dataran Rendah)

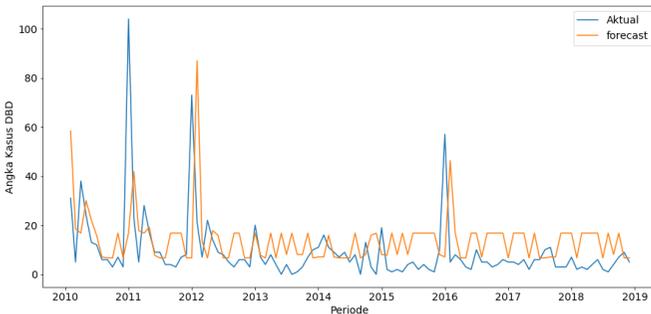
No	Kecamatan	SMAPE
1	Turen	42,84 %
2	Donomulyo	57,37%
3	Bululawang	37,43%
4	Gondanglegi	47,37%

i. Kecamatan Turen

Perbandingan jumlah kasus demam berdarah aktual dan hasil peramalan di Kecamatan Turen ditunjukkan pada Gambar 6.6. Dari gambar tersebut dapat diketahui bahwa beberapa kali model cukup baik dalam menangkap pola lonjakan jumlah kasus demam berdarah. Selain itu, dapat diketahui pula dari hasil perhitungan SMAPE yang ditampilkan pada Tabel 6.16, Kecamatan Turen memiliki SMAPE cukup tinggi jika dibandingkan dengan kecamatan lainnya yakni sebesar 42,84%. Padahal seperti yang telah dijelaskan pada subbab 6.1.4, Kecamatan Kepanjen sebagai pembentuk model memiliki korelasi yang paling tinggi terhadap kecamatan Turen yakni

dengan nilai korelasi sebesar 0.813. Maka dari itu, dapat diketahui bahwa nilai korelasi yang tinggi antarkecamatan tidak menjamin bahwa model yang akan dibangun akan menghasilkan nilai error yang lebih kecil pada kecamatan tersebut. Hal ini dapat disebabkan oleh pengaruh variabel independen yang berbeda pada tiap kecamatan sehingga mempengaruhi hasil peramalan untuk masing-masing kecamatan.

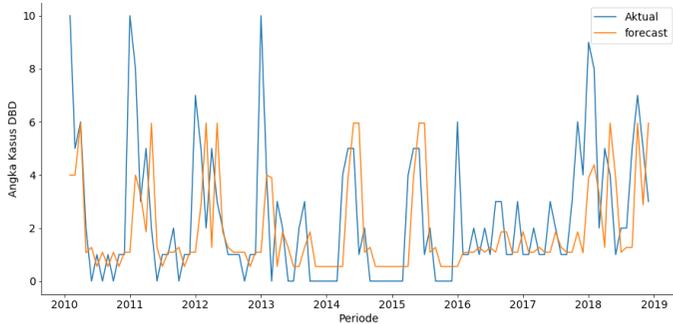
Kecamatan Turen memiliki nilai RMSE paling tinggi diantara kecamatan lainnya yaitu sebesar 16,855. Hal ini dapat disebabkan karena Kecamatan Turen memiliki banyak lonjakan angka kejadian demam berdarah hingga skala ratusan.



Gambar 6.6 Perbandingan Data Aktual dan Hasil Peramalan Kec. Turen

ii. Kecamatan Gondanglegi

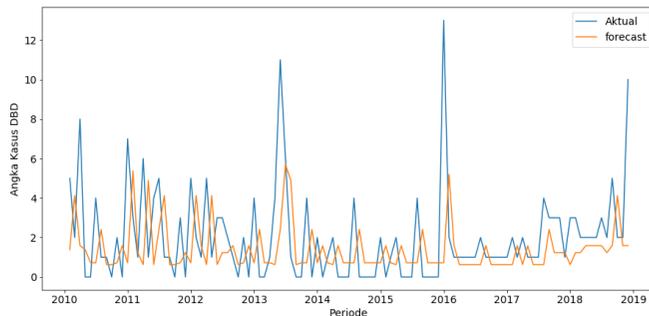
Perbandingan jumlah kasus demam berdarah aktual dan hasil peramalan di Kecamatan Gondanglegi ditunjukkan pada Gambar 6.7. Dari gambar tersebut dapat diketahui bahwa model dapat mengenali pola kenaikan dan penurunan jumlah kasus demam berdarah dengan cukup baik. Namun, dapat dilihat pula bahwa model beberapa kali gagal dalam meramalkan lonjakan pada data. Hal ini juga direpresentasikan dengan nilai RMSE sebesar 2,401 dan SMAPE sebesar 47,37%.



Gambar 6.7 Perbandingan Data Aktual dan Hasil Peramalan Kec. Gondanglegi

iii. Kecamatan Donomulyo

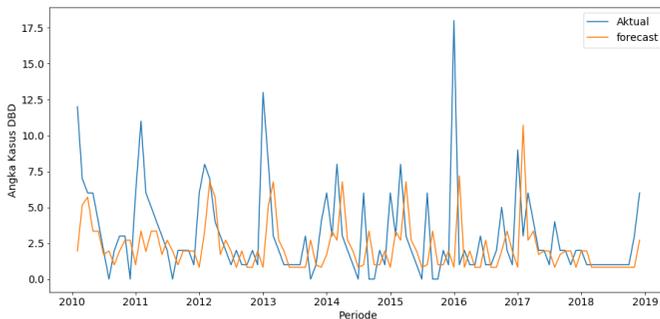
Perbandingan jumlah kasus demam berdarah aktual dan hasil peramalan di Kecamatan Donomulyo ditunjukkan pada Gambar 6.8. Dari gambar tersebut, dapat dilihat bahwa model dapat mengenali pola kenaikan dan penurunan yang terjadi pada data, namun kurang baik dalam mengenali lonjakan jumlah kasus demam berdarah yang terjadi. Hal ini direpresentasikan dengan nilai RMSE pada Kecamatan Donomulyo sebesar 2,642 dan SMAPE sebesar 57,37%.



Gambar 6.8 Perbandingan Data Aktual dan Hasil Peramalan Kec. Donomulyo

iv. Kecamatan Bululawang

Perbandingan jumlah kasus demam berdarah aktual dan hasil peramalan di Kecamatan Bululawang ditunjukkan pada Gambar 6.9. Dari gambar tersebut, dapat dilihat bahwa model dapat mengenali pola kenaikan dan penurunan yang terjadi pada data, namun beberapa kali kurang baik dalam mengenali lonjakan jumlah kasus demam berdarah yang terjadi. Hal ini direpresentasikan dengan nilai RMSE pada Kecamatan Bululawang sebesar 3,387. Di samping itu, Kecamatan Bululawang memiliki nilai SMAPE terkecil diantara kecamatan uji lainnya di dataran rendah yakni sebesar 37,43%.



Gambar 6.9 Perbandingan Data Aktual dan Hasil Peramalan Kec. Bululawang

6.2.2. Hasil Klasifikasi pada Dataran Rendah

Pada subbab ini pembahasan difokuskan pada hasil klasifikasi kejadian luar biasa yang dilakukan dengan metode Klasifikasi Berbasis Aturan di dataran rendah. Sub-bab ini terdiri dari dua bagian yaitu hasil klasifikasi pada kecamatan pembentuk model dan hasil klasifikasi pada kecamatan selain pembentuk model.

a. Hasil Uji Coba Skenario Klasifikasi

Dalam penelitian kali ini, pengukuran evaluasi klasifikasi biner yang digunakan adalah sensitivitas, spesifisitas, dan akurasi. Klasifikasi biner yang baik akan menghasilkan nilai yang tinggi untuk ketiga faktor tersebut, sedangkan klasifikasi biner yang

buruk akan menghasilkan nilai rendah pada faktor-faktor tersebut. Agar memastikan ketiga faktor tersebut baik pada suatu model, dilakukan perataan pada nilai akurasi, sensitivitas, dan spesifisitas seperti yang tertera pada Tabel 6.17 dengan atribut ‘Rata-rata’. Pada tabel tersebut, atribut ‘Model’ merujuk pada kandidat model peramalan terbaik (Tabel 6.13). Evaluasi klasifikasi untuk ketiga model terbaik dengan RMSE terkecil pada tahap peramalan dilakukan untuk mengetahui apakah model peramalan terbaik juga menghasilkan nilai evaluasi klasifikasi yang baik pula.

Tabel 6.17 Evaluasi Klasifikasi pada *Test-set* (Dataran Rendah)

Model	Akurasi	Sensitivitas	Spesifisitas	Rata-rata
1	90,9%	50,0%	100,0%	80,3%
2	83,3%	50,0%	88,5%	73,9%
3	87,5%	25,0%	96,4%	69,6%

Dari Tabel 6.17, dapat diketahui bahwa model pertama menghasilkan nilai paling tinggi pada ketiga faktor (akurasi, sensitivitas, spesifisitas) dengan nilai rata-rata sebesar 80.3%. Kemudian diikuti oleh model kedua dan ketiga dengan nilai rata-rata berturut-turut 73.9% dan 69.6%. Namun selain pengujian pada *test-set*, juga perlu dilakukan pengujian terhadap proporsi data lain dan kecamatan lainnya. Hasil pengujian tersebut dapat dilihat pada Tabel 6.18, dimana ‘mod’ menunjukkan urutan model terbaik, ‘acc’ menunjukkan rata-rata nilai akurasi yang dihasilkan pada pengujian, ‘sens’ menunjukkan rata-rata nilai sensitivitas yang dihasilkan pada pengujian, ‘spec’ menunjukkan rata-rata nilai spesifisitas yang dihasilkan pada pengujian, dan ‘avg’ menunjukkan nilai rata-rata keseluruhan pengukuran (akurasi, sensitivitas, dan spesifisitas). Penjabaran lengkap hasil pengujian yang dilakukan dilampirkan pada Lampiran C.

Tabel 6.18 Evaluasi Klasifikasi pada Uji Proporsi dan Uji Kecamatan Lain (Dataran Rendah)

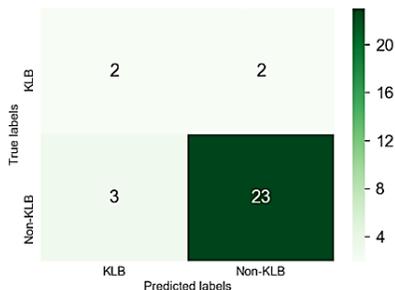
Mod	Uji Proporsi				Uji Kecamatan Lain			
	Acc	Sens	Spec	Avg	Acc	Sens	Spec	Avg
1	89,0%	64,5%	93,8%	82,4%	73,6%	20,9%	85,0%	59,8%
2	84,4%	64,5%	89,2%	79,4%	79,1%	33,4%	90,0%	67,5%
3	85,0%	53,6%	92,7%	77,1%	74,5%	30,7%	83,5%	62,9%

Dalam penelitian ini, model klasifikasi yang terbaik harus dipastikan tidak menghasilkan nilai 0% baik pada nilai sensitivitas maupun spesifisitas untuk tiap kecamatan. Karena, model yang menghasilkan nilai 0% baik pada sensitivitas maupun spesifisitas menggambarkan bahwa model tersebut tidak baik digunakan untuk memprediksi salah satu kelas, dalam hal ini KLB atau non-KLB. Seperti yang terlampir pada Lampiran C, Model 1 ternyata memiliki nilai sensitivitas sebesar 0% pada Kecamatan Bululawang, maka dari itu model 1 perlu dieliminasi dan tidak akan digunakan meskipun menghasilkan nilai evaluasi yang lebih tinggi.

Setelah tersisa 2 kandidat model terbaik yaitu model 2 dan model 3, maka model terbaik dapat dipilih diantara keduanya yang menghasilkan nilai evaluasi paling baik. Dari hasil pengujian pada Tabel 6.17, dapat diketahui bahwa model kedua menghasilkan nilai rata-rata evaluasi 79,4% untuk pengujian proporsi data dan sebesar 67,5% untuk pengujian pada kecamatan lain, dimana kedua nilai tersebut lebih tinggi dibandingkan dengan hasil evaluasi untuk model ketiga. Maka dari itu, dapat disimpulkan bahwa model yang paling baik untuk digunakan di dataran rendah adalah model yang kedua. Dalam penelitian ini, model akhir yang terpilih untuk digunakan disebut dengan model final.

Hasil klasifikasi dapat divisualisasikan melalui *confusion matrix*. *Confusion matrix* yang dibangun dengan model final untuk kecamatan pembentuk model di dataran rendah yaitu Kecamatan Kepanjen ditunjukkan pada Gambar 6.10, dimana

terdapat kelas biner ‘KLB’ dan ‘non-KLB’ sebagai label yang digunakan. *Confusion matrix* tersebut dibangun berdasarkan pengujian pada *test-set*.



Gambar 6.10 *Confusion Matrix* Kec. Kepanjen

Dari Gambar 6.10, dapat diketahui bahwa dari 4 data KLB, 2 diantaranya dapat diprediksi dengan benar sehingga nilai sensitivitas dari model yang dibangun bernilai 50%. Untuk kelas berlabel negatif, dari 26 data non-KLB, 23 diantaranya dapat diprediksi dengan benar sehingga nilai spesifisitas yang dihasilkan sebesar 88,5%. Sedangkan untuk nilai akurasi yang dihasilkan sebesar 83,33%.

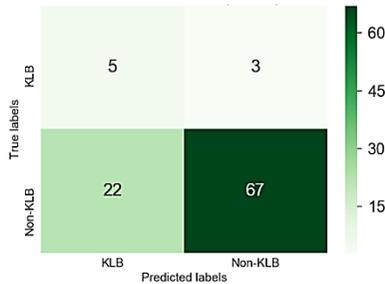
b. Hasil Klasifikasi pada Kecamatan Selain Pembentuk Model

Sub-bab ini menjelaskan tentang hasil klasifikasi pada kecamatan uji, yaitu kecamatan selain pembentuk model. Di dataran rendah, kecamatan yang berperan sebagai kecamatan uji adalah Kecamatan Turen, Kecamatan Gondanglegi, Kecamatan Donomulyo, dan Kecamatan Bululawang.

i. Kecamatan Turen

Gambar 6.11 menunjukkan hasil *confusion matrix* yang terbentuk untuk Kecamatan Turen menggunakan model terbaik. Dari *confusion matrix* tersebut dapat diketahui bahwa dari 8 data KLB pada data, model dapat memprediksi dengan benar 5 diantaranya. Sedangkan untuk label non-KLB, model dapat memprediksi dengan benar 67 data dari keseluruhan 89 data

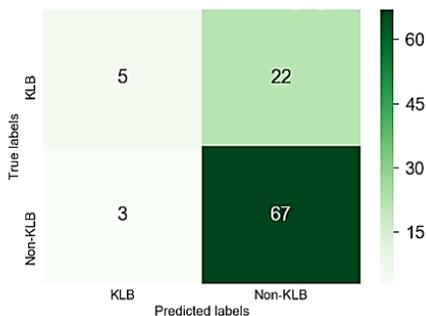
yang berlabel non-KLB. Pada kecamatan Turen, nilai sensitivitas, spesifisitas, dan akurasi yang dihasilkan berturut-turut sebesar 62.5% , 75,28%, dan 74,23%.



Gambar 6.11 *Confusion Matrix* Kec. Turen

ii. Kecamatan Gondanglegi

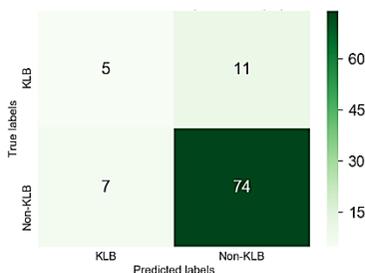
Gambar 6.12 menunjukkan hasil *confusion matrix* yang terbentuk untuk Kecamatan Gondanglegi. Dari *confusion matrix* tersebut dapat diketahui bahwa dari 27 data KLB pada data, model hanya dapat memprediksi dengan benar 5 diantaranya. Sedangkan untuk label non-KLB, model dapat memprediksi dengan benar 67 data dari keseluruhan 70 data yang berlabel non-KLB. Pada Kecamatan Gondanglegi, nilai sensitivitas, spesifisitas, dan akurasi yang dihasilkan berturut-turut sebesar 18.52%, 95.71%, dan 74.23%.



Gambar 6.12 *Confusion Matrix* Kec. Gondanglegi

iii. Kecamatan Donomulyo

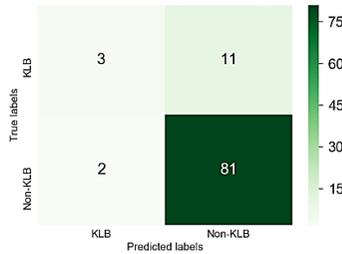
Gambar 6.13 menunjukkan hasil *confusion matrix* yang terbentuk untuk Kecamatan Donomulyo. Dari *confusion matrix* tersebut dapat diketahui bahwa dari 16 data KLB pada data, model hanya dapat memprediksi dengan benar 5 diantaranya. Sedangkan untuk label non-KLB, model dapat memprediksi dengan benar 74 data dari keseluruhan 81 data yang berlabel non-KLB. Pada Kecamatan Donomulyo, nilai sensitivitas, spesifisitas, dan akurasi yang dihasilkan berturut-turut sebesar 31.25%, 91.36% dan 81.44%.



Gambar 6.13 Confusion Matrix Kec. Donomulyo

iv. Kecamatan Bululawang

Gambar 6.14 menunjukkan hasil *confusion matrix* yang terbentuk untuk Kecamatan Bululawang. Dari *confusion matrix* tersebut dapat diketahui bahwa dari 14 data KLB pada data, model hanya dapat memprediksi dengan benar 3 diantaranya. Sedangkan untuk label non-KLB, model dapat memprediksi dengan benar 81 data dari keseluruhan 83 data yang berlabel non-KLB. Pada Kecamatan Bululawang, nilai sensitivitas, spesifisitas, dan akurasi yang dihasilkan berturut-turut sebesar 21.43%, 97.59% dan 86.60%.



Gambar 6.14 Confusion Matrix Kec. Bululawang

Meskipun memiliki nilai sensitivitas yang tidak terlalu tinggi, Kecamatan Bululawang merupakan kecamatan yang memiliki nilai akurasi yang paling tinggi untuk kecamatan uji di dataran rendah. Hal ini dapat terjadi karena prediksi salah keseluruhan yang dihasilkan oleh model di Kecamatan Bululawang hanya sebanyak 13 data dari keseluruhan 97 data.

6.2.3. Hasil Perhitungan 24 Periode Kedepan pada Dataran Rendah

Pada subbab ini dijelaskan hasil perhitungan 24 periode mendatang yakni periode Januari 2019 hingga Desember 2020 pada dataran rendah. Sub-bab ini dibagi menjadi dua bagian yaitu bagian hasil tahap peramalan dan hasil tahap klasifikasi. Hasil peramalan jumlah kejadian demam berdarah disertai label KLB pada Kecamatan Kepanjen ditunjukkan pada Tabel 6.19. Sedangkan hasil peramalan 24 periode mendatang kecamatan lainnya di dataran rendah dijabarkan pada Lampiran C.

Model yang digunakan pada tahap ini adalah model terbaik yang telah terpilih untuk keseluruhan proses (model final) di dataran rendah, yaitu dengan skenario pembagian data 70-30, dengan menggunakan variabel *input* jumlah penduduk dan kelembapan udara dengan tambahan fitur *lag* berjumlah 11. Sedangkan parameter XGBoost hasil *tuning* yang digunakan pada model ini yaitu ‘*colsample_bytree*’ sebesar 0.5, ‘*learning_rate*’ sebesar 0.5, ‘*gamma*’ sebesar 0.01, ‘*max_depth*’ sebesar 8, ‘*min_child_weight*’ sebesar 1, ‘*n_estimator*’ sebesar 50, ‘*reg_alpha*’ sebesar 0, dan ‘*subsample*’ sebesar 1.

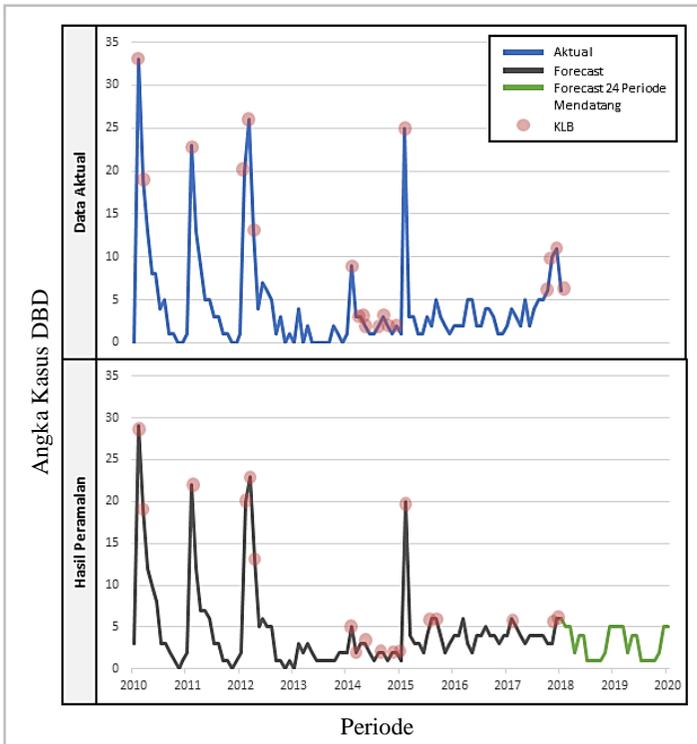
Tabel 6.19 Hasil Peramalan 2 Tahun Mendatang pada Kec. Kepanjen

Periode	Jumlah KDB	Status	Periode	Jumlah KDB	Status
Jan-19	5	non-KLB	Jan-20	5	non-KLB
Feb-19	5	non-KLB	Feb-20	5	non-KLB
Mar-19	2	non-KLB	Mar-20	2	non-KLB
Apr-19	4	non-KLB	Apr-20	4	non-KLB
Mei-19	4	non-KLB	Mei-20	4	non-KLB
Jun-19	1	non-KLB	Jun-20	1	non-KLB
Jul-19	1	non-KLB	Jul-20	1	non-KLB
Agt-19	1	non-KLB	Agt-20	1	non-KLB
Sep-19	1	non-KLB	Sep-20	1	non-KLB
Okt-19	2	non-KLB	Okt-20	2	non-KLB
Nov-19	5	non-KLB	Nov-20	5	non-KLB
Des-19	5	non-KLB	Des-20	5	non-KLB

Setelah mendapatkan data hasil peramalan untuk 24 periode mendatang, selanjutnya dapat dilakukan perbandingan data aktual angka kejadian demam berdarah dan hasil peramalan yang sudah berlabel KLB seperti yang ditunjukkan pada Gambar 6.15. Dari gambar tersebut, dapat dilihat bahwa secara keseluruhan model yang terpilih di dataran rendah dapat menghasilkan pola data yang serupa dengan data aktual. Hal ini berdampak pada label KLB yang dapat diprediksi dengan cukup baik sehingga menghasilkan nilai akurasi yang tinggi yakni sebesar 83,33% seperti yang telah dijelaskan pada subbab 6.2.2.

Dari plot yang ada pada Gambar 6.15, ada beberapa titik KLB pada jumlah kasus yang tidak terlalu tinggi, contohnya pada data aktual tahun 2015. Merujuk pada aturan pemerintah terkait KLB seperti dicantumkan pada sub-bab 2.2.1, hal ini dapat terjadi karena nilai rata-rata angka kejadian di tahun sebelumnya berada pada angka yang kecil. Rata-rata kejadian pada tahun 2014 adalah 1. Dimana yang berarti angka kejadian lebih dari 2 pada tahun 2015 akan terdeteksi sebagai KLB. Hal ini juga didukung dengan adanya faktor-faktor lain yang dapat menyebabkan suatu daerah dalam suatu kurun waktu dapat digolongkan sebagai KLB, yakni faktor politis dan faktor angka kematian. Jadi, meskipun angka kejadiannya kecil, namun

kepala daerah menetakannya menjadi KLB maka hal tersebut bisa saja terjadi. Selain itu, jika angka kejadian yang kecil, namun memiliki prosentase kematian yang tinggi, juga tetap bisa digolongkan sebagai KLB.



Gambar 6.15 Perbandingan Data Aktual dan Hasil Peramalan di Dataran Rendah (Kec. Kapanjen)

6.3. Hasil Peramalan dan Klasifikasi Dataran Sedang

Pada subbab ini dijelaskan hasil peramalan pada dataran sedang menggunakan metode XGBoost yang dilanjutkan dengan Klasifikasi Berbasis Aturan untuk menentukan status KLB dan non-KLB pada data. Sub-bab ini terdiri dari tiga bagian yaitu

hasil peramalan, hasil klasifikasi, dan hasil perhitungan 24 periode mendatang.

6.3.1. Hasil Peramalan pada Dataran Sedang

Pada subbab ini pembahasan difokuskan pada hasil peramalan yang dilakukan dengan metode XGBoost di dataran sedang. Sub-bab ini terdiri dari dua bagian yaitu hasil peramalan pada kecamatan pembentuk model, dan hasil peramalan pada kecamatan selain pembentuk model.

a. Hasil Uji Coba Skenario Peramalan

Pada proses pembentukan model peramalan di dataran sedang, terdapat beberapa skenario yang dilakukan dan masing-masing skenario menghasilkan satu model. Dari keseluruhan model yang telah dibangun tersebut kemudian dilakukan pengujian dengan membandingkan nilai RMSE tiap model.

