



**ITS**  
Institut  
Teknologi  
Sepuluh Nopember

**TUGAS AKHIR - IS184853**

**PREDIKSI KEJADIAN LUAR BIASA PADA KASUS DEMAM BERDARAH DENGUE DI KABUPATEN MALANG MENGGUNAKAN SUPPORT VECTOR MACHINES - FLOWER POLLINATION ALGORITHM (SVM - FPA)**

*DENGUE FEVER OUTBREAK PREDICTION IN MALANG USING SUPPORT VECTOR MACHINES - FLOWER POLLINATION ALGORITHM (SVM - FPA)*

**YUSNARDO TENDIO**  
NRP. 05211640000149

Dosen Pembimbing  
Prof. Ir. Arif Djunaidy, M.Sc, Ph.D

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





**TUGAS AKHIR - IS184853**

**PREDIKSI KEJADIAN LUAR BIASA PADA KASUS DEMAM BERDARAH DENGUE DI KABUPATEN MALANG MENGGUNAKAN SUPPORT VECTOR MACHINES - FLOWER POLLINATION ALGORITHM (SVM - FPA)**

**YUSNARDO TENDIO**  
**NRP. 052116400000149**

**Dosen Pembimbing**  
**Prof. Ir. Arif Djunaedi, M.Sc, Ph.D**

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

*Halaman ini sengaja dikosongkan*



UNDERGRADUATE THESIS - IS184853

***DENGUE FEVER OUTBREAK PREDICTION IN  
MALANG USING SUPPORT VECTOR  
MACHINES - FLOWER POLLINATION  
ALGORITHM (SVM - FPA)***

YUSNARDO TENDIO  
NRP. 052116400000149

**Dosen Pembimbing**  
Prof. Ir. Arif Djunaidy, M.Sc, Ph.D

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

*Halaman ini sengaja dikosongkan*

**LEMBAR PENGESAHAN****Prediksi kejadian luar biasa pada kasus demam berdarah dengue di Kabupaten Malang menggunakan Support Vector Machines - Flower Pollination Algorithm****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

**Yusnardo Tendio**  
**05211640000149**

Surabaya, 14 Agustus 2020

**Kepala Departemen Sistem Informasi**

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



*Halaman ini sengaja dikosongkan*



# LEMBAR PERSETUJUAN

## **PREDIKSI KEJADIAN LUAR BIASA PADA KASUS DEMAM BERDARAH DENGUE DI KABUPATEN MALANG MENGGUNAKAN SUPPORT VECTOR MACHINES – FLOWER POLLINATION ALGORITHM (SVM – FPA)**

### **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 :

**YUSNARDO TENDIO**  
**NRP. 05211640000085**

Disetujui Tim Penguji :

Tanggal Ujian : Juli 2020

Periode Wisuda : September 2020

**Prof. Ir. Arif Djunaidy, M.Sc, Ph.D**

  
(Pembimbing I)

**Ahmad Muklason, S.Kom, M.Sc, Ph.D**

  
(Penguji I)

**Faisal Mahananto, S.Kom, M.Eng, Ph.D**

  
(Penguji II)

**PREDIKSI KEJADIAN LUAR BIASA PADA KASUS  
DEMAM BERDARAH DENGUE DI KABUPATEN  
MALANG MENGGUNAKAN SUPPORT VECTOR  
MACHINES – FLOWER POLLINATION ALGORITHM  
(SVM – FPA)**

Nama : Yusnardo Tendio  
NRP : 05211640000149  
Departemen : Sistem Informasi ITS  
Pembimbing I : Prof. Ir. Arif Djunaidy, M.Sc, Ph.D

**ABSTRAK**

*Penyakit demam berdarah dengue (DBD) merupakan salah satu penyakit berbahaya yang menjadi sumber masalah kesehatan bagi masyarakat Indonesia. Jumlah penderita DBD mengalami peningkatan seiring berjalannya waktu. Kejadian luar biasa (KLB) pada kasus DBD dapat diprediksi dengan membuat model prediksi menggunakan keterkaitan antar variabel. Beberapa variabel yang mempengaruhi hasil prediksi antara lain adalah curah hujan, suhu, kecepatan angin, dan kelembaban. Dalam tugas akhir ini, data jumlah kasus DBD per kecamatan diperoleh dari Dinas Kesehatan Kabupaten Malang, sedangkan data curah hujan, suhu, kecepatan angin, dan kelembaban diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG).*