Data yang digunakan untuk pembentukan model pada dataran sedang yaitu data Kecamatan Pakisaji dengan menggunakan *input* sesuai dengan skenario yang telah dijelaskan pada subbab 4.1.2 bagian F. Hasil dari pengujian model pada *test-set* Kecamatan Pakisaji ditunjukkan pada Tabel 6.16. Sedangkan hasil peramalan lengkap dengan informasi parameter yang digunakan dijelaskan pada Lampiran E.

Pada Tabel 6.20, dapat diketahui nilai RMSE untuk tiap kombinasi skenario pembagian proporsi data dan pemilihan fitur diikuti dengan jumlah *lag* optimal pada skenario tersebut. Dari model-model yang dibangun, model terbaik didapatkan dari urutan nilai RMSE yang terkecil. Terdapat tiga model dengan RMSE terbaik seperti yang tertera pada Tabel 6.20, yaitu yang pertama pada skenario pembagian data 80-20 dan menggunakan *input* berupa fitur *lag* berjumlah 3 dengan nilai RMSE sebesar 2,6990. Kemudian model yang kedua yakni dengan pembagian data 80-20 yang melibatkan variabel *input* Angka Bebas Jentik dan fitur *lag* berjumlah 3 dengan nilai RMSE sebesar 2,8799. Model terbaik ketiga pada dataran sedang yaitu model dengan skenario pembagian data 70-30 dan

variabel *input* berupa fitur *lag* berjumlah 3 dengan nilai RMSE sebesar 2,8989. Ketiga model tersebut memiliki nilai SMAPE berturut-turut, 35,62%, 43,03%, dan 35,99%.

Tabel 6.20 Hasil RMSE untuk *test-set* di Kec. Pakisaji

Skenario			RMSE
Pembagian Data	Pemilihan Fitur	Windowing	
80-20	ALL-VI	3	3,6561
	TOP3-VI	3	3,1972
	TOP2-VI	3	3,1845
	TOP1-VI	2	2,8799
	NO-VI	3	2,6990
70-30	ALL-VI	3	4,0075
	TOP3-VI	2	3,2115
	TOP2-VI	3	3,4743
	TOP1-VI	2	3,1524
	NO-VI	3	2,8989
60-40	ALL-VI	1	5,3667
	TOP3-VI	2	5,2018
	TOP2-VI	3	4,4277
	TOP1-VI	3	4,3621
	NO-VI	1	4,9789
50-50	ALL-VI	1	5,3782
	TOP3-VI	11	5,1838
	TOP2-VI	3	4,5323
	TOP1-VI	3	4,5773
	NO-VI	4	5,1648

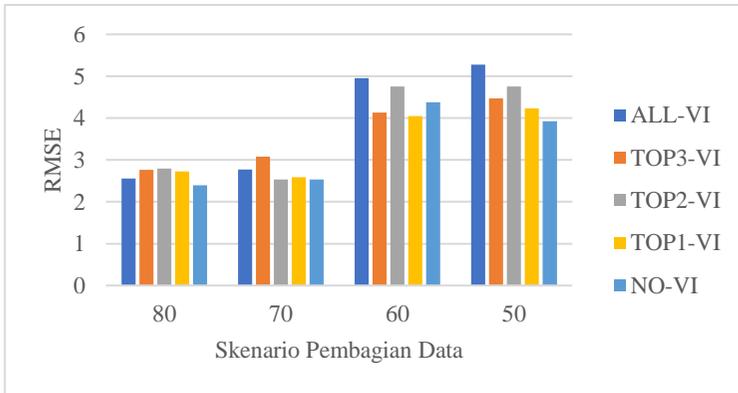
Tabel 6.21 Hasil *Tuning* Parameter di Dataran Sedang

Model	Skenario	Parameter XGBoost							
		col	lr	gm	md	mcw	n_est	alpha	sub
1	80-20	0.5	0.1	0.01	2	4	50	0	0.6
	NO-VI								
	Lag 3								
2	80-20	0.5	0.5	0.01	4	1	50	0	0.8
	TOP1-VI								
	Lag 2								
3	70-30	0.5	0.1	0	2	4	50	0	0.6
	NO-VI								
	Lag 3								

Dari hasil *tuning* parameter, dapat diketahui bahwa ketiga model dengan RMSE terkecil menggunakan parameter yang sama untuk rasio fitur yang digunakan dalam membentuk model pada parameter *colsample_bytree* sebesar 0.5, dan nilai regularisasi *reg_alpha* sebesar 1, dan jumlah estimator sebanyak 50. Hal ini berarti model hanya menggunakan separuh dari keseluruhan fitur yang ada untuk membangun model tiap iterasinya dengan jumlah iterasi *boosting* sebanyak 50 kali. Model pertama dan ketiga memiliki kesamaan pada empat parameter yakni nilai *learning_rate* sebesar 0.1, *max_depth* sebesar 2, *min_child_weight* sebesar 4 dan *subsample* sebesar 0.6. Sedangkan pada model kedua, nilai parameter *learning_rate*, *max_depth*, *min_child_weight*, dan *subsample* berturut-turut sebesar 0.5, 4, 1, dan 0.8. Untuk nilai *gamma*, model pertama dan kedua memiliki kesamaan nilai yaitu sebesar 0.01, sedangkan pada model ketiga nilai *gamma* yg digunakan yaitu 0. Informasi hasil *tuning* parameter lebih lengkap terdapat pada Lampiran E.

Dari hasil pengujian yang telah dilakukan, selain informasi mengenai model terbaik, dapat diketahui pula pola persebaran nilai RMSE yang dihasilkan. Persebaran RMSE berdasarkan skenario pembagian data di dataran sedang ditunjukkan pada Gambar 6.15. Dari gambar tersebut, dapat diketahui bahwa RMSE yang dihasilkan untuk pembagian data 80-20 dan 70-30 relatif lebih rendah dibandingkan RMSE yang dihasilkan dari pembagian data 60-40 dan 50-50. Hal ini disebabkan oleh

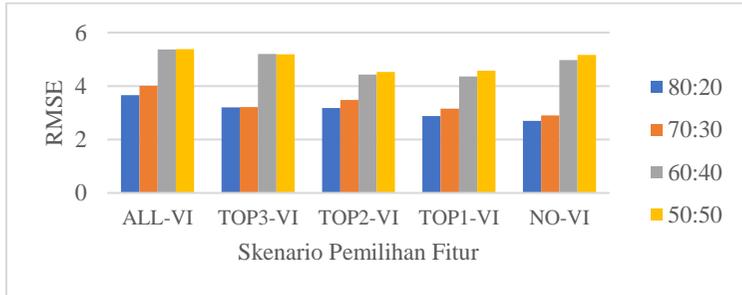
semakin sedikitnya proporsi data pelatihan, maka semakin sedikit pula data yang menjadi *input* dalam pembentukan model. Hal tersebut dapat menyebabkan model yang dihasilkan *underfit*, yakni model tidak bisa melihat logika di belakang data, sehingga tidak bisa melakukan peramalan secara tepat dan memiliki nilai *error* yang relatif tinggi.



Gambar 6.16 Persebaran RMSE berdasarkan Skenario Pembagian Data (Dataran Sedang)

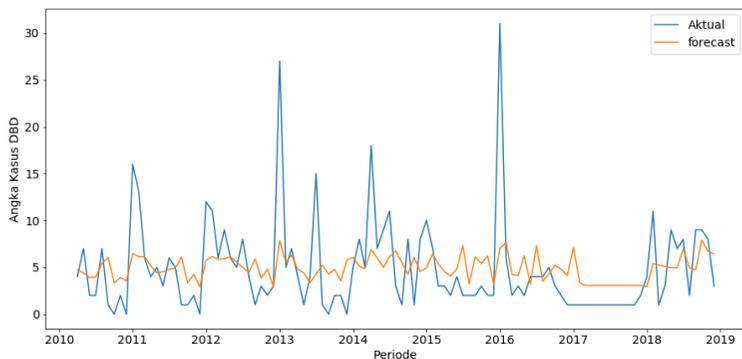
Gambar 6.17 merupakan grafik yang menunjukkan persebaran RMSE berdasarkan skenario pemilihan fitur. Dari grafik tersebut dapat diketahui bahwa RMSE yang dihasilkan pada masing-masing skenario pemilihan fitur terbagi menjadi 2 *cluster* data dengan pola yang serupa, yakni pada pembagian data 50-50 dengan 60-40 dan 80-20 dengan 70-30. Untuk proporsi data 60-40 dan 50-50, RMSE paling rendah dihasilkan oleh model yang dibangun menggunakan variabel Angka Bebas Jentik dan curah hujan, atau Angka Bebas Jentik saja ditambah dengan fitur *lag* yang optimal. Pada proporsi data tersebut, menggunakan lebih dari 3 variabel independen utama atau tidak melibatkan variabel independen utama sama sekali akan menghasilkan model dengan nilai RMSE yang lebih tinggi. Sedangkan untuk proporsi data 70-30 dan 80-20, nilai RMSE menunjukkan tren menurun seiring dengan berkurangnya jumlah variabel independen utama yang digunakan. Pada

proporsi data tersebut, model yang dibangun dengan hanya melibatkan fitur *lag* sebagai *input* menghasilkan nilai RMSE yang paling kecil.



Gambar 6.17 Persebaran RMSE berdasarkan Skenario Pemilihan Fitur (Dataran Sedang)

Perbandingan data aktual dan hasil peramalan yang dilakukan pada kecamatan pembentuk model yaitu Kecamatan Pakisaji dengan menggunakan model peramalan dengan RMSE terkecil tertera pada Gambar 6.18. Dapat dilihat pada gambar tersebut bahwa model dapat menangkap pola data dengan cukup baik, namun kurang berhasil dalam memprediksi lonjakan-lonjakan angka kejadian demam berdarah yang terjadi pada Kecamatan Pakisaji. Hal ini diperkuat dengan hasil perhitungan RMSE yaitu sebesar 2,6990 dan SMAPE sebesar 35,62%.



Gambar 6.18 Perbandingan Data Aktual dan Hasil Peramalan Kec. Pakisaji

Selain nilai RMSE pengujian pada *test-set*, tingkat *robustness* model juga perlu diperhitungkan untuk mengetahui seberapa baik model diterapkan pada data lainnya. Salah satu cara mengetahui tingkat *robustness* model yaitu dengan menguji model tersebut pada proporsi data yang berbeda. Hasil uji proporsional data untuk tiga model dengan nilai RMSE terbaik tertera pada Tabel 6.22. Hasil lengkap perbandingan model dapat dilihat pada Lampiran E.

Tabel 6.22 Hasil Uji Proporsional Data pada Kec. Pakisaji

Model Terbaik ke-	RMSE Uji Proporsional Data			Mean	St.dev
	proporsi 1	proporsi 2	proporsi 3		
1	2,8117	4,6545	4,4192	3,9618	0,8189
2	2,5576	2,3564	2,3070	2,4070	0,1084
3	2,5738	4,5603	4,3451	3,8264	0,8901

Seperti yang tertera pada Tabel 6.22, terdapat tiga uji proporsi yang dilakukan, skenario uji tersebut telah dijelaskan pada subbab 4.2.3. Dari hasil uji proporsi yang didapat kemudian dihitung nilai rata-rata dan standar deviasinya agar dapat diketahui mana model yang memiliki tingkat *robustness* paling tinggi. Mengacu pada Tabel 6.22, dua atribut yang digunakan untuk menunjang perhitungan tersebut yakni atribut ‘Mean’ yang menunjukkan nilai rata-rata hasil uji proporsi, sedangkan ‘St.dev’ yang menunjukkan nilai standar deviasi yang dihasilkan. Dapat diketahui dari hasil uji proporsi tersebut bahwa model terbaik kedua ternyata menghasilkan nilai RMSE yang paling kecil diantara dua model lainnya baik secara rata-rata maupun nilai standar deviasi RMSE dengan nilai rata-rata RMSE sebesar 2,4070 dan nilai standar deviasi RMSE sebesar 0,1084. Hal ini menunjukkan bahwa meskipun suatu model memiliki nilai RMSE paling kecil pada pengujian *test-set*, belum tentu model tersebut juga baik digunakan pada proporsi data yang lain. Dalam kasus ini, model terbaik (model dengan RMSE terkecil) ternyata memiliki tingkat *robustness* yang tidak

terlalu baik untuk proporsi data yang lain, sedangkan model yang bukan terbaik justru menghasilkan nilai *robustness* yang lebih tinggi terhadap proporsi data yang lain.

b. Hasil Peramalan pada Kecamatan Selain Pembentuk Model

Selain pengujian yang dilakukan pada proporsi pembagian data yang berbeda, model perlu diuji terhadap data kecamatan lain pada dataran yang sama. Di dataran sedang, kecamatan uji terdiri dari Kecamatan Dampit, Kecamatan Singosari, Kecamatan Tumpang, Kecamatan Lawang, Kecamatan Sumbermanjing, Kecamatan Wajak, dan Kecamatan Karangploso. Hasil perbandingan nilai RMSE yang dihasilkan pada tiap kecamatan uji untuk tiga model peramalan terbaik ditunjukkan pada Tabel 6.23. Hasil lengkap perbandingan RMSE uji kecamatan lain terlampir pada Lampiran E.

Tabel 6.23 Perbandingan Hasil RMSE pada Dataran Sedang

		Model Terbaik ke-		
		1	2	3
RMSE Uji Kecamatan Lain	Dampit	4,4113	6,0647	4,8480
	Singosari	2,8735	3,5180	3,0919
	Tumpang	4,6499	5,3434	4,6742
	Lawang	7,2460	8,8865	7,4098
	Sumbermanjing	8,2862	13,1352	9,4631
	Wajak	3,3914	3,6593	3,4150
	Karangploso	3,2199	4,1160	3,3684
Mean		4,8683	6,3890	5,1815
St.dev		1,9446	3,2471	2,2181

Dari hasil pengujian terhadap kecamatan lain, dapat diketahui bahwa model pertama menghasilkan nilai rata-rata dan standar deviasi RMSE terkecil dengan nilai rata-rata sebesar 4,8683, kemudian diikuti oleh model ketiga dengan nilai rata-rata RMSE sebesar 5,1815, lalu model kedua sebesar 6,3890. Maka dapat diartikan bahwa untuk kasus di dataran sedang, model dengan RMSE terkecil pada pengujian *test-set* juga

menunjukkan performa yang paling baik pula jika diterapkan pada kecamatan-kecamatan lainnya.

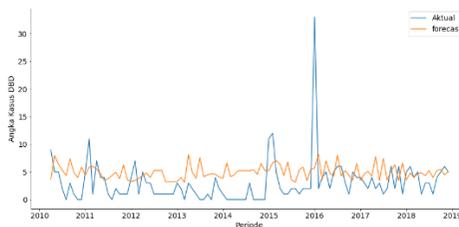
Selain perhitungan nilai RMSE, pengukuran lain yang digunakan untuk menunjang analisis permalan yakni nilai SMAPE atau rata-rata persentase kesalahan absolut simetris. Nilai SMAPE yang dihasilkan model peramalan terbaik di dataran sedang pada pengujian kecamatan selain pembentuk model ditunjukkan pada Tabel 6.24.

Tabel 6.24 Hasil Nilai SMAPE Kecamatan Lain (Dataran Sedang)

No	Kecamatan	SMAPE
1	Dampit	46,2%
2	Singosari	49,97%
3	Tumpang	34,94%
4	Lawang	43,60%
5	Sumbermanjing	59,65%
6	Wajak	46,71%
7	Karangploso	41,51%

i. Kecamatan Dampit

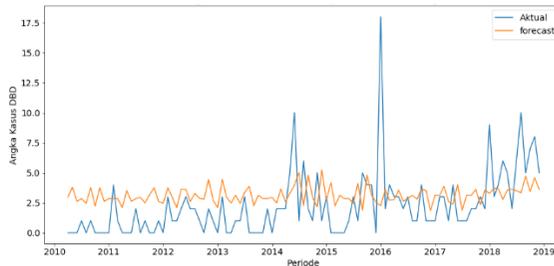
Perbandingan jumlah kasus demam berdarah aktual dan hasil peramalan di Kecamatan Dampit ditunjukkan pada Gambar 6.19. Dari gambar tersebut dapat diketahui bahwa model kurang baik dalam menangkap pola kenaikan dan penurunan jumlah kasus demam berdarah. Hal ini dapat dilihat pula dari nilai RMSE yang cukup tinggi yaitu sebesar 4,41 dan SMAPE sebesar 46,24%.



Gambar 6.19 Perbandingan Data Aktual dan Hasil Peramalan Kec. Dampit

ii. Kecamatan Singosari

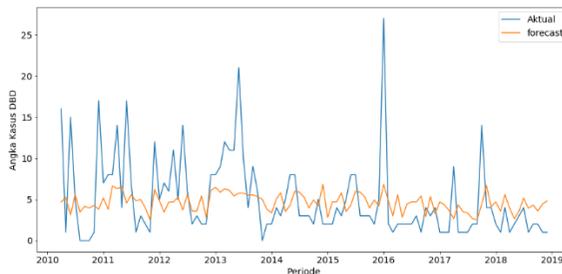
Perbandingan jumlah kasus demam berdarah aktual dan hasil peramalan di Kecamatan Singosari ditunjukkan pada Gambar 6.20. Dari gambar tersebut dapat diketahui bahwa kurang cukup baik dalam menangkap pola lonjakan jumlah kasus demam berdarah. Hal ini dapat dilihat pula dari nilai RMSE sebesar 2,873 dan SMAPE sebesar 49,97%.



Gambar 6.20 Perbandingan Data Aktual dan Hasil Peramalan Kec. Singosari

iii. Kecamatan Tumpang

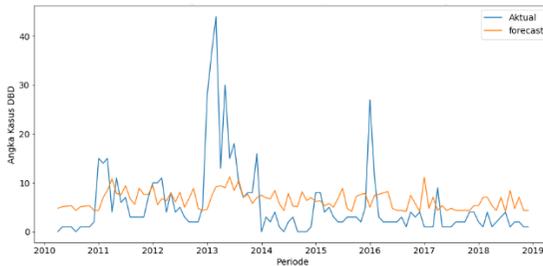
Perbandingan jumlah kasus demam berdarah aktual dan hasil peramalan di Kecamatan Tumpang ditunjukkan pada Gambar 6.21. Dari gambar tersebut dapat diketahui bahwa kurang cukup baik dalam menangkap pola lonjakan jumlah kasus demam berdarah. Hal ini dapat dilihat pula dari nilai RMSE sebesar 4,6499 dan SMAPE sebesar 34,94%.



Gambar 6.21 Perbandingan Data Aktual dan Hasil Peramalan Kec. Tumpang

iv. Kecamatan Lawang

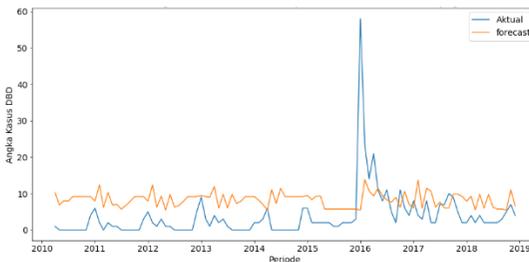
Perbandingan jumlah kasus demam berdarah aktual dan hasil peramalan di Kecamatan Lawang ditunjukkan pada Gambar 6.22. Dari gambar tersebut dapat diketahui bahwa kurang cukup baik dalam menangkap pola data jumlah kasus demam berdarah. Hal ini dapat dilihat pula dari nilai RMSE yang cukup tinggi yaitu sebesar 7,246 dan SMAPE sebesar 43,60%.



Gambar 6.22 Perbandingan Data Aktual dan Hasil Peramalan Kec. Lawang

v. Kecamatan Sumbermanjing

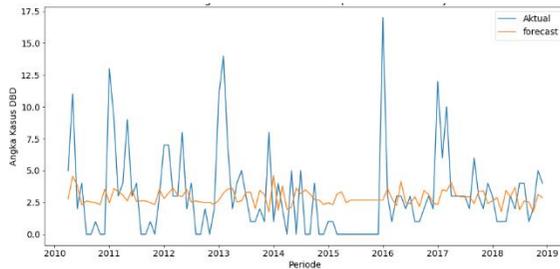
Perbandingan jumlah kasus demam berdarah aktual dan hasil peramalan di Kecamatan Sumbermanjing ditunjukkan pada Gambar 6.23. Dari gambar tersebut dapat diketahui bahwa kurang cukup baik dalam menangkap lonjakan jumlah kasus demam berdarah. Hal ini dapat dilihat pula dari nilai RMSE yang cukup tinggi yaitu sebesar 8,286 dan SMAPE sebesar 59,65%.



Gambar 6.23 Perbandingan Data Aktual dan Hasil Peramalan Kec. Sumbermanjing

vi. Kecamatan Wajak

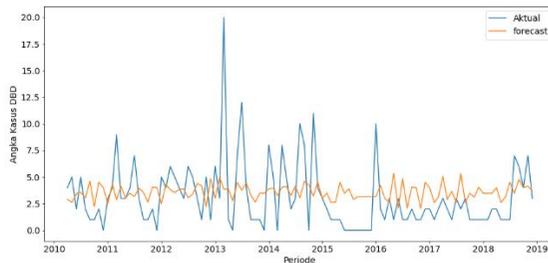
Perbandingan jumlah kasus demam berdarah aktual dan hasil peramalan di Kecamatan Wajak ditunjukkan pada Gambar 6.24. Dari gambar tersebut dapat diketahui bahwa kurang baik dalam menangkap pola data jumlah kasus demam berdarah. Hal ini dapat dilihat pula dari nilai RMSE yang cukup tinggi yaitu sebesar 3,391 dan SMAPE sebesar 46,71%.



Gambar 6.24 Perbandingan Data Aktual dan Hasil Peramalan Kec. Wajak

vii. Kecamatan Karangploso

Perbandingan jumlah kasus demam berdarah aktual dan hasil peramalan di Kecamatan Karangploso ditunjukkan pada Gambar 6.25. Dari gambar tersebut dapat diketahui bahwa kurang baik dalam menangkap pola data jumlah kasus demam berdarah. Hal ini dapat dilihat pula dari nilai RMSE yang cukup tinggi yaitu sebesar 3,219 dan SMAPE sebesar 41,51%.



Gambar 6.25 Perbandingan Data Aktual dan Hasil Peramalan Kec. Karangploso

6.3.2. Hasil Klasifikasi pada Dataran Sedang

Pada subbab ini pembahasan difokuskan pada hasil klasifikasi yang dilakukan dengan metode klasifikasi berbasis aturan di dataran rendah. Sub-bab ini terdiri dari dua bagian yaitu hasil klasifikasi pada kecamatan pembentuk model dan hasil klasifikasi pada kecamatan selain pembentuk model.

a. Hasil Uji Coba Skenario Klasifikasi

Dalam penelitian kali ini, pengukuran evaluasi klasifikasi biner yang digunakan adalah sensitivitas, spesifisitas, dan akurasi. Klasifikasi biner yang baik akan menghasilkan nilai yang tinggi untuk ketiga faktor tersebut, sedangkan klasifikasi biner yang buruk akan menghasilkan nilai rendah pada faktor-faktor tersebut. Agar memastikan ketiga faktor tersebut baik pada suatu model, dilakukan perataan pada nilai akurasi, sensitivitas, dan spesifisitas seperti yang tertera pada Tabel 6.25 dengan atribut 'Rata-rata'. Pada tabel tersebut, atribut 'Model' merujuk pada kandidat model peramalan terbaik pada Tabel 6.22. Evaluasi klasifikasi untuk ketiga model tersebut dilakukan untuk mengetahui apakah model peramalan terbaik juga menghasilkan nilai evaluasi klasifikasi yang baik pula.

Tabel 6.25 Evaluasi Klasifikasi pada *Test-set* (Dataran Sedang)

Model	Akurasi	Sensitivitas	Spesifisitas	Rata-rata
1	95,24%	100%	90%	95,08%
2	90,91%	90,91%	90,91%	90,91%
3	93,75%	100%	90,48%	94,74%

Dari Tabel 6.25, dapat diketahui bahwa model pertama menghasilkan nilai paling tinggi pada ketiga faktor (akurasi, sensitivitas, spesifisitas) dengan nilai rata-rata sebesar 95,08%. Kemudian diikuti oleh model ketiga dan kedua dengan nilai rata-rata berturut-turut 94,74% dan 90,91%. Namun selain pengujian pada data *test-set*, perlu juga dilakukan pengujian terhadap proporsi data lain dan kecamatan lainnya. Hasil

pengujian tersebut dapat dilihat pada Tabel 6.26, dimana ‘mod’ menunjukkan urutan model terbaik, ‘acc’ menunjukkan rata-rata nilai akurasi yang dihasilkan pada pengujian, ‘sens’ menunjukkan rata-rata nilai sensitivitas yang dihasilkan pada pengujian, ‘spec’ menunjukkan rata-rata nilai spesifisitas yang dihasilkan pada pengujian, dan ‘avg’ menunjukkan nilai rata-rata keseluruhan pengukuran (akurasi, sensitivitas, dan spesifisitas). Penjabaran lengkap hasil pengujian yang dilakukan dilampirkan pada Lampiran C.

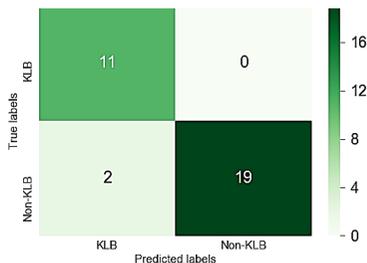
Tabel 6.26 Evalausi Klasifikasi pada Uji Proporsi dan Uji Kecamatan Lain (Dataran Sedang)

Mod	Uji Proporsi				Uji Kecamatan Lain			
	Acc	Sens	Spec	Avg	Acc	Sens	Spec	Avg
1	94,7%	94,4%	94,6%	94,6%	72,8%	59,1%	73,6%	68,5%
2	95,1%	91,4%	96,5%	94,3%	73,2%	51,7%	76,6%	67,2%
3	95,6%	100%	92,8%	96,1%	69,1%	60,4%	68,5%	66,0%

Mengacu pada Tabel 6.26, model dengan tingkat *robustness* paling tinggi adalah pada model 1 karena model tersebut menghasilkan nilai rata-rata akurasi, sensitivitas, dan spesifisitas yang paling tinggi diantara model lainnya. Namun dalam penelitian ini ditentukan bahwa model klasifikasi yang terbaik harus dipastikan tidak menghasilkan nilai 0% baik pada nilai sensitivitas maupun spesifisitas untuk tiap kecamatan. Karena, model yang menghasilkan nilai 0% baik pada sensitivitas maupun spesifisitas menggambarkan bahwa model tersebut tidak baik digunakan untuk memprediksi salah satu kelas, dalam hal ini KLB atau non-KLB. Terlebih, peramalan KLB yang dideteksi sebagai non-KLB akan sangat berbahaya, sehingga nilai sensitivitas 0% sangat tidak bisa diterima. Seperti yang telah dilampirkan pada Lampiran C, Model 1 memiliki nilai sensitivitas sebesar 0% pada Kecamatan Tumpang, maka dari itu model 1 dieliminasi dan tidak akan digunakan meskipun menghasilkan nilai evaluasi yang cukup tinggi.

Sesuai hasil yang tertera pada Tabel 6.25, nilai rata-rata indikator klasifikasi untuk model ketiga lebih baik daripada model kedua yaitu dengan selisih hingga 4,1% dan juga memiliki tingkat *robustness* yang cukup baik untuk proporsi data lain. Maka dari itu, dapat disimpulkan bahwa model yang terbaik untuk digunakan pada tahap klasifikasi di dataran sedang (model final) adalah model yang ketiga dengan skenario pembagian data 70-30 dan menggunakan *input* berupa fitur *lag* berjumlah 3 seperti yang telah dijelaskan pada Tabel 6.21.

Hasil klasifikasi dapat divisualisasikan melalui *confusion matrix*. *Confusion matrix* yang dibangun dengan model final untuk Kecamatan Pakisaji pada *test-set* ditunjukkan pada Gambar 6.26, dimana terdapat kelas biner ‘KLB’ dan ‘non-KLB’ sebagai label yang digunakan. Dari *confusion matrix* yang dibangun, dapat diketahui bahwa dari 11 data KLB, semua data KLB dapat diprediksi dengan benar sehingga nilai sensitivitas dari model yang dibangun bernilai 100%. Untuk kelas berlabel negatif, dari 21 data non-KLB, 19 diantaranya dapat diprediksi dengan benar sehingga nilai spesifisitas yang dihasilkan sebesar 90,48%. Sedangkan untuk nilai akurasi yang dihasilkan sebesar 93,75%.



Gambar 6.26 Confusion Matrix Kec. Pakisaji

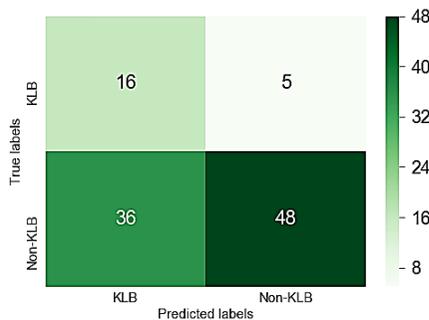
b. Hasil Klasifikasi pada Kecamatan Selain Pembentuk Model

Sub-bab ini menjelaskan tentang hasil klasifikasi pada kecamatan uji, yaitu kecamatan selain pembentuk model. Di dataran sedang, kecamatan yang berperan sebagai kecamatan uji adalah Kecamatan Dampit, Kecamatan Singosari, Kecamatan

Tumpang, Kecamatan Lawang, Kecamatan Sumbermanjing, Kecamatan Wajak, dan Kecamatan Karangploso.

i. Kecamatan Dampit

Gambar 6.27 menunjukkan hasil *confusion matrix* yang dibangun pada Kecamatan Dampit. Dari *confusion matrix* tersebut dapat diketahui bahwa dari 21 data KLB pada data, model dapat memprediksi dengan benar 16 diantaranya sehingga nilai sensitivitas yang didapat yakni 76,2%. Sedangkan untuk label non-KLB, model dapat memprediksi dengan benar 48 data dari keseluruhan 84 data yang berlabel non-KLB sehingga nilai spesifisitas yang didapat yaitu sebesar 57,1%. Dari keseluruhan kelas yang diprediksi, klasifikasi pada Kecamatan Dampit dapat menghasilkan nilai akurasi sebesar 61,0%.

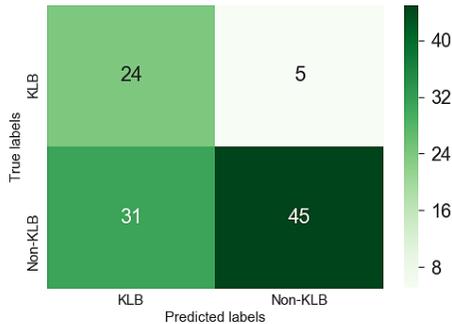


Gambar 6.27 *Confusion Matrix* Kec. Dampit

ii. Kecamatan Singosari

Gambar 6.28 menunjukkan hasil *confusion matrix* yang dibangun pada Kecamatan Singosari. Dari *confusion matrix* tersebut dapat diketahui bahwa dari 29 data KLB pada data, model dapat memprediksi dengan benar 24 diantaranya sehingga nilai sensitivitas yang didapat yakni 82,8%. Sedangkan untuk label non-KLB, model dapat memprediksi dengan benar 45 data dari keseluruhan 76 data yang berlabel non-KLB sehingga nilai spesifisitas yang didapat yaitu sebesar

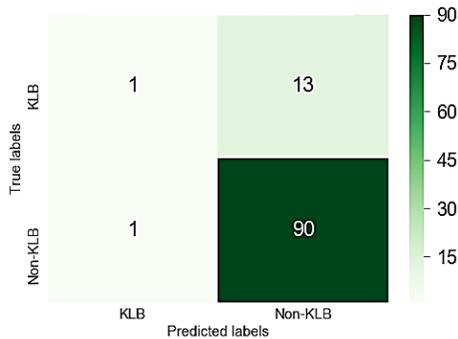
59,2%. Dari keseluruhan kelas yang diprediksi, klasifikasi pada Kecamatan Singosari dapat menghasilkan nilai akurasi sebesar 65,7%.



Gambar 6.28 *Confusion Matrix* Kec. Singosari

iii. Kecamatan Tumpang

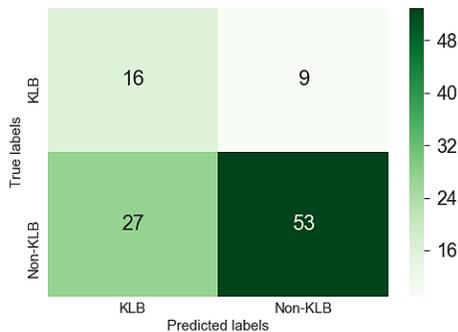
Gambar 6.29 menunjukkan hasil *confusion matrix* yang dibangun pada Kecamatan Tumpang. Dari *confusion matrix* tersebut dapat diketahui bahwa dari 14 data KLB pada data, model dapat memprediksi dengan benar 1 diantaranya sehingga nilai sensitivitas yang didapat yakni 7,1%. Angka sensitivitas tersebut merupakan yang terkecil diantara kecamatan uji lainnya di dataran sedang, hal ini disebabkan model yang dibangun kurang dapat menangkap pola data pada Kecamatan Tumpang. Sedangkan pada label non-KLB, model dapat memprediksi dengan benar 90 data dari keseluruhan 91 data yang berlabel non-KLB sehingga nilai spesifisitas yang didapat yaitu sebesar 98,9%. Hal ini menunjukkan bahwa klasifikasi yang dilakukan pada Kecamatan Tumpang baik pada kelas non-KLB, namun kurang baik pada kelas KLB. Dari keseluruhan kelas yang diprediksi, klasifikasi pada Kecamatan Singosari dapat menghasilkan nilai akurasi sebesar 86,7%.



Gambar 6.29 *Confusion Matrix* Kec. Tumpang

iv. Kecamatan Lawang

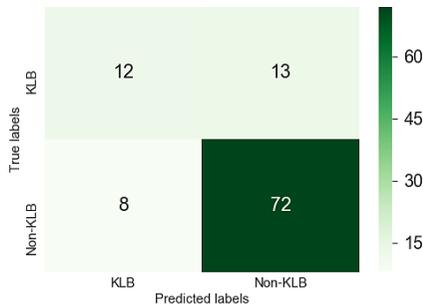
Gambar 6.30 menunjukkan hasil *confusion matrix* yang dibangun pada Kecamatan Lawang. Dari *confusion matrix* tersebut dapat diketahui bahwa dari 25 data KLB pada data, model dapat memprediksi dengan benar 16 diantaranya sehingga nilai sensitivitas yang didapat yakni 64,0%. Sedangkan pada label non-KLB, model dapat memprediksi dengan benar 53 data dari keseluruhan 80 data yang berlabel non-KLB sehingga nilai spesifisitas yang didapat yaitu sebesar 66,3%. Dari keseluruhan kelas yang diprediksi, klasifikasi pada Kecamatan Singosari dapat menghasilkan nilai akurasi sebesar 65,7%.



Gambar 6.30 *Confusion Matrix* Kec. Lawang

v. Kecamatan Wajak

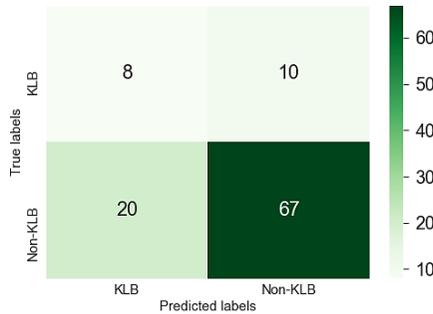
Gambar 6.31 menunjukkan hasil *confusion matrix* yang dibangun pada Kecamatan Wajak. Dari *confusion matrix* tersebut dapat diketahui bahwa dari 25 data KLB pada data, model dapat memprediksi dengan benar 12 diantaranya sehingga nilai sensitivitas yang didapat yakni 48,0%. Sedangkan pada label non-KLB, model dapat memprediksi dengan benar 72 data dari keseluruhan 80 data yang berlabel non-KLB sehingga nilai spesifisitas yang didapat yaitu sebesar 90,0%. Dari keseluruhan kelas yang diprediksi, klasifikasi pada Kecamatan Singosari dapat menghasilkan nilai akurasi sebesar 80,0%.



Gambar 6.31 *Confusion Matrix* Kec. Wajak

vi. Kecamatan Karangploso

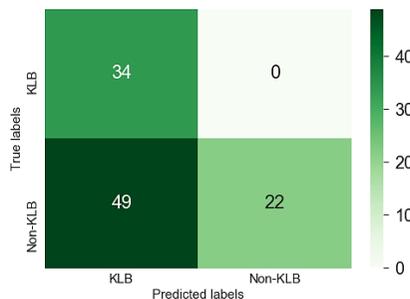
Gambar 6.32 menunjukkan hasil *confusion matrix* yang dibangun pada Kecamatan Karangploso. Dari *confusion matrix* tersebut dapat diketahui bahwa dari 18 data KLB pada data, model dapat memprediksi dengan benar 8 diantaranya sehingga nilai sensitivitas yang didapat yakni 44,4%. Sedangkan pada label non-KLB, model dapat memprediksi dengan benar 67 data dari keseluruhan 87 data yang berlabel non-KLB sehingga nilai spesifisitas yang didapat yaitu sebesar 77,0%. Dari keseluruhan kelas yang diprediksi, klasifikasi pada Kecamatan Singosari dapat menghasilkan nilai akurasi sebesar 71,4%.



Gambar 6.32 Confusion Matrix Kec. Karangploso

vii. Kecamatan Sumbermanjing

Gambar 6.33 menunjukkan hasil *confusion matrix* yang dibangun pada Kecamatan Sumbermanjing. Dari *confusion matrix* tersebut dapat diketahui bahwa dari 34 data KLB pada data, model dapat memprediksi dengan benar semua data tersebut sehingga nilai sensitivitas yang didapat yakni 100%. Sedangkan pada label non-KLB, model dapat memprediksi dengan benar 22 data dari keseluruhan 71 data yang berlabel non-KLB sehingga nilai spesifisitas yang didapat yaitu sebesar 31,0%. Hal ini menunjukkan bahwa model menunjukkan performa yang baik untuk kelas KLB namun kurang baik untuk kelas non-KLB. Dari keseluruhan kelas yang diprediksi, klasifikasi pada Kecamatan Singosari dapat menghasilkan nilai akurasi sebesar 53,3%



Gambar 6.33 Confusion Matrix Kec. Sumbermanjing

6.3.3. Hasil Perhitungan 24 Periode Kedepan pada Dataran Sedang

Pada subbab ini dijelaskan hasil perhitungan 24 periode mendatang yakni periode Januari 2019 hingga Desember 2020 pada dataran rendah. Sub-bab ini dibagi menjadi dua bagian yaitu bagian hasil tahap peramalan dan hasil tahap klasifikasi. Hasil peramalan jumlah kejadian demam berdarah disertai hasil klasifikasi KLB pada Kecamatan Pakisaji ditunjukkan pada Tabel 6.30. Sedangkan hasil peramalan untuk kecamatan lainnya di dataran rendah dijabarkan pada Lampiran C.

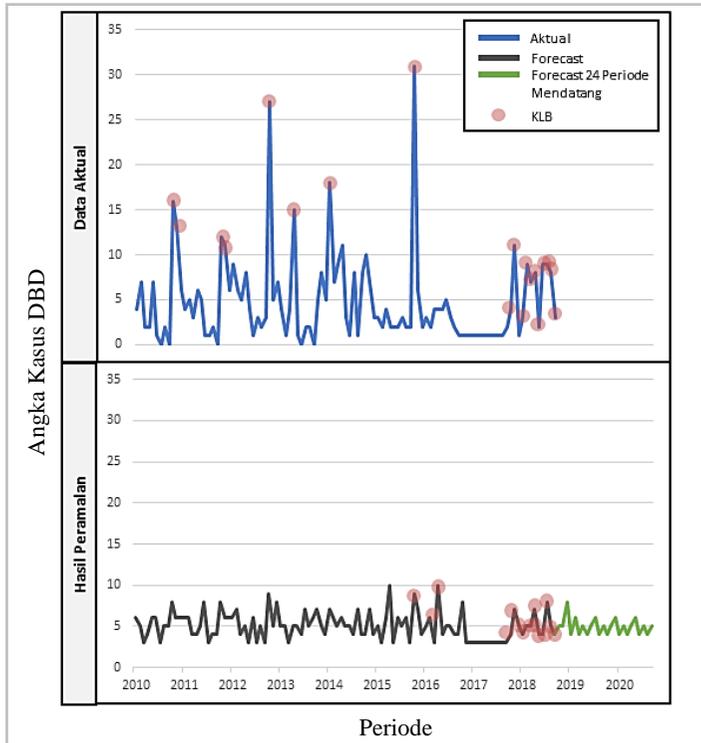
Model yang digunakan pada tahap ini adalah model terbaik yang telah terpilih untuk keseluruhan proses (model final) di dataran sedang, yaitu dengan skenario pembagian data 70-30, dengan menggunakan variabel *input* berupa fitur *lag* kejadian demam berdarah berjumlah 3. Sedangkan parameter XGBoost hasil *tuning* yang digunakan pada model ini yaitu ‘*colsample_bytree*’ sebesar 0.5, ‘*learning_rate*’ sebesar 0.1, ‘*gamma*’ sebesar 0, ‘*max_depth*’ sebesar 2, ‘*min_child_weight*’ sebesar 4, ‘*n_estimator*’ sebesar 50, ‘*reg_alpha*’ sebesar 0, dan ‘*subsample*’ sebesar 0.6.

Tabel 6.27 Hasil Peramalan 2 Tahun Mendatang pada Kec. Pakisaji

Periode	Jumlah KDB	Status
Jan-19	5	non-KLB
Feb-19	5	non-KLB
Mar-19	8	non-KLB
Apr-19	4	non-KLB
Mei-19	6	non-KLB
Jun-19	4	non-KLB
Jul-19	5	non-KLB
Agt-19	4	non-KLB
Sep-19	5	non-KLB
Okt-19	6	non-KLB
Nov-19	4	non-KLB
Des-19	5	non-KLB

Periode	Jumlah KDB	Status
Jan-20	4	non-KLB
Feb-20	5	non-KLB
Mar-20	6	non-KLB
Apr-20	4	non-KLB
Mei-20	5	non-KLB
Jun-20	4	non-KLB
Jul-20	5	non-KLB
Agt-20	6	non-KLB
Sep-20	4	non-KLB
Okt-20	5	non-KLB
Nov-20	4	non-KLB
Des-20	5	non-KLB

Dari hasil peramalan kejadian demam berdarah dan klasifikasi kejadian luar biasa yang dilakukan untuk 24 periode mendatang, didapatkan hasil perbandingan data aktual angka kejadian demam berdarah dan hasil peramalan yang sudah berlabel KLB seperti yang ditunjukkan pada Gambar 6.34.



Gambar 6.34 Perbandingan Data Aktual dan Hasil Peramalan di Dataran Sedang (Kec. Pakisaji)

Dari Gambar 6.34, dapat dilihat bahwa secara keseluruhan model yang terbaik di dataran sedang masih belum mampu menangkap pola data khususnya pada lonjakan angka kejadian demam berdarah dengan cukup baik, hal ini tergambarkan dengan nilai RMSE yang cukup tinggi jika dibandingkan dengan model pada dataran lainnya yakni sebesar 2,899. Meskipun begitu, akurasi yang dihasilkan untuk tahap

klasifikasi cukup tinggi yakni sebesar 93,75%, hal ini terjadi karena model dapat memprediksi semua label KLB dengan tepat pada *test-set* seperti yang telah dijelaskan pada subbab 6.3.2.

6.4. Hasil Peramalan dan Klasifikasi Dataran Tinggi

Pada subbab ini dijelaskan hasil peramalan pada dataran tinggi menggunakan metode XGBoost yang dilanjutkan dengan Klasifikasi Berbasis Aturan untuk menentukan status KLB dan non-KLB pada data. Sub-bab ini terdiri dari tiga bagian yaitu hasil peramalan, hasil klasifikasi, dan hasil perhitungan 24 periode mendatang.

6.4.1. Hasil Peramalan pada Dataran Tinggi

Pada subbab ini pembahasan difokuskan pada hasil peramalan yang dilakukan dengan metode XGBoost di dataran tinggi. Sub-bab ini terdiri dari dua bagian yaitu hasil peramalan pada kecamatan pembentuk model dan hasil peramalan pada kecamatan selain pembentuk model.

a. Hasil Uji Coba Skenario Peramalan

Pada proses pembentukan model peramalan di dataran tinggi, terdapat beberapa skenario yang dilakukan dan masing-masing skenario menghasilkan satu model. Dari keseluruhan model yang telah dibangun tersebut kemudian dilakukan pengujian dengan membandingkan nilai RMSE tiap model.

Data yang digunakan untuk pembentukan model pada dataran tinggi yaitu data Kecamatan Ngajum dengan menggunakan *input* sesuai dengan skenario yang telah dijelaskan pada subbab 4.1.2 bagian F. Hasil dari pengujian model pada *test-set* Kecamatan Ngajum ditunjukkan pada Tabel 6.16. Sedangkan hasil peramalan lengkap dengan informasi parameter yang digunakan dijelaskan pada Lampiran E.

Tabel 6.28 Hasil RMSE untuk *test-set* di Kec. Ngajum

Skenario			RMSE
Pembagian Data	Pemilihan Fitur	Windowing	
80-20	ALL-VI	2	1,71025
	TOP3-VI	3	1,45578
	TOP2-VI	12	1,38467
	TOP1-VI	12	1,78827
	NO-VI	10	1,53662
70-30	ALL-VI	12	1,44641
	TOP3-VI	0	1,53984
	TOP2-VI	12	1,50977
	TOP1-VI	4	1,93631
	NO-VI	10	1,93938
60-40	ALL-VI	8	4,55046
	TOP3-VI	2	4,59019
	TOP2-VI	7	4,65775
	TOP1-VI	2	4,65179
	NO-VI	10	4,57011
50-50	ALL-VI	6	4,15181
	TOP3-VI	4	4,19127
	TOP2-VI	2	4,14650
	TOP1-VI	1	4,17403
	NO-VI	10	3,98458

Pada Tabel 6.28, dapat diketahui nilai RMSE untuk tiap kombinasi skenario pembagian proporsi data dan pemilihan fitur diikuti dengan jumlah *lag* optimal pada skenario tersebut. Dari model-model yang dibangun, model terbaik didapatkan dari urutan nilai MSE yang terkecil. Terdapat tiga kandidat model terbaik seperti yang tertera pada Tabel 6.30, yaitu yang pertama pada skenario pembagian data 80-20 dan menggunakan

input berupa variabel curah hujan, jumlah penduduk dan fitur *lag* berjumlah 12 dengan nilai RMSE sebesar 1,38467. Kemudian model yang kedua yakni dengan pembagian data 70-30 yang melibatkan variabel *input* seluruh variabel independen utama dan fitur *lag* berjumlah 12 dengan nilai RMSE sebesar 1,45578. Model terbaik ketiga pada dataran sedang yaitu model dengan skenario pembagian data 80-20 dan variabel *input* berupa variabel curah hujan, jumlah penduduk, dan fitur *lag* berjumlah 3 dengan nilai RMSE sebesar 1,93938. Ketiga model tersebut memiliki nilai SMAPE berturut-turut, 22,78%, 24,13% dan 25,02%.

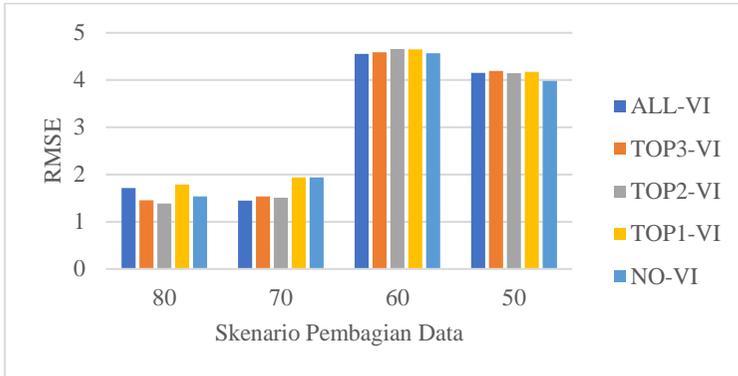
Tabel 6.29 Hasil Tuning Parameter di Dataran Tinggi

Model	Skenario	Parameter							
		col	lr	gm	md	mcw	n_est	alpha	sub
1	80-20	0.75	0.3	0.01	4	1	50	0	0.8
	TOP2-VI								
	Lag 12								
2	70-30	0.75	0.5	0	8	1	50	0	1
	ALL-VI								
	Lag 12								
3	80-20	0.75	0.5	0.01	4	1	100	0.01	0.8
	TOP3-VI								
	Lag 3								

Pada Tabel 6.29, informasi mengenai parameter hasil *tuning* menggunakan *grid-search* ditampilkan untuk tiga kandidat model peramalan terbaik di dataran rendah. Atribut ‘Model’ menunjukkan urutan model terbaik berdasarkan nilai RMSE terkecil, ‘Skenario’ menunjukkan skenario pembentukan model yang digunakan, dan ‘Parameter’ menunjukkan parameter yang dihasilkan dari proses *tuning* yang dilakukan. Parameter yang tertera pada tabel tersebut antara lain, ‘col’ merepresentasikan *colsample_bytree*, ‘lr’ merepresentasikan *learning_rate*, ‘gm’ merepresentasikan *gamma*, ‘md’ merepresentasikan *max_depth*, ‘mcw’ merepresentasikan *min_child_weight*, ‘n_est’ merepresentasikan *n_estimator*, alpha merepresentasikan *reg_alpha*, dan ‘sub’ yang merepresentasikan parameter *subsample*.