*Dalam Tugas Akhir ini, sebuah model prediksi KLB untuk kasus DBD dibangun menggunakan gabungan metode Support Vector Machine dan Flower Pollination Algorithm (SVM-FPA). Metode SVM digunakan untuk memisahkan kasus yang termasuk dalam kelas KLB dan kelas non-KLB, sedang FPA digunakan untuk memperoleh nilai parameter yang optimal dalam SVM.*

*Hasil uji coba menunjukkan bahwa nilai parameter optimal yang dihasilkan oleh FPA berupa kombinasi nilai cost dan gamma berturut-turut sebesar 2.829,0587 dan 0,002801. Kombinasi kedua nilai parameter ini memberikan hasil prediksi*

*SVM terbaik menggunakan 5-fold cross validation dengan nilai akurasi, recall, dan presisi berturut-turut sebesar 64,3%, 94,5%, dan 59,1%. Analisis menggunakan kurva precision-recall terhadap hasil prediksi KLB untuk kasus DBD yang tergolong tidak berimbang menunjukkan bahwa model SVM-FPA mempunyai kinerja yang cukup baik dan berada jauh di atas garis basis dengan nilai Average Precision (AP) sebesar 0,65.*

***Kata Kunci: demam berdarah, prediksi kejadian luar biasa, support vector machines, algoritma penyerbukan bunga***

# **DENGUE FEVER OUTBREAK PREDICTION IN MALANG USING SUPPORT VECTOR MACHINES – FLOWER POLLINATION ALGORITHM (SVM – FPA)**

Nama : Yusnardo Tendio  
NRP : 05211640000149  
Departemen : Sistem Informasi ITS  
Pembimbing I : Prof. Ir. Arif Djunaidy, M.Sc, Ph.D

## **ABSTRACT**

*Dengue fever (DF) is a dangerous disease that causes health problems for the people. The number of DF cases increases as time goes by. DF Outbreak can be predicted by making prediction models using interrelationships among variables. Some relevant variables that influence the results of predictions are rainfall, temperature, wind speed, and humidity. In this final project, the number of dengue cases per district data are obtained from the Public Health Office of Malang Region, while rainfall, temperature, wind speed, and humidity data are obtained from the Meteorology, Climatology and Geophysics Agency.*

*In this Final Project, an outbreak prediction model for DF cases is built using a combination of Support Vector Machine and Flower Pollination Algorithm (SVM-FPA). The SVM method is used to classify cases that are included in the outbreak class or not outbreak class, while the FPA is used to obtain optimal parameter values in the SVM..*

*The experimental results show that the optimal parameter values produced by the FPA consists of a combination of cost and gamma values respectively of 2829.0587 and 0.002801. The combination of these two parameter values gives the best SVM prediction results using 5-fold cross validation resulting in accuracy, recall, and precision values of 64.3%, 94.5% and 59.1%, respectively. The analysis using the precision-recall curve on the prediction results for the unbalanced DF cases shows that the SVM-FPA model has a fairly good performance*

*and is far above the baseline with an Average Precision (AP) of 0.65.*

***Keywords: dengue fever, outbreak prediction, support vector machines, flower pollination algorithm***

## SURAT PERNYATAAN BEBAS PLAGIARISME

Saya yang bertandatangan di bawah ini:

Nama : Yusnardo Tendio  
NRP : 05211640000149  
Tempat/Tanggal lahir : Palu / 07 Februari 1999  
Fakultas/Departemen : FTEIC / Sistem Informasi  
Nomor Telp/Hp/email : 082393073148/yusnardotendio@yahoo.com

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

**PREDIKSI KEJADIAN LUAR BIASA PADA KASUS DEMAM BERDARAH DENGUE DI KABUPATEN MALANG MENGGUNAKAN SUPPORT VECTOR MACHINES – FLOWER POLLINATION ALGORITHM (SVM-FPA)**

### **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, 26 Mei 2020



Yusnardo Tendio  
NRP.05211640000149

## KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Tuhan Yang Maha Esa atas izin-Nya penulis dapat menyelesaikan Tugas Akhir dengan judul **“Prediksi Kejadian Luar Biasa Pada Kasus Demam Berdarah Dengue Di Kabupaten Malang Menggunakan Support Vector Machines – Flower Pollination Algorithm (SVM – FPA)”** yang menjadi syarat kelulusan di Departemen Sistem Informasi Fakultas Teknologi Elektro dan Infomatika Cerdas Institut Teknologi Sepuluh Nopember Surabaya.

Pengerjaan Tugas Akhir ini dapat dikerjakan oleh karena dukungan, saran, motivasi, dan semangat yang diberikan oleh berbagai pihak. Secara khusus penulis akan menyampaikan ucapan terima kasih banyak kepada :

1. Seluruh keluarga besar penulis terutama orang tua dan kedua saudara penulis yang senantiasa mendoakan, memberikan motivasi dan semangat, sehingga penulis mampu menyelesaikan pendidikan sarjana ini dengan baik.
2. Bapak Arif Djunaidy, M.Sc., Ph.D, selaku dosen pembimbing yang telah memberikan arahan dan bimbingan, serta memberikan motivasi dalam pengerjaan Tugas Akhir ini.
3. Bapak Ahmad Mukhlason, S.Kom., M.Sc., Ph.D dan Faizal Mahananto, S.Kom, M.Eng., Ph.D, selaku dosen penguji yang telah memberikan kritik dan saran dalam penyempurnaan Tugas Akhir ini.
4. Ibu Wiwik Anggraeni, S.Si., M.Kom yang telah banyak meluangkan waktu untuk membimbing, mengarahkan, dan mendukung dengan memberikan ilmu, petunjuk, dan motivasi dalam penyelesaian Tugas Akhir ini.
5. Ibu Irmasari Hafid, S.Kom., M.Sc dan Bapak Achmad Holil Noor Ali, M.Kom, selaku dosen wali penulis yang telah memberikan arahan dan dukungan kepada penulis selama masa perkuliahan.

6. Teman-teman Sistem Informasi angkatan 2016 (Artemis) yang senantiasa menemani dan memberikan motivasi bagi penulis selama perkuliahan hingga dapat menyelesaikan Tugas Akhir ini.
7. Teman-teman laboratorium RDIB yang telah memberikan dukungan dan menjadi teman diskusi selama pengerjaan Tugas Akhir ini.
8. Seluruh pihak-pihak lainnya yang tidak dapat disebutkan satu per satu, yang telah membantu penulis selama perkuliahan hingga dapat menyelesaikan tugas akhir ini.

Penyusunan Tugas Akhir ini masih memiliki banyak kekurangan dan belum 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, 7 Januari 2020  
Yusnardo Tendio



## DAFTAR ISI

LEMBAR PENGESAHAN.....	V
LEMBAR PERSETUJUAN.....	VII
ABSTRAK.....	VIII
ABSTRACT.....	X
SURAT PERNYATAAN BEBAS PLAGIARISME.....	XII
KATA PENGANTAR.....	XIII
DAFTAR ISI.....	XV
DAFTAR GAMBAR.....	XVIII
DAFTAR TABEL.....	XXI
DAFTAR KODE PROGRAM.....	XXIII
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah.....	5
1.4 Tujuan.....	5
1.5 Manfaat.....	5
1.6 Relevansi.....	6
BAB II TINJAUAN PUSTAKA.....	7
2.1 Studi Literatur.....	7
2.2 Dasar Teori.....	8
2.2.1 Demam Berdarah Dengue.....	8
2.2.2 Kejadian Luar Biasa (Outbreak).....	9
2.2.3 Letak Geografis.....	10
2.2.4 Praproses Data.....	11
2.2.5 Penggalan Data.....	12
2.2.6 Flower Pollination Algorithm.....	13
2.2.7 Support Vector Machine.....	15
2.2.8 Evaluasi Kinerja.....	17
BAB III METODOLOGI.....	21
3.1 Studi Literatur.....	21
3.2 Pengumpulan Data.....	22
3.3 Praproses Data.....	22
3.4 Penyetelan Parameter.....	23
3.5 Prediksi Menggunakan Support Vector Machine.....	24
3.6 Analisa Kinerja.....	24