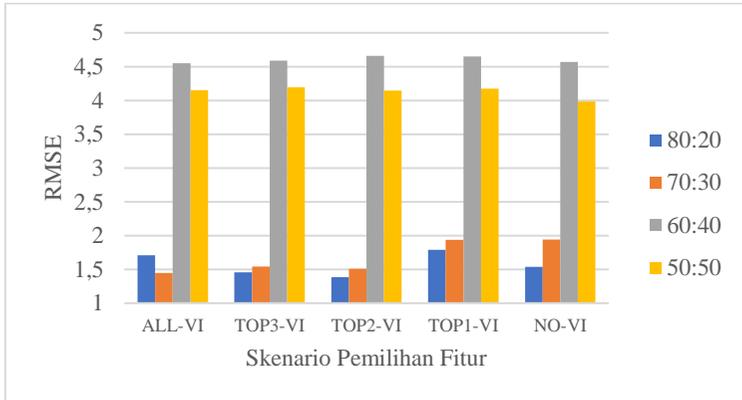
Dari hasil *tuning* parameter, dapat diketahui bahwa ketiga model dengan RMSE terkecil menggunakan parameter yang sama untuk rasio fitur yang digunakan dalam membentuk model pada parameter *colsample_bytree* sebesar 0.5, dan nilai regularisasi *reg_alpha* sebesar 1, dan jumlah estimator sebanyak 50. Hal ini berarti model hanya menggunakan separuh dari keseluruhan fitur yang ada untuk membangun model tiap iterasinya dengan jumlah iterasi *boosting* sebanyak 50 kali. Model pertama dan ketiga memiliki kesamaan pada empat parameter yakni nilai *learning_rate* sebesar 0.1, *max_depth* sebesar 2, *min_child_weight* sebesar 4 dan *subsample* sebesar 0.6. Sedangkan pada model kedua, nilai parameter *learning_rate*, *max_depth*, *min_child_weight*, dan *subsample* berturut-turut sebesar 0.5, 4, 1, dan 0.8. Untuk nilai *gamma*, model pertama dan kedua memiliki kesamaan nilai yaitu sebesar 0.01, sedangkan pada model ketiga nilai *gamma* yg digunakan yaitu 0. Informasi hasil *tuning* parameter lebih lengkap terdapat pada Lampiran E.

Dari hasil pengujian yang telah dilakukan, selain informasi mengenai model terbaik, dapat diketahui pula pola persebaran nilai RMSE yang dihasilkan. Persebaran RMSE berdasarkan skenario pembagian data di dataran sedang ditunjukkan pada Gambar 6.35. Dari visualisasi tersebut, dapat diketahui bahwa RMSE yang dihasilkan untuk pembagian data 80-20 dan 70-30 relatif lebih rendah dibandingkan RMSE yang dihasilkan dari pembagian data 60-40 dan 50-50. Hal ini disebabkan oleh semakin sedikitnya proporsi data pelatihan, maka semakin sedikit pula data yang menjadi *input* dalam pembentukan model. Hal tersebut dapat menyebabkan model yang dihasilkan *underfit*, yakni model tidak bisa melihat logika di belakang data secara baik, sehingga tidak bisa melakukan peramalan secara tepat dan memiliki nilai *error* yang relatif tinggi.



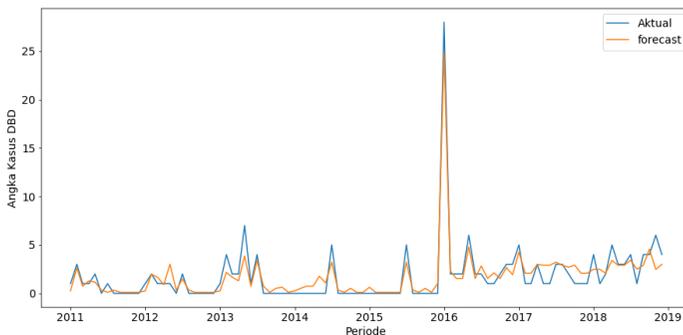
Gambar 6.35 Persebaran RMSE berdasarkan Skenario Pembagian Data (Dataran Tinggi)

Gambar 6.36 merupakan grafik yang menunjukkan persebaran RMSE berdasarkan skenario pemilihan fitur. Dari grafik tersebut dapat diketahui bahwa RMSE yang dihasilkan pada masing-masing skenario pemilihan fitur terbagi menjadi 2 cluster data dengan pola yang serupa, yakni pada pembagian data 50-50 dengan 60-40 dan 80-20 dengan 70-30. Untuk proporsi data 60-40 dan 50-50, RMSE paling rendah dihasilkan oleh model yang dibangun menggunakan variabel Angka Bebas Jentik dan curah hujan, atau Angka Bebas Jentik saja ditambah dengan fitur *lag* yang optimal. Pada proporsi data tersebut, menggunakan lebih dari 3 variabel independen utama atau tidak melibatkan variabel independen utama sama sekali akan menghasilkan model dengan nilai RMSE yang lebih tinggi. Sedangkan untuk proporsi data 70-30 dan 80-20, nilai RMSE menunjukkan tren menurun seiring dengan berkurangnya jumlah variabel independen utama yang digunakan. Pada proporsi data tersebut, model yang dibangun dengan hanya melibatkan fitur *lag* sebagai input menghasilkan nilai RMSE yang paling kecil.



Gambar 6.36 Persebaran RMSE berdasarkan Skenario Pemilihan Fitur (Dataran Tinggi)

Perbandingan data aktual dan hasil peramalan yang dilakukan pada kecamatan pembentuk model yaitu Kecamatan Ngajum dengan menggunakan model peramalan dengan RMSE terkecil tertera pada Gambar 6.37. Dapat dilihat pada gambar tersebut bahwa model bisa menangkap pola data dengan cukup baik. Hal ini dapat direpresentasikan juga dari nilai RMSE sebesar 1,38467 dan SMAPE sebesar 22,78%.



Gambar 6.37 Perbandingan Data Aktual dan Hasil Peramalan Kec. Ngajum

Selain nilai RMSE pengujian pada *test-set*, tingkat *robustness* model juga perlu diperhitungkan untuk mengetahui seberapa

baik model diterapkan pada data lainnya. Salah satu cara mengetahui tingkat *robustness* model yaitu dengan menguji model tersebut pada proporsi data yang berbeda. Hasil uji proporsional data untuk tiga model dengan nilai RMSE terbaik tertera pada Tabel 6.30. Hasil lengkap perbandingan model dapat dilihat pada Lampiran E.

Tabel 6.30 Hasil Uji Proporsional Data pada Kec. Ngajum

Model Terbaik ke-	RMSE			Mean	St.dev
	Uji Proporsional Data				
	proporsi 1	proporsi 2	proporsi 3		
1	1,2351	1,2183	1,1333	1,1956	0,0445
2	1,5007	1,2473	1,1243	1,2908	0,1567
3	1,3315	1,3196	1,1837	1,2783	0,0670

Seperti yang tertera pada Tabel 6.30, terdapat tiga uji proporsi yang dilakukan, skenario uji tersebut telah dijelaskan pada subbab 4.2.3. Dari hasil uji proporsi yang didapat kemudian dihitung nilai rata-rata dan standar deviasinya agar dapat diketahui mana model yang memiliki tingkat *robustness* paling tinggi. Mengacu pada Tabel 6.30, dua atribut yang digunakan untuk menunjang perhitungan tersebut yakni atribut ‘Mean’ yang menunjukkan nilai rata-rata hasil uji proporsi, sedangkan ‘St.dev’ yang menunjukkan nilai standar deviasi yang dihasilkan. Dapat diketahui dari hasil uji proporsi tersebut bahwa model pertama menghasilkan nilai RMSE yang paling kecil diantara dua model lainnya baik secara rata-rata RMSE maupun nilai standar deviasi. Hal ini menunjukkan bahwa model menunjukkan performa yang baik dalam melakukan peramalan di kecamatan pembentuk model, yakni Kecamatan Ngajum.

b. Hasil Peramalan pada Kecamatan Selain Pembentuk Model

Selain pengujian yang dilakukan pada proporsi pembagian data yang berbeda, model perlu diuji terhadap data kecamatan lain

pada dataran yang sama. Di dataran sedang, kecamatan uji terdiri dari Kecamatan Dampit, Kecamatan Singosari, Kecamatan Tumpang, Kecamatan Lawang, Kecamatan Sumbermanjing, Kecamatan Wajak, dan Kecamatan Karangploso. Hasil perbandingan nilai RMSE yang dihasilkan pada tiap kecamatan uji untuk tiga model peramalan terbaik ditunjukkan pada Tabel 6.31. Hasil lengkap perbandingan RMSE uji kecamatan lain terlampir pada Lampiran E.

Tabel 6.31 Perbandingan Hasil RMSE pada Dataran Tinggi

		Model Terbaik ke-		
		1	2	3
RMSE Uji Kecamatan Lain	Poncokusumo	2,333	2,460	2,338
	Jabung	2,190	2,179	2,402
Mean		2,2616	2,3193	2,3703
St.dev		0,0711	0,1404	0,0319

Dari hasil pengujian terhadap kecamatan lain, dapat diketahui bahwa model terbaik pertama menghasilkan nilai rata-rata RMSE terkecil, yaitu 2,2616. Sedangkan nilai standar deviasinya bukan yang paling kecil diantara kedua model lainnya. Hal ini berarti secara rata-rata, performa model baik dalam melakukan peramalan namun memiliki variasi angka RMSE tiap kecamatan yang sedikit lebih tinggi daripada model ketiga. Hal ini menunjukkan bahwa model yang memiliki nilai rata-rata RMSE tiap kecamatan paling kecil, belum tentu menghasilkan variasi nilai RMSE antarkecamatan yang paling kecil pula.

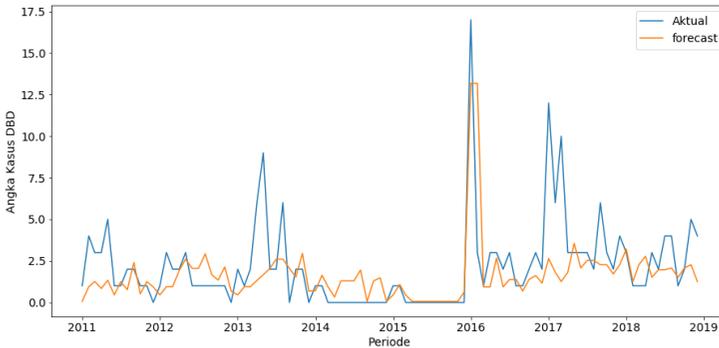
Selain perhitungan nilai RMSE, pengukuran lain yang digunakan untuk menunjang analisis peramalan yakni nilai SMAPE atau rata-rata persentase kesalahan absolut simetris. Nilai SMAPE yang dihasilkan model peramalan terbaik di dataran sedang pada pengujian kecamatan selain pembentuk model ditunjukkan pada Tabel 6.32.

Tabel 6.32 Hasil Nilai SMAPE Kecamatan Lain (Dataran Tinggi)

No	Kecamatan	SMAPE
1	Poncokusumo	47,68%
2	Jabung	42,76%

i. Kecamatan Poncokusumo

Perbandingan jumlah kasus demam berdarah aktual dan hasil peramalan di Kecamatan Poncokusumo ditunjukkan pada Gambar 6.38. Dari gambar tersebut dapat diketahui bahwa beberapa kali model cukup baik dalam menangkap pola kenaikan jumlah kasus demam berdarah. Namun disamping itu model juga terkadang gagal menangkap pola lonjakan angka kejadian demam berdarah dengan baik. Hal ini dapat digambarkan pula dengan nilai RMSE sebesar 2,3328 dan SMAPE sebesar 47,68% yang dihasilkan pada Kecamatan Poncokusumo.

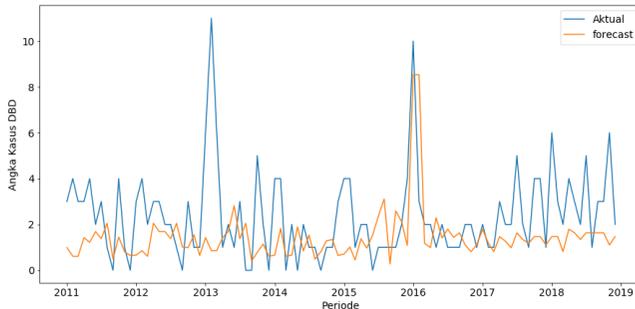


Gambar 6.38 Perbandingan Data Aktual dan Hasil Peramalan Kec. Poncokusumo

ii. Kecamatan Jabung

Perbandingan jumlah kasus demam berdarah aktual dan hasil peramalan di Kecamatan Jabung ditunjukkan pada Gambar 6.39. Dari gambar tersebut dapat diketahui bahwa beberapa kali model gagal dalam meramalkan lonjakan jumlah kasus demam berdarah pada data. Hal ini sejalan dengan hasil uji korelasi

antarkecamatan yang dapat dilihat pada Lampiran D, bahwa Kecamatan Jabung memiliki korelasi rendah yaitu sebesar 0,2 terhadap kecamatan pembentuk model yakni Kecamatan Ngajum. Nilai RMSE yang dihasilkan di Kecamatan Ngajum yaitu sebesar 2,1905 dan nilai SMAPE sebesar 42,76%.



Gambar 6.39 Perbandingan Data Aktual dan Hasil Peramalan Kec. Jabung

6.4.2. Hasil Klasifikasi pada Dataran Tinggi

Pada subbab ini pembahasan difokuskan pada hasil klasifikasi yang dilakukan dengan metode klasifikasi berbasis aturan di dataran rendah. Sub-bab ini terdiri dari dua bagian yaitu hasil klasifikasi pada kecamatan pembentuk model dan hasil klasifikasi pada kecamatan selain pembentuk model.

a. Hasil Uji Coba Skenario Klasifikasi

Dalam penelitian kali ini, pengukuran evaluasi klasifikasi biner yang digunakan adalah sensitivitas, spesifisitas, dan akurasi. Klasifikasi biner yang baik akan menghasilkan nilai yang tinggi untuk ketiga faktor tersebut, sedangkan klasifikasi biner yang buruk akan menghasilkan nilai rendah pada faktor-faktor tersebut. Agar memastikan ketiga faktor tersebut baik pada suatu model, dilakukan perataan pada nilai akurasi, sensitivitas, dan spesifisitas seperti yang tertera pada Tabel 6.34 dengan atribut 'Rata-rata'. Pada tabel tersebut, atribut 'Model' merujuk pada tiga model peramalan terbaik pada Tabel 6.30. Evaluasi klasifikasi untuk ketiga model tersebut dilakukan untuk

mengetahui apakah model peramalan terbaik juga menghasilkan nilai evaluasi klasifikasi yang baik pula.

Tabel 6.33 Evaluasi Klasifikasi pada *Test-set* (Dataran Tinggi)

Model	Akurasi	Sensitivitas	Spesifisitas	Rata-rata
1	70,00%	14,29%	100,00%	61,43%
2	86,21%	66,67%	100,00%	84,29%
3	76,19%	28,57%	100,00%	68,25%

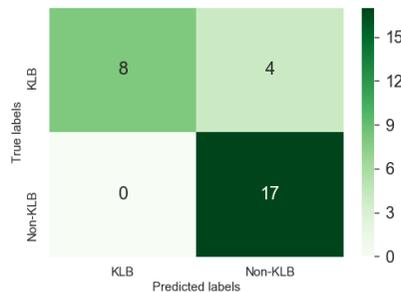
Dari Tabel 6.34, dapat diketahui bahwa model kedua menghasilkan nilai paling tinggi pada ketiga faktor (akurasi, sensitivitas, spesifisitas) dengan nilai rata-rata sebesar 84,29% dengan selisih angka yang cukup signifikan dengan kedua model lainnya. Model ketiga dan pertama memiliki nilai rata-rata berturut-turut 68,25% dan 61,43%. Namun, selain pengujian pada *test-set*, perlu juga dilakukan pengujian terhadap proporsi data lain dan kecamatan lainnya. Hasil pengujian tersebut dapat dilihat pada Tabel 6.35, dimana ‘mod’ menunjukkan urutan model terbaik, ‘acc’ menunjukkan rata-rata nilai akurasi yang dihasilkan pada pengujian, ‘sens’ menunjukkan rata-rata nilai sensitivitas yang dihasilkan pada pengujian, ‘spec’ menunjukkan rata-rata nilai spesifisitas yang dihasilkan pada pengujian, dan ‘avg’ menunjukkan nilai rata-rata keseluruhan pengukuran (akurasi, sensitivitas, dan spesifisitas). Penjabaran lengkap hasil pengujian yang dilakukan dilampirkan pada Lampiran C.

Tabel 6.34 Evaluasi Klasifikasi pada Uji Proporsi dan Uji Kecamatan Lain (Dataran Rendah)

Mod	Uji Proporsi				Uji Kecamatan Lain			
	Acc	Sens	Spec	Avg	Acc	Sens	Spec	Avg
1	80,0%	62,8%	93,1%	78,6%	79,7%	51,6%	93,9%	75,1%
2	87,1%	67,3%	100%	84,8%	85,6%	61,5%	97,2%	81,4%
3	83,4%	72,2%	92,4%	82,7%	79,4%	52,2%	92,9%	74,9%

Selain model kedua memiliki nilai yang paling tinggi untuk pengujian pada *test-set* dengan perbedaan yang cukup signifikan dengan kedua model lain, dapat diketahui pula bahwa model ini memiliki nilai yang cukup baik pula pada uji proporsi maupun pada uji kecamatan lain. Model ini mampu menghasilkan nilai rata-rata masing-masing uji berturut-turut, 84,8% dan 81,4%. Di sisi lain, model pertama yang merupakan model terbaik pada tahap peramalan justru menghasilkan nilai sensitivitas yang paling kecil yaitu hanya 14% seperti yang dapat dilihat pada Lampiran C. Hal ini menunjukkan bahwa model peramalan yang terbaik (RMSE paling kecil) belum tentu menghasilkan nilai akurasi, sensitivitas, dan spesifisitas yang baik pula. Sehingga untuk kasus di dataran tinggi, model yang akan digunakan sebagai model final adalah model yang kedua.

Hasil klasifikasi dapat divisualisasikan melalui *confusion matrix*. *Confusion matrix* yang dibangun dengan model final untuk Kecamatan Ngajum pada *test-set* ditunjukkan pada Gambar 6.39, dimana terdapat kelas biner ‘KLB’ dan ‘non-KLB’ sebagai label yang digunakan.



Gambar 6.40 Confusion Matrix Kec. Ngajum

Dari *confusion matrix* yang dibangun, dapat diketahui bahwa dari 12 data KLB, 8 diantaranya dapat diprediksi dengan benar sehingga nilai sensitivitas dari model yang dibangun bernilai 66,7%. Sedangkan untuk kelas berlabel negatif, dari 17 data non-KLB, keseluruhan data tersebut dapat diprediksi dengan

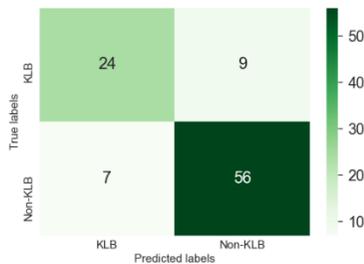
benar sehingga nilai spesifisitas yang dihasilkan sebesar 100%. Sedangkan untuk nilai akurasi yang dihasilkan sebesar 86,2%.

b. Hasil Klasifikasi pada Kecamatan Selain Pembentuk Model

Sub-bab ini menjelaskan tentang hasil klasifikasi pada kecamatan uji, yaitu kecamatan selain pembentuk model. Di dataran tinggi, kecamatan yang berperan sebagai kecamatan uji adalah Kecamatan Poncokusumo dan Kecamatan Jabung.

i. Kecamatan Poncokusumo

Gambar 6.38 menunjukkan hasil *confusion matrix* yang terbentuk untuk Kecamatan Poncokusumo. Dari *confusion matrix* tersebut dapat diketahui bahwa dari 33 data KLB pada data, model dapat memprediksi dengan benar 24 diantaranya sehingga didapatkan nilai sensitivitas sebesar 72,7%. Sedangkan untuk label non-KLB, model dapat memprediksi dengan benar 56 data dari keseluruhan 73 data yang berlabel non-KLB, sehingga nilai spesifisitas yang dihasilkan yakni sebesar 88,9%. Dengan nilai sensitivitas dan spesifisitas yang cukup tinggi tersebut, model dapat menghasilkan nilai akurasi yang cukup tinggi pula yaitu sebesar 83,3%.

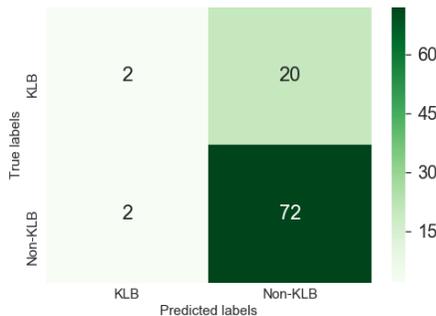


Gambar 6.41 *Confusion Matrix* Kec. Poncokusumo

ii. Kecamatan Jabung

Gambar 6.39 menunjukkan hasil *confusion matrix* yang terbentuk untuk Kecamatan Jabung. Dari *confusion matrix* tersebut dapat diketahui bahwa dari 22 data KLB pada data,

model hanya dapat memprediksi dengan benar 2 diantaranya sehingga didapatkan nilai sensitivitas sebesar 9,1%. Sedangkan untuk label non-KLB, model dapat memprediksi dengan benar 72 data dari keseluruhan 74 data yang berlabel non-KLB, sehingga nilai spesifisitas yang dihasilkan yakni sebesar 97,3%. Hal ini menunjukkan bahwa model yang dibangun hanya baik untuk memprediksi salah satu kelas di Kecamatan Jabung. Dengan nilai sensitivitas dan specificity tersebut, model ini dapat menghasilkan nilai akurasi sebesar 77,1%.



Gambar 6.42 Confusion Matrix Kec. Jabung

6.4.3. Hasil Perhitungan 24 Periode Kedepan pada Dataran Tinggi

Pada subbab ini dijelaskan hasil perhitungan 24 periode mendatang yakni periode Januari 2019 hingga Desember 2020 pada dataran tinggi. Sub-bab ini dibagi menjadi dua bagian yaitu bagian hasil tahap peramalan kejadian demam berdarah dan hasil tahap klasifikasi kejadian luar biasa. Hasil peramalan jumlah kejadian demam berdarah pada Kecamatan Ngajum ditunjukkan pada Tabel 6.36. Sedangkan hasil peramalan kecamatan lainnya di dataran rendah dijabarkan pada Lampiran C.

Model yang digunakan pada tahap ini adalah model terbaik yang telah terpilih untuk keseluruhan proses (model final) di dataran tinggi, yaitu dengan skenario pembagian data 80-20, dengan menggunakan *input* berupa variabel Angka Bebas Jentik (ABJ),

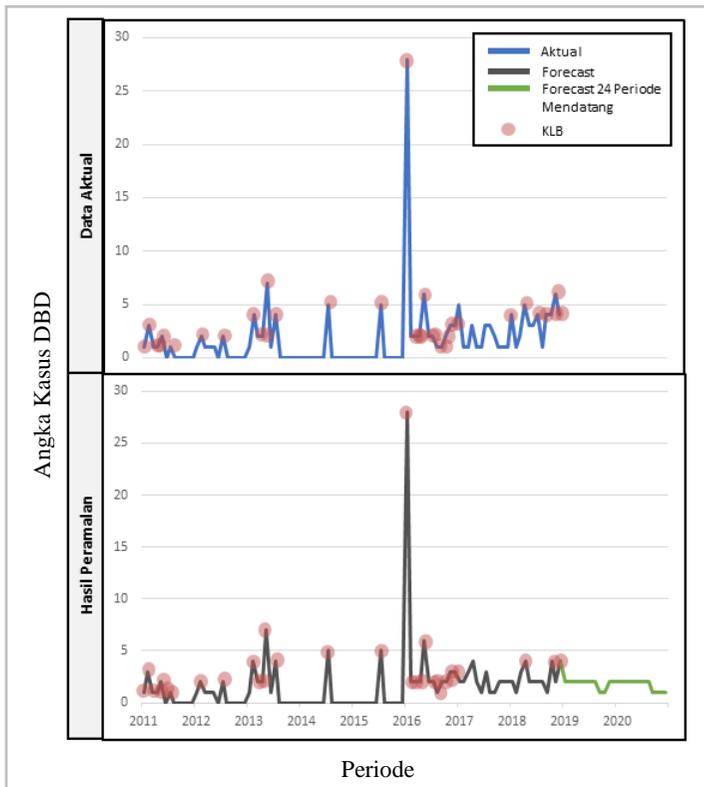
curah hujan, kelembapan udara, kecepatan angin, jumlah penduduk, dan suhu udara dengan tambahan fitur *lag* berjumlah 12. Sedangkan parameter XGBoost hasil *tuning* yang digunakan pada model ini yaitu ‘*colsample_bytree*’ sebesar 0.75, ‘*learning_rate*’ sebesar 0.5, ‘*gamma*’ sebesar 0, ‘*max_depth*’ sebesar 8, ‘*min_child_weight*’ sebesar 1, ‘*n_estimator*’ sebesar 50, ‘*reg_alpha*’ sebesar 0, dan ‘*subsample*’ sebesar 1.

Tabel 6.35 Hasil Peramalan 2 Tahun Mendatang pada Kec. Ngajum

Periode	Jumlah KDB	Status	Periode	Jumlah KDB	Status
Jan-19	2	non-KLB	Jan-20	2	non-KLB
Feb-19	2	non-KLB	Feb-20	2	non-KLB
Mar-19	2	non-KLB	Mar-20	2	non-KLB
Apr-19	2	non-KLB	Apr-20	2	non-KLB
Mei-19	2	non-KLB	Mei-20	2	non-KLB
Jun-19	2	non-KLB	Jun-20	2	non-KLB
Jul-19	2	non-KLB	Jul-20	2	non-KLB
Agt-19	2	non-KLB	Agt-20	2	non-KLB
Sep-19	1	non-KLB	Sep-20	1	non-KLB
Okt-19	1	non-KLB	Okt-20	1	non-KLB
Nov-19	2	non-KLB	Nov-20	1	non-KLB
Des-19	2	non-KLB	Des-20	1	non-KLB

Dari hasil peramalan kejadian demam berdarah dan klasifikasi kejadian luar biasa yang dilakukan untuk 24 periode mendatang, didapatkan hasil perbandingan data aktual angka kejadian demam berdarah dan hasil peramalan yang sudah berlabel KLB seperti yang ditunjukkan pada Gambar 6.43. Dari gambar tersebut dapat diketahui bahwa secara keseluruhan model yang terpilih di dataran tinggi dapat menghasilkan pola data yang serupa dengan data aktual. Hal ini berdampak pada label KLB yang dapat diprediksi dengan cukup baik sehingga

menghasilkan nilai akurasi yang tinggi yakni sebesar 86,21% seperti yang telah dijelaskan pada subbab 6.4.2.