3.7 Penyusunan Tugas Akhir.....	24
<b>BAB IV PERANCANGAN .....</b>	<b>27</b>
4.1 Pengumpulan Data dan Eksplorasi.....	27
4.2 Praproses Data.....	27
4.2.1 Mengubah Data Harian Variabel Iklim .....	28
4.2.2 Menangani Data Yang Hilang .....	28
4.2.3 Normalisasi Data .....	28
4.2.4 Oversampling Data.....	29
4.2.5 Membagi Data Latih dan Uji.....	29
4.3 Penyetelan Parameter .....	30
4.4 Prediksi Menggunakan Support Vector Machine.....	31
4.5 Analisa Kinerja.....	31
<b>BAB V IMPLEMENTASI .....</b>	<b>33</b>
5.1 Persiapan Implementasi.....	33
5.2 Pengumpulan dan Eksplorasi Data.....	33
5.2.1 Ringkasan Statistik .....	34
5.2.2 Distribusi dan Hubungan Antar Variabel .....	34
5.3 Praproses Data .....	34
5.3.1 Mengubah Data Harian Variabel Iklim .....	35
5.3.2 Menangani Data Yang Hilang .....	36
5.3.3 Normalisasi Data .....	37
5.3.4 Oversampling Data.....	37
5.3.5 Pembagian Data Latih dan Uji.....	38
5.4 Penyetelan Parameter .....	38
5.4.1 Konfigurasi Parameter FPA .....	38
5.4.2 Pencarian Parameter dengan FPA .....	39
5.5 Prediksi Support Vector Machine .....	43
5.5.1 Load Data .....	44
5.5.2 Prediksi SVM .....	44
5.6 Analisa Kinerja.....	45
<b>BAB VI HASIL DAN PEMBAHASAN.....</b>	<b>47</b>
6.1 Hasil Eksplorasi Data .....	47
6.1.1 Visualisasi Jumlah Kasus Demam Berdarah .....	47
6.1.2 Ringkasan Statistik .....	55
6.1.3 Distribusi Kelas Target.....	56
6.1.4 Distribusi Tiap Variabel .....	56

6.2 Hasil Mengubah Data Harian Variabel Iklim .....	61
6.3 Hasil Pengisian Data Yang Hilang.....	62
6.4 Hasil Normalisasi Data .....	63
6.5 Hasil Implementasi Oversampling .....	65
6.6 Hasil Pembagian Data Latih dan Uji.....	66
6.7 Hasil Penyetelan Parameter.....	66
6.7.1 Percobaan dengan 10 iterasi.....	67
6.7.2 Percobaan dengan 20 iterasi.....	69
6.7.3 Percobaan dengan 50 iterasi.....	71
6.7.4 Percobaan dengan 100 iterasi.....	73
6.8 Prediksi SVM dan Evaluasi .....	75
6.8.1 Prediksi SVM dengan Parameter Default .....	75
6.8.2 Prediksi SVM dengan Parameter Hasil Optimasi	77
<b>BAB VII KESIMPULAN DAN SARAN .....</b>	<b>81</b>
7.1 Kesimpulan .....	81
7.2 Saran.....	82
<b>DAFTAR PUSTAKA .....</b>	<b>84</b>
<b>BIODATA PENULIS .....</b>	<b>89</b>
<b>LAMPIRAN A HASIL NORMALISASI .....</b>	<b>91</b>

## DAFTAR GAMBAR

Gambar 2.1 Pseudo-code Flower Pollination Algorithm .....	15
Gambar 2.2 Contoh Plot PRC.....	20
Gambar 3.1 Tahapan Pengerjaan.....	21
Gambar 4.1 Persebaran Data Variabel Target (KLB) .....	29
Gambar 5.1 Data Iklim .....	35
Gambar 6.1 Plot Kasus Demam Berdarah di Kecamatan Turen .....	47
Gambar 6.2 Plot Kasus Demam Berdarah di Kecamatan Ngajum .....	48
Gambar 6.3 Plot Kasus Demam Berdarah di Kecamatan Singosari .....	48
Gambar 6.4 Plot Kasus Demam Berdarah di Kecamatan Gondanglegi .....	49
Gambar 6.5 Plot Kasus Demam Berdarah di Kecamatan Bululawang.....	49
Gambar 6.6 Plot Kasus Demam Berdarah di Kecamatan Dampit .....	50
Gambar 6.7 Plot Kasus Demam Berdarah di Kecamatan Kepanjen.....	50
Gambar 6.8 Plot Kasus Demam Berdarah di Kecamatan Pakisaji .....	51
Gambar 6.9 Plot Kasus Demam Berdarah di Kecamatan Donomulyo.....	51
Gambar 6.10 Plot Kasus Demam Berdarah di Kecamatan Sumbermanjing Wetan .....	52
Gambar 6.11 Plot Kasus Demam Berdarah di Kecamatan Wajak.....	52
Gambar 6.12 Plot Kasus Demam Berdarah di Kecamatan Jabung.....	53
Gambar 6.13 Plot Kasus Demam Berdarah di Kecamatan Lawang .....	53
Gambar 6.14 Plot Kasus Demam Berdarah di Kecamatan Poncokusumo .....	54

Gambar 6.15 Plot Kasus Demam Berdarah di Kecamatan Karangploso .....	54
Gambar 6.16 Plot Kasus Demam Berdarah di Kecamatan Tumpang .....	55
Gambar 6.17 Distribusi Kelas Target.....	56
Gambar 6.18 Distribusi Variabel Suhu .....	56
Gambar 6.19 Rata-rata Suhu di Malang pada Tahun 2010-2018.....	57
Gambar 6.20 Distribusi Kelembaban Udara .....	58
Gambar 6.21 Rata-Rata Kelembaban Udara di Kabupaten Malang pada Tahun 2010-2018.....	58
Gambar 6.22 Distribusi Variabel Curah Hujan .....	59
Gambar 6.23 Rata-rata Curah Hujan pada Tahun 2010-2018	59
Gambar 6.24 Distribusi Variabel Kecepatan Angin.....	60
Gambar 6.25 Rata-rata Kecepatan Angin pada Tahun 2010-2018.....	60
Gambar 6.26 Tren Skor Berdasarkan Populasi Pada Percobaan dengan 10 Iterasi .....	68
Gambar 6.27 Tren Skor Berdasarkan Switch Probability Pada Percobaan dengan 10 Iterasi.....	68
Gambar 6.28 Tren Skor Berdasarkan Populasi Pada Percobaan dengan 20 Iterasi .....	70
Gambar 6.29 Tren Skor Berdasarkan Switch Probability Pada Percobaan dengan 10 Iterasi.....	70
Gambar 6.30 Tren Skor Berdasarkan Populasi Pada Percobaan dengan 50 Iterasi .....	72
Gambar 6.31 Tren Skor Berdasarkan Switch Probability Pada Percobaan dengan 50 Iterasi.....	72
Gambar 6.32 Tren Skor Berdasarkan Populasi Pada Percobaan dengan 100 Iterasi .....	74
Gambar 6.33 Tren Skor Berdasarkan Switch Probability Pada Percobaan dengan 100 Iterasi.....	74
Gambar 6.34 Tren Skor Berdasarkan Jumlah Iterasi .....	75
Gambar 6.35 Confusion Matrix Model SVM dengan Default Parameter.....	76

Gambar 6.36 Confusion Matrix Model SVM dengan Parameter Optimal Pada Data Validasi .....	77
Gambar 6.37 Dampak Optimisasi Parameter pada Model SVM .....	78
Gambar 6.38 Precision-recall Curve dengan 5-fold Cross Validation .....	80

## DAFTAR TABEL

Tabel 2.1 Confusion Matrix .....	17
Tabel 2.2 Data Berat Tikus .....	19
Tabel 3.1 Deskripsi Umum Atribut Yang Digunakan.....	22
Tabel 4.1 Sumber Data yang Digunakan .....	27
Tabel 4.2 Parameter FPA .....	30
Tabel 4.3 Daftar Percobaan Pada FPA.....	31
Tabel 5.1 Spesifikasi Perangkat Keras Yang Digunakan.....	33
Tabel 5.2 Perangkat Lunak Yang Digunakan .....	33
Tabel 5.3 Formula Mengubah Data Harian Variabel Iklim ...	36
Tabel 5.4 Contoh Penangan Data Yang Hilang Pada Data Iklim .....	36
Tabel 5.5 Default Parameter SVM.....	44
Tabel 6.1 Ringkasan Statistik Variabel Numerik.....	55
Tabel 6.2 Hasil Mengubah Data Harian Variabel Iklim .....	61
Tabel 6.3 Hasil Pengisian Data Yang Hilang Pada Data Iklim .....	62
Tabel 6.4 Hasil Pengisian Data yang Hilang pada Variabel ABJ.....	63
Tabel 6.5 Hasil Normalisasi Data .....	63
Tabel 6.6 Validasi Hasil Normalisasi.....	65
Tabel 6.7 Hasil Implementasi Oversampling .....	65
Tabel 6.8 Ukuran Data Latih, Data Validasi dan Data Uji.....	66
Tabel 6.9 5-fold cross validation.....	66
Tabel 6.10 Skenario Inisiasi Parameter FPA.....	67
Tabel 6.11 Pencarian Parameter dengan 10 Iterasi .....	67
Tabel 6.12 Pencarian Parameter dengan 20 Iterasi .....	69
Tabel 6.13 Pencarian Parameter dengan 50 Iterasi .....	71
Tabel 6.14 Pencarian Parameter dengan 100 Iterasi .....	73
Tabel 6.15 Hasil Evaluasi Prediksi SVM dengan Parameter Default.....	76
Tabel 6.16 Hasil Evaluasi Prediksi SVM dengan Parameter Hasil Optimasi.....	77