Gambar 6.43 Perbandingan Data Aktual dan Hasil Peramalan di Dataran Tinggi (Kec. Ngajum)

6.5. Hasil Pencarian Satu Model Terbaik untuk Seluruh Kecamatan

Selain dilihat dari sudut pandang masing-masing dataran, model juga dapat diujicobakan untuk keseluruhan data kecamatan tanpa melibatkan pengelompokan data berdasarkan dataran. Dari hasil uji coba tersebut akan diketahui mana model yang menjadi model paling baik untuk keseluruhan data kecamatan.

Hasil ujicoba ditunjukkan pada tabel 6.36 dimana ditampilkan nilai RMSE dan SMAPE untuk sembilan model yang didapat dari tiga model terbaik pada tiap dataran. Perhitungan RMSE dan SMAPE disajikan dalam bentuk nilai rata-rata serta nilai median yang dihasilkan masing-masing model saat diterapkan pada 15 kecamatan lainnya.

Tabel 6.36 Hasil Pencarian Satu Model Terbaik

Model	RMSE		SMAPE (%)	
	Mean	Median	Mean	Median
Model DR 1	4,83	3,56	46,32	44,44
Model DR 2	4,95	3,48	49,17	45,50
Model DR 3	4,84	3,50	46,02	44,81
Model DS 1	4,99	3,39	47,06	47,18
Model DS 2	6,20	4,40	48,32	49,57
Model DS 3	5,36	3,41	47,89	47,62
Model DT 1	4,40	3,43	43,96	43,07
Model DT 2	4,52	3,60	46,62	44,48
Model DT 3	4,71	3,73	51,64	48,93

Dari Tabel 6.36 dapat diketahui bahwa Model DT 1 yakni model yang dibentuk dari data Kecamatan Ngajum menghasilkan nilai rata-rata RMSE, rata-rata SMAPE dan median SMAPE yang paling kecil diantara model lainnya. Hal ini menunjukkan bahwa ternyata model dengan nilai RMSE yang paling kecil pada *test-set* dari keseluruhan model yang terbentuk merupakan model yang paling *robust* saat diujicobakan pada keseluruhan data kecamatan. Dalam pembentukannya, model ini menggunakan skenario pembagian data 80-20 dengan melibatkan variabel curah hujan, jumlah penduduk dan fitur lag berjumlah 12. Sedangkan jika dilihat dari segi klasifikasi Kejadian Luar Biasa, model ini dapat menghasilkan nilai rata-rata performa klasifikasi yang cukup baik saat diujicobakan pada kecamatan lain yakni sebesar 81,4% dengan rincian nilai

akurasi sebesar 85,6%, sensitivitas sebesar 61,5% dan spesifisitas sebesar 97,2%.

6.6. Hasil Perbandingan dengan Metode Lain

Pada subbab ini dijelaskan perbandingan antara performa peramalan dan klasifikasi metode yang digunakan pada penelitian dengan metode lainnya. Penjelasan dibagi menjadi dua bagian yakni perbandingan performa peramalan dan perbandingan performa klasifikasi.

6.6.1. Perbandingan Performa Peramalan

Untuk mengetahui lebih lanjut apakah performa peramalan dengan metode XGBoost yang telah dilakukan sudah cukup baik, maka hasil peramalan yang telah didapat perlu dibandingkan dengan metode lainnya. Perbandingan RMSE yang dihasilkan oleh metode XGBoost dan metode lain yakni KNN Regression, Random Forest, Decision Tree, dan Lasso Regression tertera pada Tabel 6.40.

Tabel 6.37 Perbandingan RMSE Terbaik dengan Algoritma Lain

Metode	Dataran			Avg
	Rendah	Sedang	Tinggi	
XGBoost	2,399	2,699	1,38467	2,161
Decision Tree	2,481	4,337	2,419	3,079
Linear Regression	2,008	3,530	2,424	2,654
RBFNN	2,165	2,687	1,9039	2,252
KNN Regression	2,133	2,293	1,543	1,989
Random Forest	2,257	2,596	1,344	2,066

Dari Tabel 6.40, dapat diketahui bahwa peramalan jumlah kejadian demam berdarah yang dilakukan dengan metode XGBoost menghasilkan performa yang lebih baik daripada beberapa algoritma *machine learning* lain yang lebih sederhana yaitu Decision Tree dan Linear Regression. Sedangkan metode XGBoost memiliki selisih nilai RMSE lebih rendah namun tidak berbeda secara signifikan dengan algoritma *machine*

learning lainnya yaitu KNN Regression dan RBFNN yakni dengan selisih rata-rata RMSE sebesar $\pm 0,3$. Sedangkan jika dibandingkan dengan algoritma *ensemble* lainnya yaitu Random Forest, XGBoost menghasilkan nilai RMSE yang sedikit lebih kecil dengan selisih rata-rata RMSE hanya sebesar $\pm 0,1$.

6.6.2. Perbandingan Performa Klasifikasi

Perbandingan hasil evaluasi klasifikasi juga dilakukan untuk mengetahui seberapa baik metode yang diusulkan pada penelitian ini jika dibandingkan dengan metode lain, dalam kasus ini yaitu klasifikasi yang dilakukan secara langsung pada data KLB tanpa melewati proses regresi kejadian demam berdarah terlebih dahulu. Perbandingan ini menggunakan algoritma yang sama yaitu XGBoost, karena selain dapat digunakan untuk kasus regresi, XGBoost juga dapat digunakan sebagai penggolong (*classifier*). Perbandingan hasil klasifikasi dengan metode tersebut ditunjukkan pada Tabel 6.41. Pada tabel tersebut, atribut 'avg' digunakan untuk merata-rata tiga faktor penilaian yang digunakan, yakni akurasi, sensitivitas, dan spesifisitas agar dapat dilihat seberapa baik model yang dibangun secara keseluruhan.

Tabel 6.38 Perbandingan Performa Klasifikasi dengan Metode Lain

Metode	Pengukuran	Dataran		
		Rendah	Sedang	Tinggi
Regresi XGBoost - Klasifikasi Berbasis Aturan	akurasi	83,30%	93,80%	86,21%
	sensitivitas	50,00%	100,00%	66,70%
	spesifisitas	88,50%	90,50%	100,00%
	avg	73,90%	94,70%	84,30%
Klasifikasi XGBoost	akurasi	76,00%	68,80%	44,80%
	sensitivitas	0,00%	9,10%	100,00%
	spesifisitas	88,40%	100,00%	5,90%
	avg	54,80%	59,30%	50,20%

Dari Tabel 6.41, dapat diketahui bahwa secara umum metode yang diusulkan yakni regresi XGBoost yang dilanjutkan dengan Klasifikasi Berbasis Aturan menghasilkan nilai ketiga faktor (akurasi, sensitivitas, dan spesifisitas) yang lebih tinggi jika dibandingkan nilai yang dihasilkan pada metode klasifikasi XGBoost secara langsung. Kedua metode tersebut memiliki nilai selisih untuk ketiga faktor pengukuran yang cukup jauh, yakni sekitar 20%-35%.

6.7. Kesimpulan Hasil Percobaan

Setelah melakukan semua proses peramalan dan klasifikasi, maka didapatkan model yang terbaik atau yang disebut di penelitian ini dengan ‘model final’. Ketiga model final pada tiap dataran disertai dengan masing-masing skenario pembentukan model tertera pada Tabel 6.38. Atribut parameter XGBoost ‘col’ merepresentasikan *colsample_bytree*, ‘lr’ merepresentasikan *learning_rate*, ‘gm’ merepresentasikan *gamma*, ‘md’ merepresentasikan *max_depth*, ‘mcw’ merepresentasikan *min_child_weight*, ‘n_est’ merepresentasikan *n_estimator*, alpha merepresentasikan *reg_alpha*, dan ‘sub’ yang merepresentasikan parameter *subsampling*.

Tabel 6.39 Parameter pada Model terbaik Tiap Dataran

Dataran	Skenario	Parameter XGBoost							
		col	lr	gm	md	mcw	n_est	alpha	sub
Rendah	70-30	0,5	0,5	0,01	8	1	50	0	1
	TOP-2								
	Lag 11								
Sedang	70-30	0,5	0,1	0	2	4	50	0	0,6
	NO-VI								
	Lag 3								
Tinggi	70-30	0,75	0,5	0	8	1	50	0	1
	ALL-VI								
	Lag 12								

Performa hasil model final tertera pada Tabel 6.39. Pengukuran yang dicantumkan pada tabel tersebut yakni nilai *Root Mean Square Error* (RMSE) dan *Symmetric Mean Absolute Percentage Error* (SMAPE) sebagai pengukuran pada hasil peramalan, sedangkan nilai akurasi, sensitivitas, dan spesifisitas

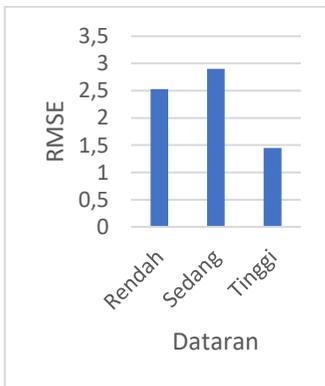
sebagai pengukuran yang digunakan untuk evaluasi hasil klasifikasi.

Tabel 6.40 Model Final pada Tiap Dataran

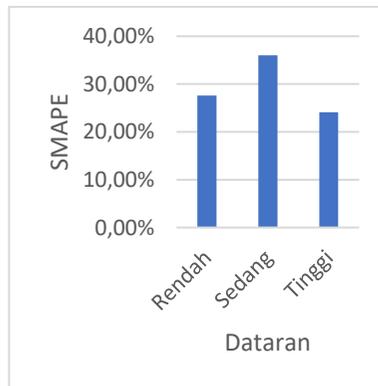
Pengukuran		Dataran		
		Rendah	Sedang	Tinggi
Tahap Peramalan	RMSE	2,532	2,899	1,446
	SMAPE	27,57%	35,99%	24,13%
Tahap Klasifikasi	Akurasi	83,33%	93,75%	86,21%
	Sensitivitas	50,00%	100,00%	66,67%
	Spesifisitas	88,46%	90,48%	100,00%

Visualisasi untuk data evaluasi tahap peramalan pada tabel 6.39 ditunjukkan pada Gambar 6.44. Sedangkan pengukuran tahap klasifikasi divisualisasikan pada Gambar 6.45.

(a) Pengukuran RMSE



(b) Pengukuran SMAPE



Gambar 6.44 Hasil Performa Peramalan (a) RMSE dan (b) SMAPE tiap Dataran

Pemilihan model dengan menggunakan ukuran RMSE dan SMAPE tersebut tidak didasarkan pada *threshold* nilai tertentu yang dapat dijadikan acuan dalam menentukan model baik atau tidak. Melainkan, kedua pengukuran tersebut hanya dapat digunakan untuk mencari model yang paling baik secara relatif

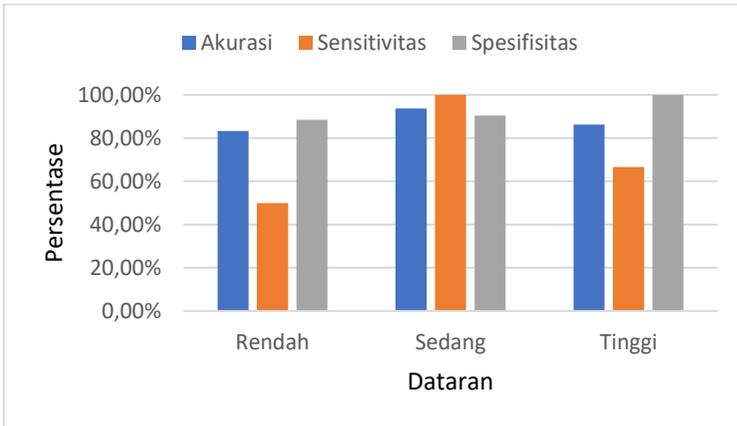
jika dibandingkan dengan model lainnya yang terbentuk. Sehingga, pemilihan model pada penelitian ini didasarkan pada model yang menghasilkan performa yang paling baik pada kecamatan pembentuk model terlebih dahulu dengan harapan dapat menghilangkan model dengan performa yang kurang baik. Namun setelah diujicobakan ke data lainnya, diketahui pula bahwa model terpilih ternyata tidak signifikan terhadap hasil yang diinginkan pada uji pembagian data lain dan uji kecamatan lain seperti yang ditunjukkan pada tabel 6.41, dimana secara umum performa yang dihasilkan pada data uji di kecamatan lain lebih rendah daripada hasil di *test-set*. Maka dapat disimpulkan bahwa model dengan hasil akurasi terbaik saat diuji pada *test-set* belum tentu memiliki tingkat *robustness* yang baik jika diuji pada proporsi data lain dan kecamatan yang lain. Meskipun begitu, model tersebut tetap terpilih dengan mempertimbangkan pula hasil plot data dan klasifikasi KLB pada data.

Tabel 6.41 Tingkat Robustness Model Final

Dataran	Robustness Model	
	Terhadap pembagian data	Terhadap kecamatan lain
Rendah	Robust	Robust
Sedang	Bukan yang paling <i>robust</i>	Bukan yang paling <i>robust</i>
Tinggi	Bukan yang paling <i>robust</i>	Robust untuk nilai standar deviasi

Model-model yang terpilih pada tiap dataran didapatkan dari model yang memiliki nilai performa peramalan dan klasifikasi yang baik. Dapat dilihat pada Gambar 6.44 bahwa pada tahap peramalan, performa paling baik ditunjukkan pada model di dataran tinggi, dimana pada dataran tersebut terdapat nilai RMSE dan SMAPE paling kecil. Sedangkan performa model pada tahap peramalan yang paling buruk terdapat pada dataran sedang dengan nilai RMSE dan SMAPE tertinggi. Meskipun begitu, pada tahap klasifikasi (Gambar 6.45), model di dataran

sedang dapat menghasilkan nilai akurasi, spesifisitas, dan sensitivitas yang relatif tinggi dibandingkan dataran lain. Hal ini dapat disebabkan oleh hasil prediksi yang baik pada *test-set*, seperti yang telah dijelaskan pada subbab 6.3.2.



Gambar 6.45 Hasil Performa Klasifikasi tiap Dataran

Selain informasi mengenai hasil model terbaik, dari hasil keseluruhan proses penelitian dapat diketahui beberapa hal lainnya.

1. Dalam pembentukan model XGBoost, skenario pembagian data 60-40 dan 50-50 menghasilkan nilai *error* yang relatif tinggi jika dibandingkan skenario 80-20 dan 70-30. Hal ini dapat disebabkan oleh jumlah data pelatihan yang terlalu sedikit pada skenario 60-40 dan 50-50 cenderung menghasilkan model yang *underfit*, yakni model tidak bisa melihat logika di belakang data, sehingga tidak bisa melakukan peramalan secara tepat dan memiliki nilai *error* yang relatif tinggi.
2. Jumlah variabel independen yang digunakan dalam membangun model peramalan XGBoost dapat mempengaruhi besar *error* yang dihasilkan. Penggunaan banyak variabel independen tidak memastikan akan menghasilkan nilai akurasi yang lebih baik daripada

skenario lainnya. Hal ini disebabkan oleh perbedaan pembobotan tiap fitur pada model yang terbentuk.

3. Setiap skenario pembagian data dan pemilihan fitur akan memiliki nilai *lag* optimal yang berbeda, begitupula dengan nilai *hyper-parameter* yang dihasilkan.
4. Performa peramalan dengan menggunakan algoritma XGBoost lebih baik jika dibandingkan dengan algoritma peramalan yang lebih sederhana seperti Decision Tree dan Linear Regression. Sedangkan, jika dibandingkan dengan algoritma *ensemble* lainnya yakni Random Forest, XGBoost masih menghasilkan nilai *error* yang sedikit lebih tinggi dengan selisih rata-rata RMSE sebesar $\pm 0,1$.
5. Jika dibandingkan dengan metode pengklasikasian secara langsung pada data berlabel KLB, metode gabungan regresi XGBoost dan Klasifikasi Berbasis Aturan menghasilkan nilai akurasi, sensitivitas, dan spesifisitas yang lebih baik dengan selisih nilai pada ketiga faktor tersebut sebesar 20-35%.

6.8. Analisis Hasil Percobaan

Pada sub-bab ini dijelaskan mengenai *insights* yang dapat diambil dari keseluruhan hasil percobaan yang telah dilakukan. Pembahasan pada sub-bab ini dibedakan menjadi dua bagian yakni *insights* dari segi manajerial dan analisis segi metode. bagian Analisis Segi Manajerial difokuskan membahas hasil peramalan yang dilakukan jika dilihat dari sudut pandang manajerial, yakni berisi *insights* yang bermanfaat bagi pemerintah. Sedangkan pada bagian Analisis Segi Metode akan dibahas bagaimana *insights* yang beraikatan dengan data dan metode yang digunakan.

6.8.1. Analisis Segi Manajerial

Hasil peramalan untuk kasus demam berdarah pada tahun 2019 hingga tahun 2020 di Kabupaten Malang ditunjukkan pada Tabel 6.42. Di tabel tersebut dapat terlihat bahwa pada tahun 2020, diramalkan akan terjadi sebanyak 708 kasus demam

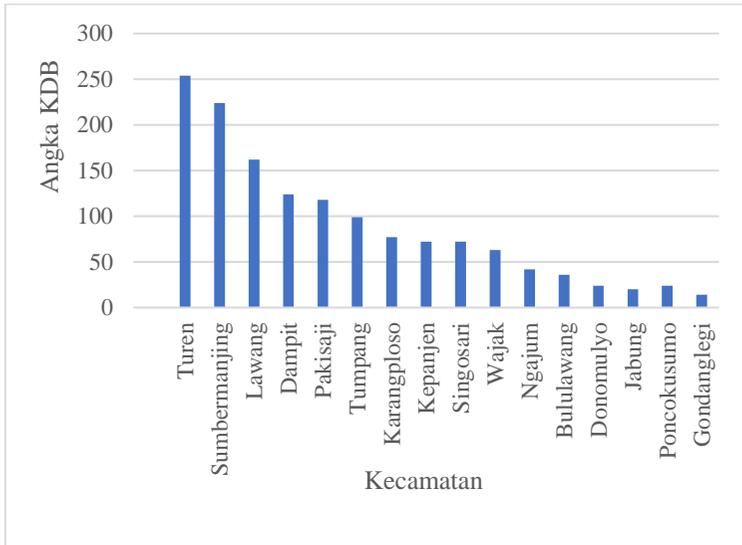
berdarah. Visualisasi persebaran kasus demam berdarah berdasarkan daerah ditunjukkan pada Gambar 6.46.

Tabel 6.42 Angka Kasus Demam Berdarah Per Kecamatan pada Tahun 2019 - 2020

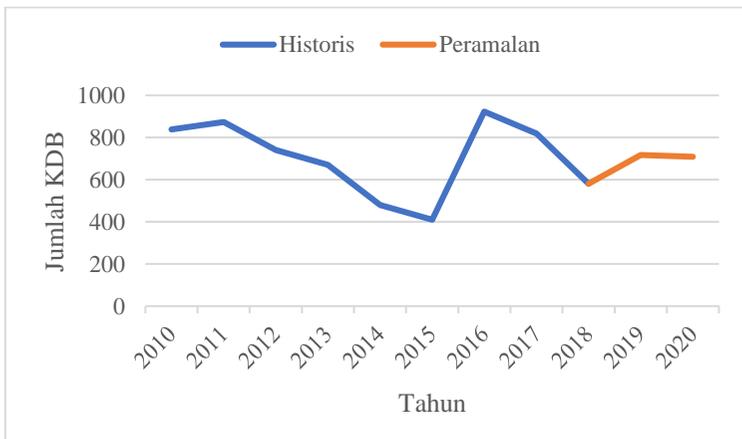
Kecamatan	Angka Kasus DBD		Jumlah
	Tahun 2019	Tahun 2020	
Turen	127	127	254
Sumbermanjing	112	112	224
Lawang	81	81	162
Dampit	62	62	124
Pakisaji	61	57	118
Tumpang	49	50	99
Karangploso	38	39	77
Kepanjen	36	36	72
Singosari	36	36	72
Wajak	32	31	63
Ngajum	22	20	42
Bululawang	18	18	36
Donomulyo	12	12	24
Jabung	12	8	20
Poncokusumo	12	12	24
Gondanglegi	7	7	14

Dari visualisasi pada Gambar 6.46, dapat diketahui bahwa kecamatan yang diramalkan memiliki angka kasus demam berdarah paling tinggi secara kumulatif dari tahun 2019-2020 yakni pada kecamatan Turen dengan angka kasus sebesar 254 kasus hingga akhir tahun 2020, selanjutnya diikuti oleh kecamatan Sumbermanjing dengan 224 kasus, dan kecamatan Lawang dengan 162 kasus. Kemudian diikuti pula oleh kecamatan Dampit dan Pakisaji yang diramalkan memiliki jumlah kasus demam berdarah lebih dari 100 kasus. Sedangkan jika dilihat dari hasil peramalan KLB, terdapat kecamatan yang diramalkan akan mengalami KLB pada bulan November hingga Desember pada tahun 2020 yakni kecamatan Donomulyo. Maka dari itu, pemerintah sebaiknya menjadikan kecamatan-

kecamatan tersebut sebagai daerah prioritas dalam melakukan program-program pencegahan untuk menekan angka kasus demam berdarah.



Gambar 6.46 Persebaran Kejadian Demam Berdarah Berdasarkan Kecamatan (Tahun 2019-2020)

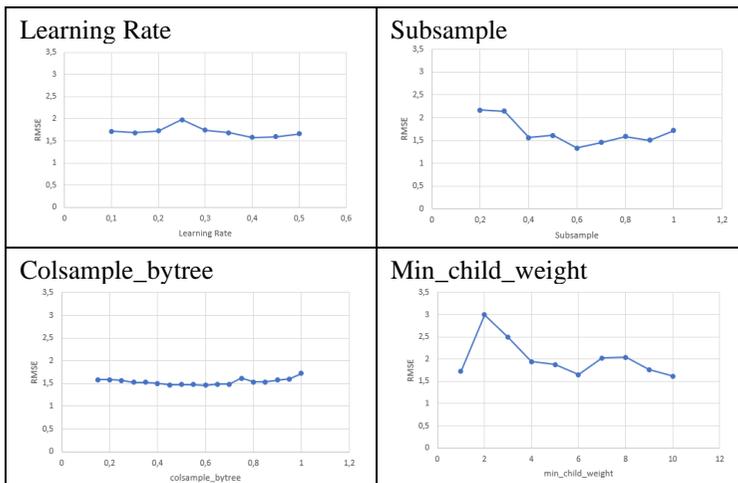


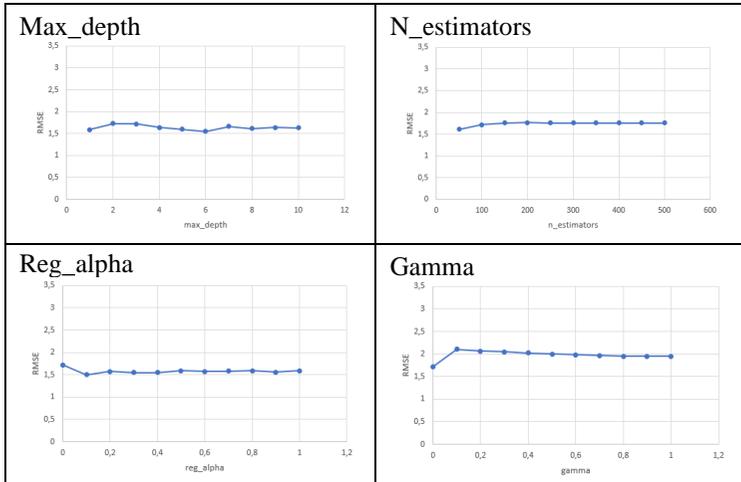
Gambar 6.47 Tren Angka Kejadian Demam Berdarah

Gambar 6.47 menunjukkan tren angka kejadian demam berdarah untuk tahun 2010 hingga 2020. Dari gambar tersebut dapat diketahui ternyata pada tahun 2019 dan 2020 diramalkan masih belum ada penurunan angka kejadian demam berdarah yang signifikan. Namun di sisi lain, angka kasus pada tahun 2019 dan 2020 juga bukan merupakan lonjakan tertinggi dalam kurun waktu 10 tahun terakhir. Hal ini menunjukkan bahwa program-program pemerintah sudah menunjukkan *progres*, namun masih perlu digalakkan kembali untuk tahun-tahun selanjutnya agar penurunan kasus demam berdarah dapat terjadi lebih signifikan.