Tabel 6.17 Hasil Evaluasi Prediksi SVM dengan Parameter Hasil Optimasi pada 5-fold cross validation.....	78
Tabel 6.18 Hasil Pencarian Feature Importance.....	79



## DAFTAR KODE PROGRAM

Kode Program 5.1 Potongan Kode Program Ringkasan Statistik.....	34
Kode Program 5.2 Potongan Kode Visualisasi dan Eksplorasi .....	34
Kode Program 5.3 Penanganan Missing Value Pada Data ABJ .....	37
Kode Program 5.4 Potongan Kode Normalisasi Data .....	37
Kode Program 5.5 Potongan Kode Oversampling Data.....	38
Kode Program 5.6 Potongan Kode Pembagian Data Latih dan Uji.....	38
Kode Program 5.7 Potongan Kode Random Levy 2 Dimension .....	40
Kode Program 5.8 Potongan Kode Flower Pollination Algorithm .....	41
Kode Program 5.9 Potongan Kode Fungsi Objektif.....	42
Kode Program 5.10 Inisiasi Variabel Skenario Pencarian Parameter.....	42
Kode Program 5.11 Kode Program Skenario Pencarian Parameter.....	43
Kode Program 5.12 Potongan Kode Import Library Yang Dibutuhkan .....	44
Kode Program 5.13 Load Data.....	44
Kode Program 5.14 Prediksi SVM dengan Default Parameter .....	45
Kode Program 5.15 Prediksi SVM dengan Parameter Hasil FPA .....	45
Kode Program 5.16 Prediksi SVM pada 5-fold cross validation .....	45
Kode Program 5.17 Evaluasi Kinerja Model Prediksi .....	45

*Halaman ini sengaja dikosongkan*

# **BAB I**

## **PENDAHULUAN**

1.1.1 Pada bab pendahuluan ini akan menjelaskan tentang proses identifikasi masalah penelitian yang meliputi latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan relevansi terhadap pengerjaan tugas akhir. Harapannya dengan berdasarkan uraian pada bab ini gambaran umum permasalahan dan pemecahan masalah tugas akhir dapat dipahami.

### **1.1 Latar Belakang**

Demam Berdarah Dengue atau DBD adalah penyakit arboviral yang dapat dibawa oleh nyamuk *Aedes Aegypti* dan dapat ditularkan ke alam maupun manusia melalui nyamuk betina. Gejala dari penyakit ini biasanya dapat berupa sakit kepala, nyeri pada sendi dan otot, serta munculnya ruam pada kulit. *World Health Organization* (WHO) melaporkan bahwa setelah Kejadian Luar Biasa (KLB) Dengue yang terjadi di Filipina pada tahun 1955 mengakibatkan peningkatan kejadian demam berdarah. Berdasarkan hasil estimasi kasus demam berdarah yang dilakukan oleh WHO, terdapat sekitar 390 juta kasus demam berdarah setiap tahunnya [1].

Indonesia merupakan salah satu negara dengan banyak laporan kasus demam berdarah. Awal mula munculnya kasus demam di Indonesia terjadi di Kota Surabaya dan Jakarta pada tahun 1968. Dan sejak saat itu, penduduk yang terinfeksi penyakit demam berdarah terus meningkat [2]. Berdasarkan Pusat Data dan Informasi Kementerian Kesehatan RI, pada tahun 2016 terdapat 204.171 kasus demam berdarah serta terdapat 1.500 kasus kematian yang dilaporkan [3]. Dengan jumlah kasus demam berdarah yang banyak, tidak heran apabila WHO mencatat negara Indonesia sebagai negara yang memiliki kasus demam berdarah terbanyak di Asia Tenggara [1].

Jawa Timur merupakan salah satu provinsi dengan jumlah kasus demam berdarah terbanyak di Indonesia. Pada tahun 2018, tercatat ada 1.081 kasus demam berdarah yang tersebar di Jawa Timur. Hingga akhir Januari 2019, tercatat sebanyak 72 orang terserang penyakit DBD di Kabupaten Malang [4].

DBD menimbulkan beban ekonomi terhadap individu, rumah tangga, maupun terhadap negara. Bahkan sampai sekarang belum terdapat obat maupun vaksin yang dapat mencegah maupun mengobati. Pengobatan saat ini dilakukan hanya untuk mencegah infeksi tersebut semakin parah. Di Indonesia terdapat program untuk mengurangi pertambahan populasi nyamuk demam berdarah yang bernama program Pemberantasan Sarang Nyamuk. Program ini diusung pada tahun 1992 dengan melibatkan masyarakat untuk melakukan gerakan 3M (Menguras-Menutup-Mengubur). Gerakan tersebut mengalami perkembangan dengan penggunaan larvasida untuk membunuh larva nyamuk yang diusung pada tahun 2000. Namun dengan peningkatan yang terjadi pada kasus demam berdarah maka dapat dikatakan program tersebut masih belum berhasil untuk mencapai target yang diinginkan [5].

Menanggapi ancaman serius dari demam berdarah, banyak orang telah menerapkan metode-metode penggalian data untuk memprediksi *outbreak* atau kejadian luar biasa pada demam berdarah. Prediksi pada kasus demam berdarah telah banyak dilakukan dengan berbagai metode yang berkembang, hal ini dilakukan untuk mendapatkan model dengan akurasi yang optimal. Pada tugas akhir ini akan dilakukan prediksi kejadian luar biasa demam berdarah yang mana hasilnya akan berguna untuk dijadikan sebagai dasar pengambilan keputusan pencegahan demam berdarah [6].

Perkembangan metode dari penggalian data sangat pesat, terbukti dari banyaknya ragam metode yang telah digunakan dan bermanfaat, seperti *naïve bayes*, *decision tree*,

*backpropagation*, *neural network*, *Support Vector Machines* (SVM), serta metode-metode lainnya. SVM merupakan sebuah metode yang dibuat oleh Boster, Guyon, dan Vapnik. Metode ini awalnya dibuat untuk memecahkan masalah *overfitting*, masalah *local minima*, dan masalah dimensi pada ANN. SVM mampu menghasilkan prediksi paling akurat setelah dibandingkan dengan 5 algoritma berbeda pada penelitian yang dilakukan oleh Zhang dan Kusiakm. SVM memiliki beberapa parameter yang harus ditentukan nilainya, namun sangat sulit untuk menentukan nilai parameter yang akan menghasilkan akurasi terbaik [6].

Penggunaan metode SVM telah dilakukan untuk meramalkan wabah demam berdarah, seperti penelitian yang dilakukan pada data 5 distrik di Selangor menggunakan *Least Squares – Support Vector Machines* (LSSVM) menghasilkan akurasi sebesar 86,84% [7], Penelitian menggunakan data kasus demam berdarah di Thailand dengan metode *Radial Basis Function - Support Vector Machines* (RBF-SVM) menghasilkan akurasi sebesar 88,37% [8], dan juga penelitian yang membandingkan beberapa algoritma untuk mendapatkan model terbaik menggunakan data dari WHO. SVM memiliki beberapa kelebihan apabila dibandingkan dengan metode lainnya, seperti hasil peramalan yang lebih akurat, biaya komputasi yang rendah, dan landasan teori yang jelas. Namun SVM juga memiliki beberapa kelemahan, seperti waktu pelatihan yang panjang, tidak mudah untuk menggabungkan domain, dan harus menentukan nilai parameter untuk menghasilkan model terbaik. Penentuan nilai parameter untuk menghasilkan model terbaik memang sulit tetapi terdapat metode optimasi sebagai solusinya [8].