6.8.2. Analisis Segi Metode

Analisis yang pertama dilakukan dari segi metode adalah melihat parameter-parameter mana saja yang paling berpengaruh dengan performa peramalan dengan metode XGBoost dengan *sensitivity analysis*. Parameter yang akan diujicoba adalah *learning_rate*, *subsample*, *colsample_bytree*, *min_child_weight*, *n_estimators*, *max_depth*, *gamma* dan *reg_alpha*. Hasil dari ujicoba ditunjukkan pada Gambar 6.48.



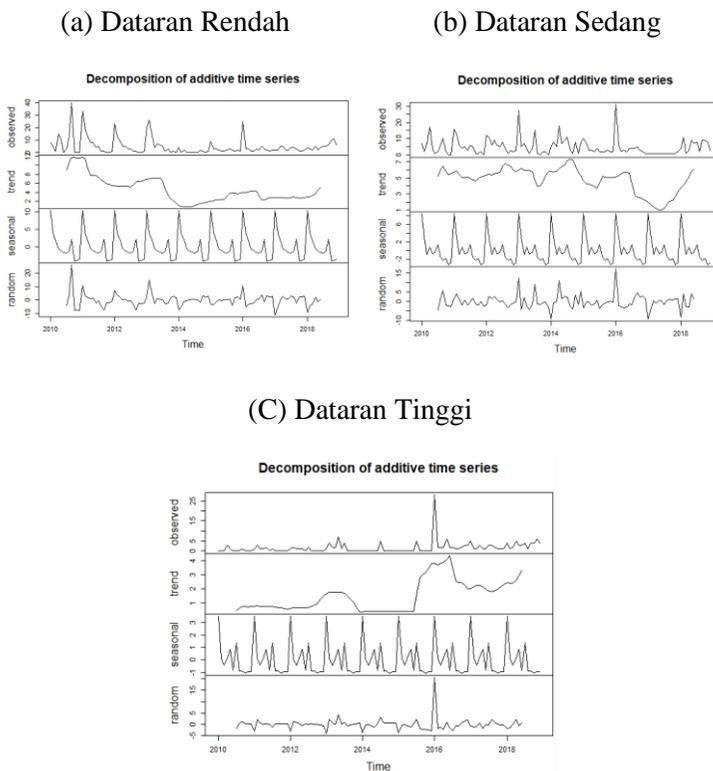


Gambar 6.48 Hasil Sensitivity Analysis pada Parameter XGBoost

Dari gambar 6.48 dapat diketahui bahwa nilai gamma, alpha, $n_estimators$, max_depth , $colsample_bytree$, dan learning rate ternyata memiliki pengaruh yang tidak terlalu signifikan dengan performa yang dihasilkan. Hal ini direpresentasikan dengan nilai RMSE yang cukup stabil saat dilakukan perubahan pada parameter-parameter tersebut. Sedangkan, nilai RMSE terlihat sensitif dengan perubahan parameter min_child_weight dan $subsample$. Seperti yang telah dijelaskan pada sub-bab 4.2.1, $subsample$ merupakan parameter yang mendefinisikan rasio sampel atau baris yang digunakan dalam membangun pohon keputusan setiap ronde *boosting* pada model XGBoost. Hal ini menunjukkan bahwa parameter ini ternyata memiliki sensitivitas yang tinggi terhadap performa yang dihasilkan untuk pembentukan pada data yang relatif sedikit (kurang dari 100 baris). Selain itu, min_child_weight atau parameter yang menentukan jumlah minimum berat (hessian) yang diperlukan pada *child node* juga dapat mempengaruhi performa yang dihasilkan pada data.

Selain itu, dapat diketahui pula bahwa persebaran dan variasi nilai pada data dapat mempengaruhi akurasi peramalan yang

dihasilkan oleh XGBoost, dimana pada studi kasus tugas akhir ini model yang dibangun pada data dengan persebaran data yang kecil menghasilkan performa yang lebih baik. Hal ini mengacu pada hasil performa model di dataran tinggi pada Gambar 6.44, dimana dataran tinggi memiliki persebaran data kejadian demam berdarah paling kecil. Karakteristik persebaran data tersebut dapat direpresentasikan dengan nilai standar deviasi. Dimana pada kecamatan pembentuk model di dataran tinggi, nilai standar deviasi yang dihasilkan adalah 3,05 dimana kedua dataran lainnya memiliki standar deviasi yang lebih besar yakni 5,10 dan 6,84.



Gambar 6.49 Dekomposisi Tren & Siklus pada Dataran (a) Rendah (b) Sedang (c) Tinggi

Pada Gambar 6.49 ditunjukkan hasil dekomposisi tren dan musiman data untuk kecamatan pembentuk model di dataran rendah, sedang, dan tinggi. Dari visualisasi tersebut, dapat diketahui bahwa ketiga data tersebut memiliki karakteristik yang hampir sama yakni memiliki unsur musiman dan tren. Hal yang berbeda dapat dilihat pada lonjakan yang terjadi di tiap dataran, dimana pada dataran tinggi lonjakan kejadian demam berdarah relatif berada pada frekuensi yang tidak banyak jika dibandingkan dengan kedua dataran lainnya. Hal ini dapat mendukung nilai *error* peramalan yang kecil, karena semakin banyak lonjakan maka semakin tinggi pula kemungkinan selisih data aktual dan peramalan yang dihasilkan.

Halaman ini sengaja dikosongkan

BAB VII

KESIMPULAN DAN SARAN

Bab ini membahas tentang kesimpulan dari semua proses yang telah dilakukan dan saran yang dapat diberikan untuk pengembangan penelitian selanjutnya.

7.1. Kesimpulan

Kesimpulan yang didapatkan dari proses pengerjaan tugas akhir yang telah dilakukan antara lain :

1. Pada dataran rendah, gabungan metode XGBoost dan Klasifikasi Berbasis Aturan yang digunakan untuk meramalkan KLB demam berdarah menghasilkan model terbaik dengan skenario pembagian proporsi data 70% - 30%, serta melibatkan variabel jumlah penduduk, kelembapan udara dan 11 fitur *lag* tambahan sebagai variabel independen untuk *input* proses peramalan. Pada tahap peramalan, dihasilkan nilai RMSE sebesar 2,532 dan SMAPE sebesar 27,57%. Sedangkan pada tahap klasifikasi, dihasilkan nilai akurasi sebesar 83,33%, sensitivitas sebesar 50,0%, dan nilai spesifisitas sebesar 88,5%.
2. Pada dataran sedang, gabungan metode XGBoost dan Klasifikasi Berbasis Aturan yang digunakan untuk meramalkan KLB demam berdarah menghasilkan model terbaik dengan skenario pembagian proporsi data 70% - 30%, serta melibatkan 3 fitur *lag* tambahan sebagai variabel independen untuk *input* proses peramalan. Pada tahap peramalan, dihasilkan nilai RMSE sebesar 2,899 dan SMAPE sebesar 35,99%. Sedangkan pada tahap klasifikasi, dihasilkan nilai akurasi sebesar 93,75%, sensitivitas sebesar 100,00%, dan nilai spesifisitas sebesar 90,48%.
3. Pada dataran tinggi, gabungan metode XGBoost dan Klasifikasi Berbasis Aturan yang digunakan untuk

meramalkan KLB demam berdarah menghasilkan model terbaik dengan skenario pembagian proporsi data 70% - 30%, serta melibatkan variabel suhu, curah hujan, kelembapan udara, kecepatan angin, jumlah penduduk, Angka Bebas Jentik (ABJ), dan 12 fitur *lag* tambahan sebagai variabel independen untuk *input* proses peramalan. Pada tahap peramalan, dihasilkan nilai RMSE sebesar 1,446 dan SMAPE sebesar 24,13%. Sedangkan pada tahap klasifikasi, dihasilkan nilai akurasi sebesar 86,21%, sensitivitas sebesar 66,67%, dan nilai spesifisitas sebesar 100,00%.

4. Hasil peramalan hingga akhir tahun 2020 dengan menggunakan metode XGBoost-Klasifikasi Berbasis Aturan menunjukkan bahwa belum adanya penurunan angka kejadian demam berdarah yang signifikan. Terhitung hingga akhir tahun 2020, diramalkan akan terjadi kasus demam berdarah secara kumulatif dalam setahun pada 16 kecamatan di Kabupaten Malang adalah 708 kasus. Sehingga pemerintah masih perlu menggalakkan kembali program-program pencegahan dan penanganan yang diprioritaskan pada kecamatan-kecamatan yang diramalkan akan memiliki jumlah kasus demam berdarah tertinggi yakni Turen, Sumbermanjing dan Lawang serta kecamatan yang diramalkan akan terjadi KLB pada akhir tahun 2020 yakni Donomulyo.
5. Persebaran dan variasi nilai pada data kecamatan pembentuk model dapat mempengaruhi hasil kinerja peramalan dengan menggunakan algoritma XGBoost. Sedangkan, parameter yang sensitif terhadap performa model XGBoost pada tugas akhir ini adalah *min_child_weight* dan *subsample*.

7.2. Saran

Dari pengerjaan tugas akhir ini, adapun beberapa saran untuk pengembangan penelitian ke depan.

1. Penambahan jumlah periode atau baris data yang digunakan sebagai data pelatihan, sehingga logika dibalik data dapat dilihat lebih baik.
2. Penambahan atau penggantian variabel bebas yang digunakan, termasuk memilih variabel independen yang berkorelasi kuat terhadap fluktuasi jumlah kejadian demam berdarah. Hal ini dimaksudkan agar penelitian selanjutnya dapat menghasilkan peramalan yang benar-benar akurat dan representatif.
3. Melakukan perbandingan dengan penggunaan algoritma optimasi lain untuk mengoptimasi parameter-parameter yang ada pada metode XGBoost, agar memastikan parameter yang digunakan sudah cukup optimal.

Halaman ini sengaja dikosongkan

DAFTAR PUSTAKA

- [1] Pusat Data dan Informasi Kementerian Kesehatan RI, “InfoDatin : Situasi Penyakit Demam Berdarah di Indonesia Tahun 2017,” 22 April 2018. [Online]. Available: <http://www.depkes.go.id/>. [Diakses 30 October 2019].
- [2] Kementerian Kesehatan Republik Indonesia, “Kasus DBD Terus Bertambah, Anung Imbau Masyarakat Maksimalkan PSN,” Kementerian Kesehatan Republik Indonesia, 4 Februari 2019. [Online]. Available: kemkes.go.id/article/view/19020600004/kasus-dbd-terus-bertambah-anung-imbau-masyarakat-maksimalkan-psn.html. [Diakses 30 Oktober 2019].
- [3] B. Widhayasa, “5 Wabah Penyakit dengan Status Kejadian Luar Biasa di Indonesia,” IDN Times, 13 Juni 2019. [Online]. Available: <https://www.idntimes.com/science/discovery/bayu-widhayasa/5-wabah-penyakit-dengan-status-kejadian-luar-biasa-di-indonesia-exp-c1c2>. [Diakses 30 Oktober 2019].
- [4] Buku Pedoman Penyelidikan dan Penanggulangan Kejadian Luar Biasa Penyakit Menular dan Keracunan Pangan, Kementerian Kesehatan Republik Indonesia, 2011.
- [5] A. Gama dan B. Faizah, “Analisis Faktor Risiko Kejadian DBD di Desa Mojosongo,” *Eksplanasi*, vol. V, no. 2, 2010.
- [6] S. Desi, “Meningkat, Kabupaten Tertinggi Se-Jatim,” RADARMALANG.ID, 1 Februari 2019. [Online]. Available: <https://radarmalang.id/meningkat-kabupaten-tertinggi-se-jatim/>. [Diakses 30 Oktober 2019].

- [7] Y. Yusof dan Z. Mustaffa, "Dengue Outbreak Prediction: A Least Squares Support Vector Machines," Januari 2011.
- [8] J. Nan, L. Xianyi, C. Jie, C. Xiangping, D. Guanghui, L. Kangkang dan H. Gang, "Using Climate Factors to Predict the Outbreak of Dengue Fever," dalam *2018 7th International Conference on Digital Home (ICDH)*, China, 2018.
- [9] S. Chen, X. Jian, C. Lili, Z. Xi, Z. Li dan L. Jinfeng, "A Regularization-Based eXtreme Gradient Boosting Approach in Foodborne Disease," *MEDINFO 2019: Health and Wellbeing e-Networks for All*, 2019.
- [10] R. Badkundri, V. Valbuena, S. Pinnamareddy, B. Cantrell dan J. Standeven, "Forecasting the Yemen Cholera Outbreak with Machine Learning," 2018.
- [11] V. Morde, "XGBoost Algorithm: Long May She Reign!," 8 April 2019. [Online]. Available: <https://towardsdatascience.com/https-medium-com-vishalorde-xgboost-algorithm-long-she-may-rein-edd9f99be63d>. [Diakses 30 Oktober 2019].
- [12] E. W. Sulistyowati, "Peramalan KLB Campak Menggunakan Gabungan Metode JST Backpropagation dan CART," *IJCCS*, vol. 8, no. 1, pp. 49-58, 2014.
- [13] Misnadiarly, *Demam Berdarah Dengue (DBD)*, Jakarta: Pustaka Populer Obor, 2019.
- [14] V. Kotu dan B. Deshpande, "Time Series Forecasting," dalam *Data Science: Concepts and Practice*, 2nd penyunt., Cambridge, Morgan Kaufmann, 2019, pp. 430-439.
- [15] O. A. S. dan I. I. A. , "A Comparative Study between Univariate and Multivariate Linear Stationary Time Series Models," *American Journal of Mathematics and Statistics*, vol. VI, no. 5, pp. 203-212, 2016.
- [16] H. Junaedi, H. Budiarto, I. Maryati dan Y. Melani, "Data Transformation Pada Data Mining," dalam *Prosiding*

Konferensi Nasional “Inovasi dalam Desain dan Teknologi”, 2011.

- [17] J. Han, M. Kamber dan J. Pei, *Data Mining : Concepts and Techniques*, 3rd penyunt., Waltham, Massachusetts: Elsevier, 2012, pp. 96-114.
- [18] Z.-H. Zhou, *Ensemble Methods Foundations and Algorithms*, Cambridge: Chapman & Hall/CRC, 2012.
- [19] xgboost developers, “Introduction to Boosted Trees,” [Online]. Available: <https://xgboost.readthedocs.io/en/latest/tutorials/model.html>. [Diakses 30 Mei 2020].
- [20] Oracle Help Center, “Data Mining Concepts,” Oracle, [Online]. Available: https://docs.oracle.com/cd/B28359_01/datamine.111/b28129/classify.htm#DMCON035. [Diakses 30 Oktober 2019].
- [21] T. A.K.H., “Rule-based Classification,,” dalam *LIU L., ÖZSU M.T. (eds) Encyclopedia of Database Systems*, Boston, Springer, 2009.
- [22] J. Heizer dan B. Render, *Operations Management*. Tenth Edition., New Jersey: Pearson, 2011.
- [23] A. H. Hakim dan Y. Prasetyawan, *Perencanaan dan Pengendalian Produksi*, Yogyakarta: Graha Ilmu, 2008.
- [24] V. Kreinovich, H. T. Nguyen dan R. Ouncharoen, “How to Estimate Forecasting Quality: A System-Motivated Derivation of Symmetric Mean Absolute Percentage Error (SMAPE) and Other Similar Characteristics,” *Semantic Scholar*, Mexico, 2014.
- [25] M. Sokolova dan G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Information Processing and Management*, vol. 45, no. 4, pp. 427-437, 2009.
- [26] XGBoost Developers, “XGBoost Parameters,” [Online]. Available:

<https://xgboost.readthedocs.io/en/latest/parameter.html#>.
[Diakses 10 November 2019].

- [27] xgboost developers, “Python API Reference,” [Online]. Available:
https://xgboost.readthedocs.io/en/latest/python/python_api.html. [Diakses 20 Mei 2020].
- [28] W. A. I Putu Agus Aditya Pramana, “Peramalan Jumlah Kasus Demam Berdarah di Kabupaten Malang Menggunakan Metode Fuzzy Inference System,” *JURNAL TEKNIK ITS*, vol. V, no. 1, 2016.
- [29] M. G. Dunn, *Exploring Your World: The Adventure of Geography*, Washington, D.C.: National Geographic Society, 1993.

BIODATA PENULIS



Humaira Nur Pradani (Mira), lahir di Situbondo pada tanggal 18 April 1998 merupakan anak ketiga dari tiga bersaudara, pasangan dari Bapak Andi Prahjono dan Ibu Dewi Ramdani. Penulis menempuh pendidikan dimulai di TK Al-Hikmah Surabaya (lulus tahun 2004), SD Al-Hikmah Surabaya (lulus tahun 2010), SMP Hikmah Surabaya (lulus tahun 2013), SMA Al-Hikmah Surabaya (lulus tahun 2016), hingga akhirnya bisa

menempuh masa kuliah di Fakultas Teknologi Elektro dan Informatika Cerdas (F-Electics) Institut Teknologi Sepuluh Nopember Surabaya.

Sejak tahun 2017, penulis dipercaya untuk menjadi salah satu *awardee* Beasiswa Unggulan oleh Kementerian Pendidikan dan Kebudayaan. Selama masa perkuliahannya, penulis aktif mengikuti berbagai kegiatan kemahasiswaan baik dalam bentuk kepanitiaan maupun organisasi. Dalam dunia organisasi, penulis pernah menjabat sebagai staff dan sekretaris di Departemen Kesejahteraan Mahasiswa HMSI (Himpunan Mahasiswa Sistem Informasi) ITS. Selain itu, dalam dunia akademik, penulis juga aktif menjadi asisten dosen pada mata kuliah Statistika, Teknik Peramalan dan Sistem Enterprise.

Semoga dengan adanya penulisan tugas akhir ini mampu memberikan kontribusi yang positif bagi dunia pendidikan. Jika ingin memberikan saran terkait tugas akhir ini, dapat menghubungi penulis melalui email hnpradani@gmail.com.

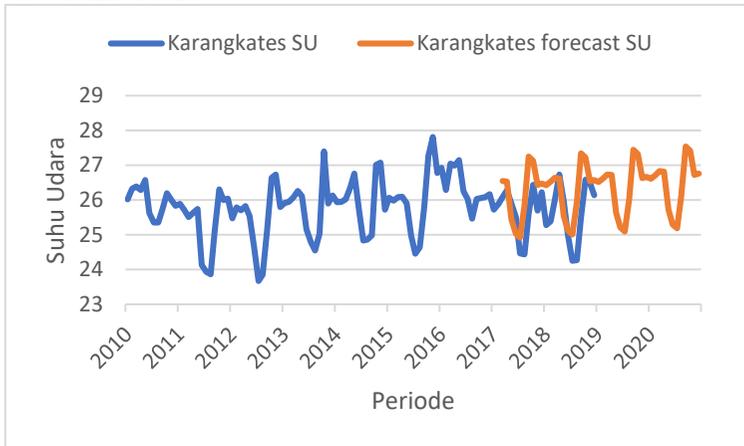
Halaman ini sengaja dikosongkan

LAMPIRAN A

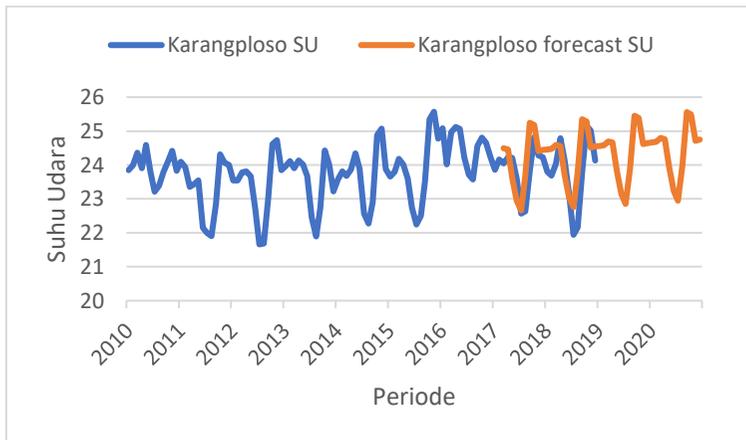
Gambar plot peramalan 24 periode mendatang untuk variabel independen yang digunakan dalam tugas akhir.

1. Variabel Independen Data Iklim

a. Suhu Udara

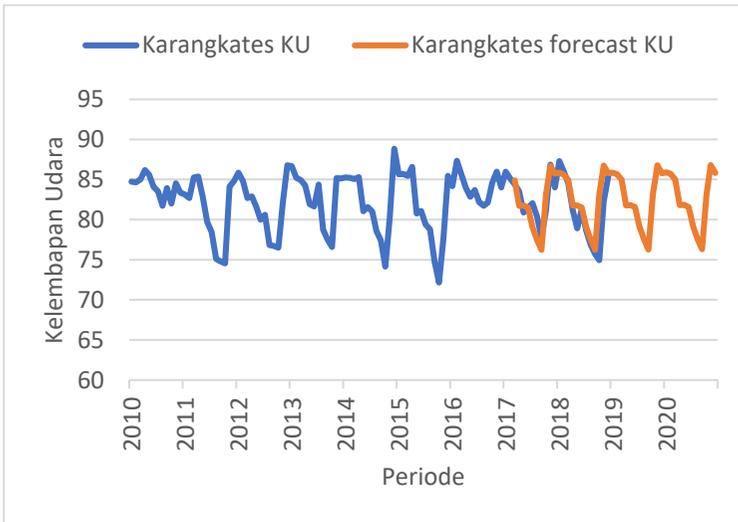


Gambar A.1 Peramalan Suhu Udara dengan Metode Dekomposisi Multiplikatif (Stasiun Karangkates)

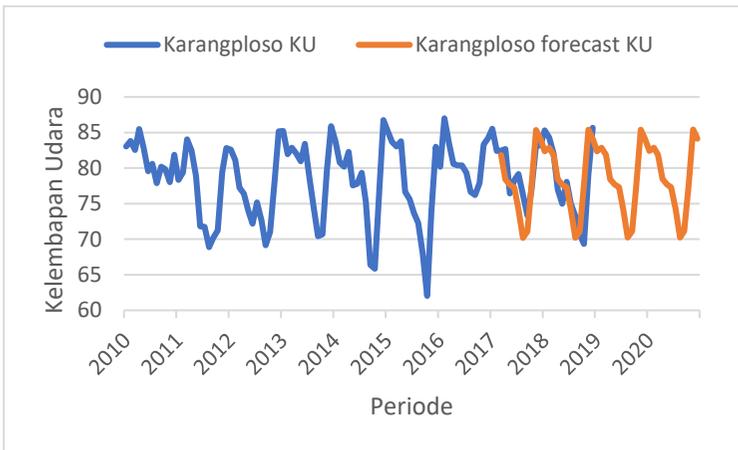


Gambar A.2 Peramalan Suhu Udara dengan Metode Dekomposisi Multiplikatif (Stasiun Karangploso)

b. Kelembapan Udara

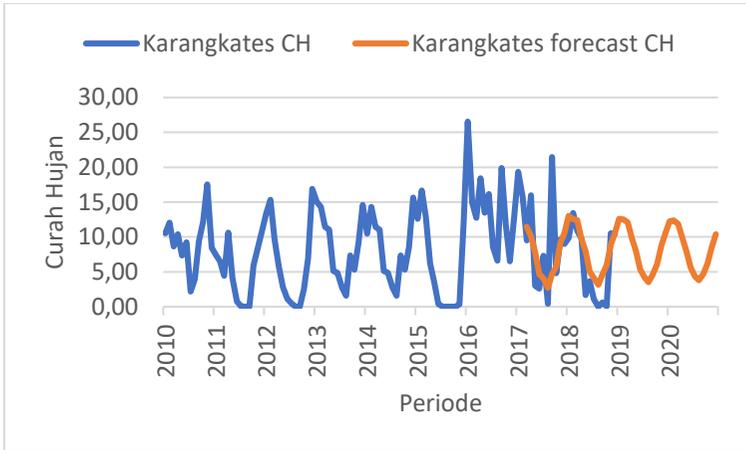


Gambar A.3 Peramalan Kelembapan dengan Metode Dekomposisi Multiplikatif (Stasiun Karangates)

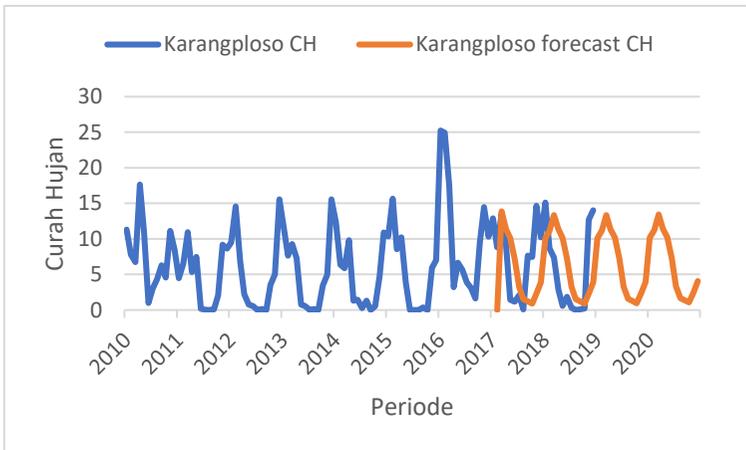


Gambar A.4 Peramalan Kelembapan dengan Metode Dekomposisi Multiplikatif (Stasiun Karangploso)

c. Curah Hujan

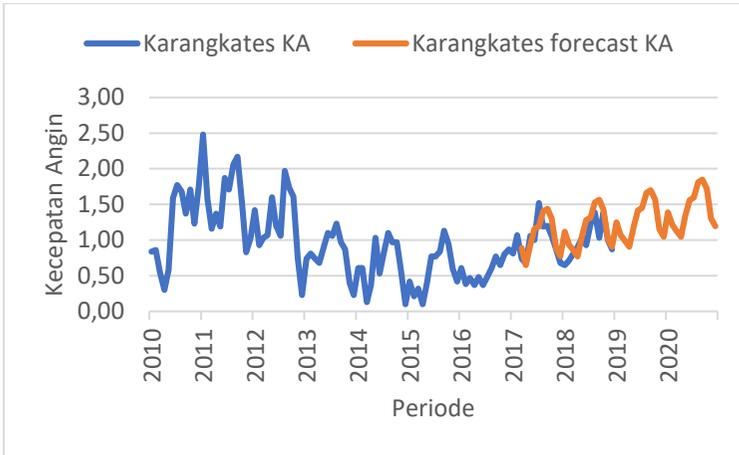


Gambar A.5 Peramalan Curah Hujan dengan Metode ARIMA (Stasiun Karangates)

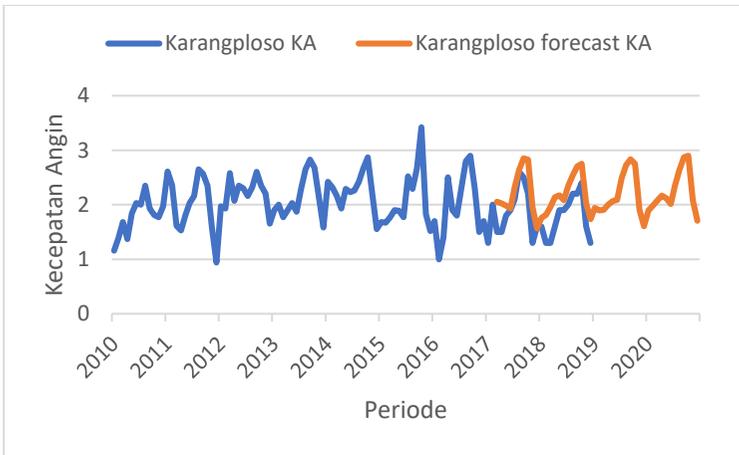


Gambar A.6 Peramalan Curah Hujan dengan Metode ARIMA (Stasiun Karangploso)

d. Kecepatan Angin

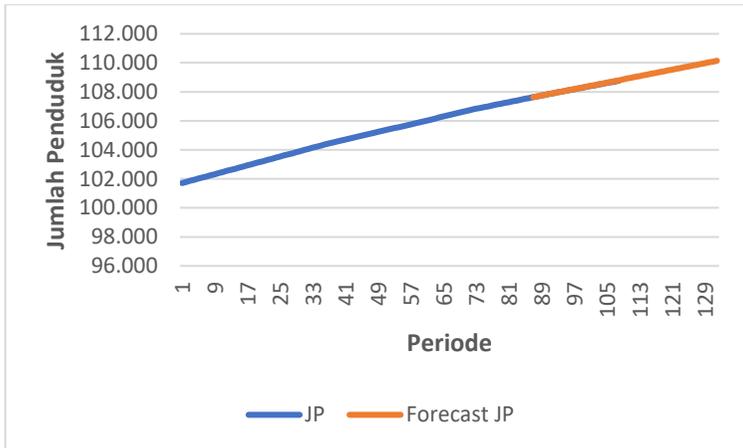


Gambar A.7 Peramalan Kecepatan Angin dengan Metode ARIMA (Stasiun Karangates)



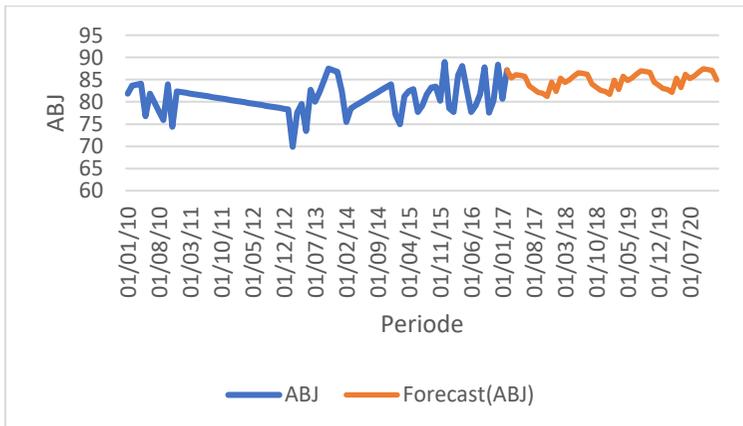
Gambar A.8 Peramalan Kecepatan Angin Metode ARIMA (Stasiun Karangploso)

- Variabel Independen Jumlah Penduduk (contoh pada Kecamatan Kepanjen)



Gambar A.9 Peramalan Jumlah Penduduk dengan Metode Double Exponential Smoothing

- Variabel Independen Angka Bebas Jentik (contoh pada Kecamatan Kepanjen)



Gambar A.10 Peramalan Angka Bebas Jentik dengan Metode Exponential Smoothing

Halaman ini sengaja dikosongkan

LAMPIRAN B

Hasil Peramalan 24 Periode Kedepan

1. Dataran Rendah

Tabel B.1 Hasil Peramalan menggunakan XGBoost pada Dataran Rendah

a. Kecamatan Kepanjen

Periode	Jumlah KDB	Status
Jan-19	5	non-KLB
Feb-19	5	non-KLB
Mar-19	2	non-KLB
Apr-19	4	non-KLB
Mei-19	4	non-KLB
Jun-19	1	non-KLB
Jul-19	1	non-KLB
Agt-19	1	non-KLB
Sep-19	1	non-KLB
Okt-19	2	non-KLB
Nov-19	5	non-KLB
Des-19	5	non-KLB

Periode	Jumlah KDB	Status
Jan-20	5	non-KLB
Feb-20	5	non-KLB
Mar-20	2	non-KLB
Apr-20	4	non-KLB
Mei-20	4	non-KLB
Jun-20	1	non-KLB
Jul-20	1	non-KLB
Agt-20	1	non-KLB
Sep-20	1	non-KLB
Okt-20	2	non-KLB
Nov-20	5	non-KLB
Des-20	5	non-KLB

b. Kecamatan Turen

Periode	Jumlah KDB	Status
Jan-19	17	KLB
Feb-19	17	KLB
Mar-19	7	non-KLB
Apr-19	13	KLB
Mei-19	13	KLB
Jun-19	5	non-KLB
Jul-19	5	non-KLB
Agt-19	5	non-KLB
Sep-19	5	non-KLB
Okt-19	6	non-KLB
Nov-19	17	KLB
Des-19	17	KLB

Periode	Jumlah KDB	Status
Jan-20	17	non-KLB
Feb-20	17	non-KLB
Mar-20	7	non-KLB
Apr-20	13	non-KLB
Mei-20	13	non-KLB
Jun-20	5	non-KLB
Jul-20	5	non-KLB
Agt-20	5	non-KLB
Sep-20	5	non-KLB
Okt-20	6	non-KLB
Nov-20	17	non-KLB
Des-20	17	non-KLB

c. Kecamatan Gondanglegi

Periode	Jumlah KDB	Status
Jan-19	1	non-KLB
Feb-19	1	non-KLB
Mar-19	1	non-KLB
Apr-19	1	non-KLB

Periode	Jumlah KDB	Status
Jan-20	1	non-KLB
Feb-20	1	non-KLB
Mar-20	1	non-KLB
Apr-20	1	non-KLB

Periode	Jumlah KDB	Status
Mei-19	1	non-KLB
Jun-19	0	non-KLB
Jul-19	0	non-KLB
Agt-19	0	non-KLB
Sep-19	0	non-KLB
Okt-19	0	non-KLB
Nov-19	1	non-KLB
Des-19	1	non-KLB

Periode	Jumlah KDB	Status
Mei-20	1	non-KLB
Jun-20	0	non-KLB
Jul-20	0	non-KLB
Agt-20	0	non-KLB
Sep-20	0	non-KLB
Okt-20	0	non-KLB
Nov-20	1	non-KLB
Des-20	1	non-KLB

d. Kecamatan Donomulyo

Periode	Jumlah KDB	Status
Jan-19	2	non-KLB
Feb-19	2	non-KLB
Mar-19	1	non-KLB
Apr-19	1	non-KLB
Mei-19	1	non-KLB
Jun-19	0	non-KLB
Jul-19	0	non-KLB
Agt-19	0	non-KLB
Sep-19	0	non-KLB

Periode	Jumlah KDB	Status
Jan-20	2	KLB
Feb-20	2	KLB
Mar-20	1	non-KLB
Apr-20	1	non-KLB
Mei-20	1	non-KLB
Jun-20	0	non-KLB
Jul-20	0	non-KLB
Agt-20	0	non-KLB
Sep-20	0	non-KLB

Periode	Jumlah KDB	Status
Okt-19	1	non-KLB
Nov-19	2	non-KLB
Des-19	2	non-KLB

Periode	Jumlah KDB	Status
Okt-20	1	non-KLB
Nov-20	2	KLB
Des-20	2	KLB

e. Kecamatan Bululawang

Periode	Jumlah KDB	Status
Jan-19	2	non-KLB
Feb-19	2	non-KLB
Mar-19	1	non-KLB
Apr-19	2	non-KLB
Mei-19	2	non-KLB
Jun-19	1	non-KLB
Jul-19	1	non-KLB
Agt-19	1	non-KLB
Sep-19	1	non-KLB
Okt-19	1	non-KLB
Nov-19	2	non-KLB
Des-19	2	non-KLB

Periode	Jumlah KDB	Status
Jan-20	2	non-KLB
Feb-20	2	non-KLB
Mar-20	1	non-KLB
Apr-20	2	non-KLB
Mei-20	2	non-KLB
Jun-20	1	non-KLB
Jul-20	1	non-KLB
Agt-20	1	non-KLB
Sep-20	1	non-KLB
Okt-20	1	non-KLB
Nov-20	2	non-KLB
Des-20	2	non-KLB

2. Dataran Sedang

Tabel B.2 Hasil Peramalan menggunakan XGBoost pada Dataran Sedang

a. Kecamatan Pakisaji

Periode	Jumlah KDB	Status
Jan-19	5	non-KLB
Feb-19	5	non-KLB
Mar-19	8	non-KLB
Apr-19	4	non-KLB
Mei-19	6	non-KLB
Jun-19	4	non-KLB
Jul-19	5	non-KLB
Agt-19	4	non-KLB
Sep-19	5	non-KLB
Okt-19	6	non-KLB
Nov-19	4	non-KLB
Des-19	5	non-KLB

Periode	Jumlah KDB	Status
Jan-20	4	non-KLB
Feb-20	5	non-KLB
Mar-20	6	non-KLB
Apr-20	4	non-KLB
Mei-20	5	non-KLB
Jun-20	4	non-KLB
Jul-20	5	non-KLB
Agt-20	6	non-KLB
Sep-20	4	non-KLB
Okt-20	5	non-KLB
Nov-20	4	non-KLB
Des-20	5	non-KLB

b. Kecamatan Dampit

Periode	Jumlah KDB	Status
Jan-19	6	non-KLB
Feb-19	4	non-KLB

Periode	Jumlah KDB	Status
Jan-20	5	non-KLB
Feb-20	5	non-KLB

Periode	Jumlah KDB	Status
Mar-19	5	non-KLB
Apr-19	5	non-KLB
Mei-19	5	non-KLB
Jun-19	6	non-KLB
Jul-19	5	non-KLB
Agt-19	5	non-KLB
Sep-19	5	non-KLB
Okt-19	5	non-KLB
Nov-19	6	non-KLB
Des-19	5	non-KLB

Periode	Jumlah KDB	Status
Mar-20	5	non-KLB
Apr-20	6	non-KLB
Mei-20	5	non-KLB
Jun-20	5	non-KLB
Jul-20	5	non-KLB
Agt-20	5	non-KLB
Sep-20	6	non-KLB
Okt-20	5	non-KLB
Nov-20	5	non-KLB
Des-20	5	non-KLB

c. Kecamatan Wajak

Periode	Jumlah KDB	Status
Jan-19	2	non-KLB
Feb-19	3	non-KLB
Mar-19	2	non-KLB
Apr-19	3	non-KLB
Mei-19	3	non-KLB
Jun-19	3	non-KLB

Periode	Jumlah KDB	Status
Jan-20	2	non-KLB
Feb-20	3	non-KLB
Mar-20	2	non-KLB
Apr-20	3	non-KLB
Mei-20	3	non-KLB
Jun-20	2	non-KLB

Periode	Jumlah KDB	Status
Jul-19	3	non-KLB
Agt-19	2	non-KLB
Sep-19	3	non-KLB
Okt-19	2	non-KLB
Nov-19	3	non-KLB
Des-19	3	non-KLB

Periode	Jumlah KDB	Status
Jul-20	3	non-KLB
Agt-20	2	non-KLB
Sep-20	3	non-KLB
Okt-20	3	non-KLB
Nov-20	2	non-KLB
Des-20	3	non-KLB

d. Kecamatan Singosari

Periode	Jumlah KDB	Status
Jan-19	4	non-KLB
Feb-19	3	non-KLB
Mar-19	3	non-KLB
Apr-19	3	non-KLB
Mei-19	3	non-KLB
Jun-19	3	non-KLB
Jul-19	3	non-KLB
Agt-19	2	non-KLB
Sep-19	3	non-KLB
Okt-19	3	non-KLB

Periode	Jumlah KDB	Status
Jan-20	3	non-KLB
Feb-20	3	non-KLB
Mar-20	3	non-KLB
Apr-20	3	non-KLB
Mei-20	3	non-KLB
Jun-20	3	non-KLB
Jul-20	3	non-KLB
Agt-20	3	non-KLB
Sep-20	3	non-KLB
Okt-20	3	non-KLB

Periode	Jumlah KDB	Status
Nov-19	3	non-KLB
Des-19	3	non-KLB

Periode	Jumlah KDB	Status
Nov-20	3	non-KLB
Des-20	3	non-KLB

e. Kecamatan Tumpang

Periode	Jumlah KDB	Status
Jan-19	2	non-KLB
Feb-19	4	KLB
Mar-19	4	KLB
Apr-19	6	KLB
Mei-19	3	non-KLB
Jun-19	4	KLB
Jul-19	4	KLB
Agt-19	5	KLB
Sep-19	5	KLB
Okt-19	4	KLB
Nov-19	5	KLB
Des-19	3	non-KLB

Periode	Jumlah KDB	Status
Jan-20	4	non-KLB
Feb-20	4	non-KLB
Mar-20	4	non-KLB
Apr-20	5	non-KLB
Mei-20	4	non-KLB
Jun-20	4	non-KLB
Jul-20	4	non-KLB
Agt-20	4	non-KLB
Sep-20	5	non-KLB
Okt-20	4	non-KLB
Nov-20	4	non-KLB
Des-20	4	non-KLB

f. Kecamatan Sumbermanjing

Periode	Jumlah KDB	Status
Jan-19	12	KLB
Feb-19	13	KLB
Mar-19	6	KLB
Apr-19	6	KLB
Mei-19	7	KLB
Jun-19	12	KLB
Jul-19	12	KLB
Agt-19	13	KLB
Sep-19	6	KLB
Okt-19	6	KLB
Nov-19	7	KLB
Des-19	12	KLB

Periode	Jumlah KDB	Status
Jan-20	12	non-KLB
Feb-20	13	non-KLB
Mar-20	6	non-KLB
Apr-20	6	non-KLB
Mei-20	7	non-KLB
Jun-20	12	non-KLB
Jul-20	12	non-KLB
Agt-20	13	non-KLB
Sep-20	6	non-KLB
Okt-20	6	non-KLB
Nov-20	7	non-KLB
Des-20	12	non-KLB

g. Kecamatan Lawang

Periode	Jumlah KDB	Status
Jan-19	5	KLB
Feb-19	6	KLB
Mar-19	6	KLB
Apr-19	10	KLB

Periode	Jumlah KDB	Status
Jan-20	7	non-KLB
Feb-20	6	non-KLB
Mar-20	7	non-KLB
Apr-20	8	non-KLB

Periode	Jumlah KDB	Status
Mei-19	6	KLB
Jun-19	7	KLB
Jul-19	6	KLB
Agt-19	7	KLB
Sep-19	8	KLB
Okt-19	7	KLB
Nov-19	8	KLB
Des-19	5	KLB

Periode	Jumlah KDB	Status
Mei-20	6	non-KLB
Jun-20	7	non-KLB
Jul-20	6	non-KLB
Agt-20	7	non-KLB
Sep-20	8	non-KLB
Okt-20	6	non-KLB
Nov-20	7	non-KLB
Des-20	6	non-KLB

h. Kecamatan Karangploso

Periode	Jumlah KDB	Status
Jan-19	3	non-KLB
Feb-19	4	non-KLB
Mar-19	3	non-KLB
Apr-19	4	non-KLB
Mei-19	2	non-KLB
Jun-19	3	non-KLB
Jul-19	3	non-KLB
Agt-19	3	non-KLB
Sep-19	4	non-KLB

Periode	Jumlah KDB	Status
Jan-20	3	non-KLB
Feb-20	4	non-KLB
Mar-20	3	non-KLB
Apr-20	3	non-KLB
Mei-20	3	non-KLB
Jun-20	3	non-KLB
Jul-20	4	non-KLB
Agt-20	3	non-KLB
Sep-20	3	non-KLB

Periode	Jumlah KDB	Status
Okt-19	3	non-KLB
Nov-19	3	non-KLB
Des-19	3	non-KLB

Periode	Jumlah KDB	Status
Okt-20	3	non-KLB
Nov-20	3	non-KLB
Des-20	4	non-KLB

3. Dataran Tinggi

Tabel B.3 Hasil Peramalan menggunakan XGBoost pada Dataran Tinggi

a. Kecamatan Ngajum

Periode	Jumlah KDB	Status
Jan-19	2	non-KLB
Feb-19	2	non-KLB
Mar-19	2	non-KLB
Apr-19	2	non-KLB
Mei-19	2	non-KLB
Jun-19	2	non-KLB
Jul-19	2	non-KLB
Agt-19	2	non-KLB
Sep-19	1	non-KLB
Okt-19	1	non-KLB
Nov-19	2	non-KLB
Des-19	2	non-KLB

Periode	Jumlah KDB	Status
Jan-20	2	non-KLB
Feb-20	2	non-KLB
Mar-20	2	non-KLB
Apr-20	2	non-KLB
Mei-20	2	non-KLB
Jun-20	2	non-KLB
Jul-20	2	non-KLB
Agt-20	2	non-KLB
Sep-20	1	non-KLB
Okt-20	1	non-KLB
Nov-20	1	non-KLB
Des-20	1	non-KLB

b. Kecamatan Poncokusumo

Periode	Jumlah KDB	Status
Jan-19	1	non-KLB
Feb-19	1	non-KLB
Mar-19	1	non-KLB
Apr-19	1	non-KLB
Mei-19	1	non-KLB
Jun-19	1	non-KLB
Jul-19	1	non-KLB
Agt-19	1	non-KLB
Sep-19	1	non-KLB
Okt-19	1	non-KLB
Nov-19	1	non-KLB
Des-19	1	non-KLB

Periode	Jumlah KDB	Status
Jan-20	1	non-KLB
Feb-20	1	non-KLB
Mar-20	1	non-KLB
Apr-20	1	non-KLB
Mei-20	1	non-KLB
Jun-20	1	non-KLB
Jul-20	1	non-KLB
Agt-20	1	non-KLB
Sep-20	1	non-KLB
Okt-20	1	non-KLB
Nov-20	1	non-KLB
Des-20	1	non-KLB

c. Kecamatan Jabung

Periode	Jumlah KDB	Status
Jan-19	1	non-KLB
Feb-19	1	non-KLB
Mar-19	1	non-KLB
Apr-19	1	non-KLB

Periode	Jumlah KDB	Status
Jan-20	1	non-KLB
Feb-20	1	non-KLB
Mar-20	1	non-KLB
Apr-20	1	non-KLB

Periode	Jumlah KDB	Status
Mei-19	1	non-KLB
Jun-19	1	non-KLB
Jul-19	1	non-KLB
Agt-19	1	non-KLB
Sep-19	1	non-KLB
Okt-19	1	non-KLB
Nov-19	1	non-KLB
Des-19	1	non-KLB

Periode	Jumlah KDB	Status
Mei-20	0	non-KLB
Jun-20	0	non-KLB
Jul-20	0	non-KLB
Agt-20	1	non-KLB
Sep-20	0	non-KLB
Okt-20	1	non-KLB
Nov-20	1	non-KLB
Des-20	1	non-KLB

Halaman ini sengaja dikosongkan

LAMPIRAN C

Hasil Klasifikasi

1. Dataran Rendah

- Evaluasi Klasifikasi pada Uji Proporsi

Tabel C.1 Evaluasi Klasifikasi pada Uji Proporsi Data di Dataran Rendah

	Proporsi	Model 1	Model 2	Model 3
Akurasi	80-20	90,9%	85,0%	86,4%
	70-30	93,9%	83,3%	87,5%
	60-40	86,0%	84,6%	83,7%
	50-50	87,0%	83,7%	84,9%
Sensitivitas	80-20	50,0%	50,0%	25,0%
	70-30	50,0%	50,0%	25,0%
	60-40	66,7%	66,7%	66,7%
	50-50	76,9%	76,9%	69,2%
Spesifisitas	80-20	100,0%	93,8%	100,0%
	70-30	100,0%	88,5%	96,4%
	60-40	91,2%	87,9%	88,2%
	50-50	90,2%	86,1%	90,0%

- Evaluasi Klasifikasi pada Uji Kecamatan Lain

Tabel C.2 Evaluasi Klasifikasi pada Uji Kecamatan Lain di Dataran Rendah

	Kecamatan	Model 1	Model 2	Model 3
Akurasi	Turen	67,3%	74,2%	64,2%
	Donomulyo	76,6%	81,4%	81,1%
	Bululawang	81,3%	86,6%	84,9%
	Gondanglegi	69,2%	74,2%	67,9%
Sensitivitas	Turen	45,5%	62,5%	50,0%

	Kecamatan	Model 1	Model 2	Model 3
Sensitivitas	Donomulyo	10,5%	31,3%	33,3%
	Bululawang	0,0%	21,4%	7,1%
	Gondanglegi	27,6%	18,5%	32,1%
Spesifisitas	Turen	69,8%	75,3%	65,6%
	Donomulyo	90,9%	91,4%	90,9%
	Bululawang	94,6%	97,6%	96,7%
	Gondanglegi	84,6%	95,7%	80,8%

2. Dataran Sedang

- Evaluasi Klasifikasi pada Uji Proporsi

Tabel C.3 Evaluasi Klasifikasi pada Uji Proporsi Data di Dataran Sedang

	Proporsi	Model 1	Model 2	Model 3
Akurasi	80-20	95,2%	90,9%	95,2%
	70-30	96,9%	93,8%	93,8%
	60-40	92,9%	95,3%	95,2%
	50-50	94,3%	96,2%	96,2%
Sensitivitas	80-20	100,0%	90,9%	100,0%
	70-30	100,0%	90,9%	100,0%
	60-40	91,7%	91,7%	100,0%
	50-50	91,7%	91,7%	100,0%
Spesifisitas	80-20	90,0%	90,9%	90,0%
	70-30	95,2%	95,2%	90,5%
	60-40	93,3%	96,8%	93,3%
	50-50	95,1%	97,6%	95,1%

- Evaluasi Klasifikasi pada Uji Kecamatan Lain

Tabel C.4 Evaluasi Klasifikasi pada Uji Kecamatan Lain di Dataran Sedang

	Kecamatan	Model 1	Model 2	Model 3
Akurasi	Dampit	67,6%	72,6%	61,0%
	Singosari	68,6%	70,8%	65,7%
	Tumpang	85,7%	82,1%	86,7%
	Lawang	66,7%	75,5%	65,7%
	Sumbermanjing	55,2%	55,7%	53,3%
	Wajak	87,6%	84,0%	80,0%
Sensitivitas	Dampit	78,1%	71,7%	71,4%
	Singosari	76,2%	71,4%	76,2%
	Tumpang	79,3%	62,1%	82,8%
	Lawang	0,0%	7,1%	7,1%
	Sumbermanjing	60,0%	44,0%	64,0%
	Wajak	100,0%	91,2%	100,0%
Spesifisitas	Dampit	48,0%	64,0%	48,0%
	Singosari	50,0%	22,2%	44,4%
	Tumpang	65,5%	72,9%	57,1%
	Lawang	64,5%	74,0%	59,2%
	Sumbermanjing	98,9%	93,5%	98,9%
	Wajak	68,8%	85,2%	66,3%

4. Dataran Tinggi

- Evaluasi Klasifikasi pada Uji Proporsi

Tabel C.5 Evaluasi Klasifikasi pada Uji Proporsi Data di Dataran Tinggi

	Proporsi	Model 1	Model 2	Model 3
Akurasi	80-20	70,0%	80,0%	76,2%
	70-30	79,3%	86,2%	84,4%
	60-40	79,5%	89,7%	81,0%
	50-50	81,3%	91,7%	84,9%
Sensitivitas	80-20	14,3%	42,9%	28,6%
	70-30	50,0%	66,7%	66,7%
	60-40	68,4%	78,9%	75,0%
	50-50	70,0%	80,0%	75,0%
Spesifisitas	80-20	100,0%	100,0%	100,0%
	70-30	100,0%	100,0%	100,0%
	60-40	90,0%	100,0%	86,4%
	50-50	89,3%	100,0%	90,9%

- Evaluasi Klasifikasi pada Uji Kecamatan Lain

Tabel C.6 Evaluasi Klasifikasi pada Uji Kecamatan Lain di Dataran Tinggi

	Kecamatan	Model 1	Model 2	Model 3
Akurasi	Poncokusumo	84,4%	83,3%	72,4%
	Jabung	74,0%	77,1%	74,3%
Sensitivitas	Poncokusumo	60,6%	72,7%	37,1%
	Jabung	9,1%	9,1%	7,4%
Spesifisitas	Poncokusumo	96,8%	88,9%	90,0%
	Jabung	93,2%	97,3%	97,4%

LAMPIRAN D

Uji Korelasi Pemilihan Kecamatan

1. Dataran Rendah

Tabel D.1 Uji Korelasi Antarkecamatan pada Dataran Rendah

	<i>Turen</i>	<i>Kepanjen</i>	<i>Gondanglegi</i>	<i>Bululawang</i>	<i>Donomulyo</i>
<i>Turen</i>	1	0.813	0.559	0.341	0.538
<i>Kepanjen</i>	0.813	1	0.639	0.575	0.544
<i>Gondanglegi</i>	0.559	0.639	1	0.413	0.445
<i>Bululawang</i>	0.340	0.575	0.413	1	0.390
<i>Donomulyo</i>	0.538	0.544	0.445	0.390	1

2. Dataran Sedang

Tabel D.2 Uji Korelasi pada Dataran Sedang

	<i>Dampit</i>	<i>Singosari</i>	<i>Pakisaji</i>	<i>Tumpang</i>	<i>Wajak</i>	<i>Sumbermanjing</i>	<i>Lawang</i>	<i>Karangploso</i>
<i>Dampit</i>	1	0,472	0,517	0,351	0,452	0,661	0,199	0,080
<i>Singosari</i>	0,472	1	0,345	0,144	0,198	0,511	0,027	0,179
<i>Pakisaji</i>	0,517	0,345	1	0,305	0,492	0,430	0,393	0,397
<i>Tumpang</i>	0,351	0,144	0,305	1	0,310	0,302	0,448	0,284
<i>Wajak</i>	0,452	0,198	0,492	0,310	1	0,389	0,537	0,231
<i>Sumbermanjing</i>	0,661	0,511	0,430	0,302	0,389	1	0,242	0,091
<i>Lawang</i>	0,199	0,027	0,393	0,448	0,537	0,242	1	0,428
<i>Karangploso</i>	0,080	0,179	0,397	0,284	0,231	0,091	0,428	1

3. Dataran Tinggi

Tabel D.3 Uji Korelasi pada Dataran Tinggi

	<i>Ngajum</i>	<i>Poncokusumo</i>	<i>Jabung</i>
<i>Ngajum</i>	1	0,684	0,467
<i>Poncokusumo</i>	0,684	1	0,271
<i>Jabung</i>	0,467	0,271	1

Halaman ini sengaja dikosongkan

LAMPIRAN E

Hasil Peramalan dengan pengukuran RMSE dan SMAPE

1. Dataran Rendah
 - Pengujian Pada Kecamatan Pembangun Model (Test-Set)

Tabel E.1 Hasil Uji pada *Test-set* (Dataran Rendah)

Pembagian Data	Skenario		Parameter								RMSE	SMAPE
	Pemilihan Fitur	Windowing	col	lr	gm	md	mcw	n_est	alpha	sub		
80-20	ALL-VI	6	1	0,5	0	8	1	100	0	0,6	2,55	31,31
	TOP3-VI	3	0,5	0,5	0	4	1	50	0	1	2,77	26,56
	TOP2-VI	7	0,75	0,1	0	2	4	50	0	0,6	2,79	25,54
	TOP1-VI	1	0,5	0,5	0	8	1	100	0	0,8	2,72	28,36
	NO-VI	1	0,5	0,5	0	2	1	100	0,01	0,6	2,40	26,74
70-30	ALL-VI	6	0,5	0,1	0	8	1	100	0	0,6	2,77	26,34
	TOP3-VI	10	0,75	0,1	0,01	8	1	50	0,01	0,6	3,08	29,55
	TOP2-VI	11	0,5	0,5	0,01	8	1	50	0	1	2,53	27,57
	TOP1-VI	2	0,5	0,3	0,01	8	1	100	0	0,6	2,59	25,71
	NO-VI	2	0,5	0,1	0,01	8	1	50	0,01	0,6	2,53	25,60
60-40	ALL-VI	0	0,5	0,1	0,01	2	4	50	0	0,6	4,95	34,23
	TOP3-VI	12	0,5	0,1	0	4	1	100	0	1	4,13	30,84

Skenario			Parameter								RMSE	SMAPE
Pembagian Data	Pemilihan Fitur	Windowing	col	lr	gm	md	mcw	n_est	alpha	sub		
	TOP2-VI	2	0,5	0,1	0,01	2	4	50	0	0,6	4,75	68,21
	TOP1-VI	0	0,5	0,5	0	2	4	50	0	0,6	4,05	26,70
	NO-VI	1	0,5	0,3	0,01	4	4	50	0	0,8	4,38	32,90
50-50	ALL-VI	12	1	0,5	0,01	4	2	100	0	0,6	5,28	62,37
	TOP3-VI	0	0,5	0,1	0,01	2	4	50	0,01	0,6	4,47	60,81
	TOP2-VI	0	0,5	0,1	0,01	2	4	50	0	0,6	4,75	68,21
	TOP1-VI	0	0,5	0,5	0,01	4	1	100	0	0,6	4,23	38,17
	NO-VI	1	0,5	0,5	0,01	4	1	50	0,01	1	3,93	34,12

Keterangan :

col = colsample_bytree, lr = learning_rate, gm = gamma, md = max_depth, mcw = min_child_weight, n_est = n_estimator, alpha = reg_alpha, Sub = subsample

- Hasil Uji Proporsional (Dataran Rendah)

Tabel E.2 Hasil Uji Proporsional Data (Dataran Rendah)

Pembagian Data	Skenario		RMSE (Uji Proporsi)			SMAPE (Uji Proporsi)		
	Pemilihan Fitur	Windowing	Proporsi 1	Proporsi 2	Proporsi 3	Proporsi 1	Proporsi 2	Proporsi 3
80-20	ALL-VI	6	2,102	1,827	1,638	21,302	16,219	15,097
	TOP3-VI	3	2,242	1,957	1,743	17,687	13,696	14,780
	TOP2-VI	7	2,529	3,578	3,289	26,203	27,121	27,628
	TOP1-VI	1	2,223	1,947	1,738	19,240	14,964	17,608
	NO-VI	1	2,211	3,765	3,508	25,755	27,566	30,631
70-30	ALL-VI	6	2,920	2,418	2,168	23,370	20,686	18,716
	TOP3-VI	10	3,153	2,912	2,736	25,764	29,843	28,346
	TOP2-VI	11	2,621	2,394	2,221	26,144	25,176	23,932
	TOP1-VI	2	2,648	2,558	2,389	23,973	29,147	33,631
	NO-VI	2	2,655	3,858	3,779	25,285	28,442	30,147
60-40	ALL-VI	0	4,211	4,153	4,555	37,625	36,631	36,959
	TOP3-VI	12	3,181	3,017	3,724	28,623	29,501	25,410
	TOP2-VI	2	4,612	4,097	5,156	73,157	64,507	68,564
	TOP1-VI	0	2,951	2,506	3,782	27,693	25,897	30,036
	NO-VI	1	2,707	2,551	4,020	32,583	31,616	33,864
50-50	ALL-VI	12	4,582	4,340	5,695	80,404	71,459	65,649
	TOP3-VI	0	4,086	3,691	4,826	62,442	54,001	61,059
	TOP2-VI	0	4,612	4,097	5,156	73,157	64,507	68,564

Skenario		RMSE (Uji Proporsi)				SMAPE (Uji Proporsi)		
	TOP1-VI	0	3,754	3,166	4,518	43,411	37,135	35,111
	NO-VI	1	2,092	2,043	4,151	22,718	25,759	29,443

- Hasil Uji Kecamatan Lain (Dataran Rendah)

Tabel E.3 Hasil Uji Kecamatan Lain (Dataran Rendah)

Skenario			RMSE				SMAPE			
Pembagian Data	Pemilihan Fitur	Windowing	Tur	Don	Bul	Gon	Tur	Don	Bul	Gon
80-20	ALL-VI	6	11,10	2,26	2,96	1,88	49,08	48,84	54,21	50,46
	TOP3-VI	3	12,04	2,15	2,57	2,04	46,15	45,21	38,88	51,94
	TOP2-VI	7	14,20	2,34	3,05	2,37	45,29	45,08	40,63	53,49
	TOP1-VI	1	15,07	2,64	2,71	2,12	51,79	43,83	48,64	59,14
	NO-VI	1	16,86	2,64	3,39	2,40	42,84	57,37	37,42	47
70-30	ALL-VI	6	12,07	2,16	2,75	2,08	45,59	48,15	41,87	48,96
	TOP3-VI	10	11,85	2,35	2,73	2,25	42,83	50,73	37,81	51,14
	TOP2-VI	11	15,48	2,33	2,54	2,53	41,78	47,07	41,98	56,85
	TOP1-VI	2	15,70	2,54	2,58	2,06	53,08	44,22	40,71	51,63
	NO-VI	2	16,79	2,63	3,09	2,38	45,50	54,97	33,98	49,76
60-40	ALL-VI	0	18,12	2,45	2,80	2,59	43,16	48,32	36,33	59,38
	TOP3-VI	12	10,54	2,35	2,71	2,12	39,84	45,93	36,62	49,12
	TOP2-VI	2	17,48	2,51	2,99	2,77	58,28	42,65	51,06	66,07
	TOP1-VI	0	19,59	2,68	2,95	2,49	46,04	45,50	41,59	59,96
	NO-VI	1	17,14	2,70	3,39	2,35	43,18	56,63	40,83	50,50

Skenario			RMSE				SMAPE			
50-50	ALL-VI	12	10,34	2,79	3,01	2,31	48,21	53,15	43,77	61,48
	TOP3-VI	0	17,33	2,46	2,87	2,70	56,26	44,10	44,47	63,41
	TOP2-VI	0	17,48	2,51	2,99	2,77	58,28	42,65	51,06	66,07
	TOP1-VI	0	16,90	2,62	3,36	2,51	35,06	49,47	49,40	66,34
	NO-VI	1	16,70	2,54	3,17	2,38	47,07	49,91	32,69	47,29159265

Keterangan : Tur = Kecamatan Turen, Don = Kecamatan Donomulyo, Bul = Kecamatan Bululawang, Gon = Kecamatan Gondanglegi

2. Dataran Sedang

- Pengujian Pada Kecamatan Pembangun Model (Test-Set)

Tabel E.4 Hasil Uji pada *Test-set* (Dataran Sedang)

Pembagian Data	Skenario		Parameter								RMSE	SMAPE
	Pemilihan Fitur	Windowing	col	lr	gm	md	mcw	n_est	alpha	sub		
80-20	ALL-VI	3	0,5	0,1	0,01	8	4	50	0,01	1	3,656	39,907
	TOP3-VI	3	0,5	0,1	0,01	8	1	50	0,01	0,8	3,197	33,739
	TOP2-VI	3	0,75	0,5	0,01	2	1	100	0	0,8	3,184	41,974
	TOP1-VI	2	0,5	0,5	0,01	4	1	50	0	0,8	2,880	43,029
	NO-VI	3	0,5	0,1	0,01	2	4	50	0	0,6	2,699	35,620
70-30	ALL-VI	3	0,75	0,1	0,01	2	4	50	0,01	0,6	4,007	40,186
	TOP3-VI	2	0,5	0,1	0	4	1	50	0	0,6	3,212	38,409
	TOP2-VI	3	0,75	0,5	0	4	1	50	0,01	1	3,474	43,390
	TOP1-VI	2	1	0,5	0,01	4	1	50	0	0,6	3,152	44,941
	NO-VI	3	0,5	0,1	0	2	4	50	0	0,6	2,899	35,988
60-40	ALL-VI	1	0,75	0,1	0,01	4	4	50	0	0,6	5,367	40,903
	TOP3-VI	2	0,5	0,1	0,01	4	4	50	0	0,6	5,202	41,288
	TOP2-VI	3	0,5	0,1	0	4	1	50	0,01	0,6	4,428	38,173
	TOP1-VI	3	1	0,1	0	2	1	50	0,01	0,6	4,362	34,695
	NO-VI	1	0,5	0,1	0,01	2	4	100	0,01	1	4,979	35,159
50-50	ALL-VI	1	0,5	0,1	0	4	4	50	0,01	0,6	5,378	41,443
	TOP3-VI	11	1	0,1	0	4	1	50	0	0,8	5,184	39,333

Skenario			Parameter								RMSE	SMAPE
Pembagian Data	Pemilihan Fitur	Windowing	col	lr	gm	md	mcw	n_est	alpha	sub		
	TOP2-VI	3	0,5	0,1	0	2	4	50	0	0,6	4,532	37,793
	TOP1-VI	3	0,75	0,1	0	2	4	50	0	0,6	4,577	36,451
	NO-VI	4	1	0,3	0,01	8	2	100	0	1	5,165	36,342

Keterangan :

col = colsample_bytree, lr = learning_rate, gm = gamma, md = max_depth, mcw = min_child_weight, n_est = n_estimator, alpha = reg_alpha, Sub = subsample

- Hasil Uji Proporsional (Dataran Rendah)

Tabel E.5 Hasil Uji Proporsional Data (Dataran Rendah)

Pembagian Data	Skenario		RMSE (Uji Proporsi)			SMAPE (Uji Proporsi)		
	Pemilihan Fitur	Windowing	Proporsi 1	Proporsi 2	Proporsi 3	Proporsi 1	Proporsi 2	Proporsi 3
80-20	ALL-VI	3	3,043	3,066	2,782	32,299	27,782	24,709
	TOP3-VI	3	2,667	2,435	2,284	28,985	25,124	23,689
	TOP2-VI	3	2,630	2,360	2,191	35,136	30,261	27,183
	TOP1-VI	2	2,558	2,356	2,307	36,950	32,533	30,348
	NO-VI	3	2,812	4,654	4,419	36,349	36,293	35,704
70-30	ALL-VI	3	3,958	4,261	3,924	44,651	36,287	33,925
	TOP3-VI	2	3,100	2,955	2,743	38,801	32,101	29,419
	TOP2-VI	3	3,620	3,033	2,700	41,740	33,291	26,856
	TOP1-VI	2	3,417	2,788	2,567	48,063	38,393	33,789
	NO-VI	3	2,574	4,560	4,345	34,959	35,858	35,227
60-40	ALL-VI	1	3,866	3,857	4,849	43,567	42,887	36,288
	TOP3-VI	2	4,221	4,285	4,785	45,084	43,306	38,128
	TOP2-VI	3	3,137	3,083	4,002	39,811	36,629	33,909
	TOP1-VI	3	3,421	3,231	4,017	36,387	34,599	32,778
	NO-VI	1	2,901	2,771	4,759	37,569	36,009	35,749
50-50	ALL-VI	1	4,140	4,450	5,753	42,744	46,360	44,005
	TOP3-VI	11	4,744	4,208	5,260	44,459	43,948	41,551
	TOP2-VI	3	3,454	3,634	4,809	40,419	38,552	38,499

Skenario		RMSE (Uji Proporsi)				SMAPE (Uji Proporsi)		
	TOP1-VI	3	3,468	3,441	4,788	36,059	34,543	35,819
	NO-VI	4	3,556	3,237	5,412	36,738	35,341	35,611

- Hasil Uji Kecamatan Lain (Dataran Rendah)

Tabel E.6 Hasil Uji Kecamatan Lain (Dataran Rendah)

Skenario			RMSE							SMAPE (%)						
Pembagian Data	Pemilihan Fitur	Windowing	Dp	Sg	Tp	Lw	Sm	Wj	Kp	Dp	Sg	Tp	Lw	Sm	Wj	Kp
80-20	ALL-VI	3	4,6	3,3	5,0	7,8	10,6	3,3	3,6	44	54	36	40	62	46	43
	TOP3-VI	3	5,5	3,5	4,7	8,1	12,6	3,6	3,6	44	53	34	40	59	46	41
	TOP2-VI	3	6,3	3,7	5,1	8,0	13,5	3,9	3,9	51	54	38	46	60	49	42
	TOP1-VI	2	6,1	3,5	5,3	8,9	13,1	3,7	4,1	50	51	39	40	56	49	42
	NO-VI	3	4,4	2,9	4,6	7,2	8,3	3,4	3,2	46	50	35	44	60	47	42
70-30	ALL-VI	3	4,7	3,1	5,2	7,9	10,0	3,2	3,6	46	53	39	44	63	46	42
	TOP3-VI	2	4,7	3,1	4,9	7,3	9,3	3,3	3,4	46	53	37	43	62	48	43
	TOP2-VI	3	5,8	3,7	5,0	7,7	12,6	3,7	3,9	49	57	40	38	57	49	44
	TOP1-VI	2	7,1	3,9	5,0	9,1	15,2	3,9	4,0	56	53	39	58	70	50	41
	NO-VI	3	4,8	3,1	4,7	7,4	9,5	3,4	3,4	48	51	35	45	61	46	42
60-40	ALL-VI	1	4,9	3,1	5,2	8,0	10,3	3,3	3,4	46	51	38	46	64	46	41
	TOP3-VI	2	5,4	3,4	5,4	8,2	11,3	3,5	3,5	50	53	40	49	68	47	44
	TOP2-VI	3	5,6	3,5	5,0	7,4	10,6	3,4	3,6	51	53	37	44	61	46	44
	TOP1-VI	3	5,8	3,4	5,3	7,1	9,8	3,5	3,5	50	53	39	39	57	47	42
	NO-VI	1	4,9	2,9	4,8	6,9	9,4	3,4	3,0	48	50	34	43	61	45	38

Skenario		RMSE								SMAPE (%)							
50-50	ALL-VI	1	5,4	3,4	5,5	8,2	11,2	3,5	3,7	49	52	40	47	66	47	49	
	TOP3-VI	11	6,6	3,7	5,5	8,9	14,8	4,0	4,1	50	51	36	47	67	48	50	
	TOP2-VI	3	5,1	3,1	5,1	7,3	8,8	3,3	3,3	51	52	39	41	60	46	51	
	TOP1-VI	3	5,6	3,2	5,0	7,1	9,4	3,5	3,5	52	53	38	38	55	47	52	
	NO-VI	4	7,3	4,2	5,4	9,4	15,4	3,7	3,8	50	54	36	42	58	46	50	

Keterangan : Dp = Kecamatan Dampit, Sg = Kecamatan Singosari, Tp = Kecamatan Tumpang, Lw = Kecamatan Lawang, Sm = Kecamatan Sumbermanjing, Wj = Kecamatan Wajak, Kp = Kecamatan Karangploso

3. Dataran Tinggi

- Pengujian Pada Kecamatan Pembangun Model (Test-Set)

Tabel E.7 Hasil Uji pada *Test-set* (Dataran Tinggi)

Pembagian Data	Skenario		Parameter								RMSE	SMAPE
	Pemilihan Fitur	Windowing	col	lr	gm	md	mcw	n_est	alpha	sub		
80-20	ALL-VI	2	0,5	0,5	0,01	4	1	50	0	1	1,71	29,40
	TOP3-VI	3	0,75	0,5	0,01	4	1	100	0,01	0,8	1,46	25,02
	TOP2-VI	12	0,75	0,3	0,01	4	1	50	0	0,8	1,38	22,78
	TOP1-VI	12	1	0,1	0	8	4	100	0,01	0,8	1,79	27,47
	NO-VI	10	0,75	0,1	0	8	1	50	0,01	1	1,54	27,17
70-30	ALL-VI	12	0,75	0,5	0	8	1	50	0	1	1,45	24,13
	TOP3-VI	0	0,5	0,1	0,01	2	1	50	0	1	1,54	26,08
	TOP2-VI	12	1	0,1	0	4	1	100	0,01	0,8	1,51	24,59
	TOP1-VI	4	0,75	0,1	0	8	1	50	0	1	1,94	32,07
	NO-VI	10	1	0,1	0	4	4	50	0,01	0,6	1,94	29,58
60-40	ALL-VI	8	0,75	0,5	0	4	2	50	0	0,8	4,55	39,66
	TOP3-VI	2	0,5	0,5	0,01	2	2	50	0	0,8	4,59	42,21
	TOP2-VI	7	0,5	0,3	0	4	2	100	0,01	1	4,66	42,68
	TOP1-VI	2	1	0,3	0	8	4	50	0	0,6	4,65	45,97
	NO-VI	10	0,5	0,3	0	2	4	50	0	0,8	4,57	39,52
50-50	ALL-VI	6	0,75	0,5	0,01	2	4	50	0	1	4,15	49,75
	TOP3-VI	4	1	0,1	0	2	4	50	0,01	0,6	4,19	52,58

Skenario			Parameter								RMSE	SMAPE
Pembagian Data	Pemilihan Fitur	Windowing	col	lr	gm	md	mcw	n_est	alpha	sub		
	TOP2-VI	2	0,75	0,1	0	2	4	50	0	0,6	4,15	53,58
	TOP1-VI	1	1	0,1	0	4	4	50	0	0,8	4,17	58,08
	NO-VI	10	0,75	0,3	0	4	4	100	0,01	0,6	3,98	50,32

Keterangan :

col = colsample_bytree, lr = learning_rate, gm = gamma, md = max_depth, mcw = min_child_weight, n_est = n_estimator, alpha = reg_alpha, Sub = subsample

- Hasil Uji Proporsional (Dataran Tinggi)

Tabel E.8 Hasil Uji Proporsional Data (Dataran Tinggi)

Pembagian Data	Skenario		RMSE (Uji Proporsi)			SMAPE (Uji Proporsi)		
	Pemilihan Fitur	Windowing	Proporsi 1	Proporsi 2	Proporsi 3	Proporsi 1	Proporsi 2	Proporsi 3
80-20	ALL-VI	2	1,559	1,477	1,332	25,867	33,927	46,393
	TOP3-VI	3	1,332	1,320	1,184	23,475	31,366	45,611
	TOP2-VI	12	1,235	1,218	1,133	21,779	25,952	38,207
	TOP1-VI	12	1,617	2,279	2,085	22,630	27,075	38,762
	NO-VI	10	1,258	2,553	3,060	19,593	25,941	38,664
70-30	ALL-VI	12	1,501	1,247	1,124	25,584	25,661	37,517
	TOP3-VI	0	1,524	1,602	1,533	27,538	37,536	48,003
	TOP2-VI	12	1,555	1,306	1,178	23,659	26,333	38,083
	TOP1-VI	4	1,957	1,932	1,744	34,360	37,813	49,772
	NO-VI	10	1,760	3,608	3,518	27,110	35,336	46,401
60-40	ALL-VI	8	1,927	1,910	4,070	33,973	32,745	49,729
	TOP3-VI	2	1,973	1,914	4,136	35,152	31,952	53,117
	TOP2-VI	7	2,048	1,928	4,179	35,440	34,717	52,118
	TOP1-VI	2	2,126	1,972	4,191	38,960	35,347	56,163
	NO-VI	10	1,853	1,752	4,180	31,358	32,251	49,984
50-50	ALL-VI	6	1,912	1,795	4,554	34,199	31,120	37,945
	TOP3-VI	4	1,973	1,930	4,650	33,239	32,361	41,284
	TOP2-VI	2	1,969	1,944	4,593	33,459	32,377	42,781

Skenario		RMSE (Uji Proporsi)			SMAPE (Uji Proporsi)			
	TOP1-VI	1	2,286	2,059	4,631	44,769	38,165	47,998
	NO-VI	10	1,940	1,890	4,322	31,621	35,375	40,523

- Hasil Uji Kecamatan Lain (Dataran Tinggi)

Tabel E.9 Hasil Uji Kecamatan Lain (Dataran Tinggi)

Skenario		RMSE			SMAPE	
Pembagian Data	Pemilihan Fitur	Windowing	Poncokusumo	Jabung	Poncokusumo	Jabung
80-20	ALL-VI	2	2,26	2,31	56,88	53,03
	TOP3-VI	3	2,34	2,40	56,79	55,20
	TOP2-VI	12	2,33	2,19	47,68	42,76
	TOP1-VI	12	2,49	2,26	48,19	42,87
	NO-VI	10	2,72	2,23	47,92	38,27
70-30	ALL-VI	12	2,46	2,18	46,65	41,55
	TOP3-VI	0	2,42	2,48	60,62	62,15
	TOP2-VI	12	2,38	2,14	44,70	38,34
	TOP1-VI	4	2,49	2,42	59,50	55,56
	NO-VI	10	2,70	2,23	49,41	43,47
60-40	ALL-VI	8	2,71	2,45	56,12	55,77
	TOP3-VI	2	2,74	2,53	57,14	58,32
	TOP2-VI	7	2,71	2,36	53,73	48,91
	TOP1-VI	2	2,74	2,49	57,90	57,38
	NO-VI	10	2,72	2,48	53,70	53,47

Skenario			RMSE		SMAPE	
50-50	ALL-VI	6	2,82	2,57	58,69	61,50
	TOP3-VI	4	2,75	2,47	54,37	53,25
	TOP2-VI	2	2,78	2,56	57,18	58,78
	TOP1-VI	1	2,94	2,67	69,21	71,82
	NO-VI	10	2,66	2,37	51,36	50,09