Optimasi merupakan usaha untuk mendapatkan hasil terbaik dengan batasan-batasan yang ada. Perkembangan metode optimasi sangat cepat, terbukti dengan terdapat banyak sekali metode dari optimasi, seperti algoritma genetika, algoritma

koloni semut, dan *Flower Pollination Algorithm*. *Flower Pollination Algorithm* merupakan algoritma optimasi dikembangkan pada tahun 2012 yang terinspirasi dari alam. Parameter yang terdapat pada SVM dengan kernel *Radial Basis Function* adalah parameter *cost* (C) dan  $\gamma$ . Rentang nilai dari parameter C yang baik adalah 0.03125 – 32768 sedangkan rentang nilai dari parameter  $\gamma$  yang baik adalah 0.00003 – 8. Dalam pengerjaan tugas akhir ini, FPA digunakan sebagai metode untuk mencari kombinasi parameter C dan gamma pada SVM. FPA menginisialisasi populasi acak seragam (n) menggunakan *Levy random walk*. SVM dilatih dan diuji n kali dan akurasi terbaik dipilih, dimana parameter C dan  $\gamma$  disimpan sebagai parameter optimal pada populasi, sehingga algoritma ini dapat menentukan kombinasi parameter dengan bilangan riil terbaik untuk digunakan oleh SVM. Model yang berkinerja buruk diganti secara iteratif dengan model yang mengadopsi parameter dengan fungsi objektif terbaik [9]. Oleh sebab itu, tugas akhir ini mengusulkan untuk memprediksi kejadian luar biasa pada demam berdarah di Kabupaten Malang menggunakan metode SVM yang di optimasi dengan *Flower Pollination Algorithm* [10].

Hasil tugas akhir ini akan menghasilkan model prediksi kejadian luar biasa dengan metode SVM dan FPA yang mampu memprediksi KLB dengan akurat agar dapat digunakan sebagai acuan dalam penentuan prioritas dalam pencegahan sebelum melonjaknya penderita demam berdarah pada suatu daerah di Kabupaten Malang.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah yang akan diangkat pada tugas akhir ini adalah bagaimana hasil prediksi kejadian luar biasa demam berdarah dengan menggunakan metode *Support Vector Machine* – *Flower Pollination Algorithm*.

### **1.3 Batasan Masalah**

Berdasarkan rumusan masalah yang telah disebutkan, adapun batasan masalah terkait pengerjaan tugas akhir ini adalah sebagai berikut.

1. Data yang digunakan adalah jumlah penderita demam berdarah, suhu udara, kelembaban, curah hujan, kecepatan angin, jumlah penduduk, angka bebas jentik dan jumlah kejadian demam berdarah di beberapa daerah Kabupaten Malang dari tahun 2010 hingga 2018.
2. Prediksi kejadian luar biasa pada kasus demam berdarah di Kabupaten Malang menggunakan data dengan periode bulanan.
3. Data jumlah penderita demam berdarah didapatkan dari Dinas Kesehatan Kabupaten Malang.
4. Penentuan awal kejadian luar biasa didasarkan pada Peraturan Menteri Kesehatan Republik Indonesia dimana terdapat kemungkinan pada suatu periode dengan jumlah kasus DBD sebanyak 2 diklasifikasikan sebagai kejadian luar biasa apabila periode sebelumnya tidak terdapat kasus demam berdarah atau hanya terdapat 1 kasus.
5. Data suhu udara, kelembaban, indeks curah hujan, kecepatan angin didapatkan dari Badan Meteorologi, Klimatologi, dan Geofisika Indonesia.
6. Perangkat lunak yang digunakan adalah Jupyter Notebook dengan bahasa Python.

### **1.4 Tujuan**

Tujuan dari tugas akhir ini adalah membangun model prediksi kejadian luar biasa untuk kasus demam berdarah dengue di Kabupaten Malang menggunakan metode SVM-FPA.

### **1.5 Manfaat**

Manfaat yang diperoleh apabila tujuan tugas akhir dapat tercapai adalah Dinas Kesehatan Kabupaten Malang dapat memprediksi kejadian luar biasa demam berdarah dengue sehingga dapat fokus untuk melakukan pencegahan pada daerah tersebut.

## **1.6 Relevansi**

Penyakit demam berdarah telah banyak menyebabkan kematian oleh karena itu dibutuhkan informasi prediksi kejadian luar biasa penyakit demam berdarah sehingga dapat dilakukan pencegahan. Tugas akhir ini termasuk dalam topik penelitian yang terdapat pada laboratorium Rekayasa Data dan Intelegensi Bisnis (RDIB) pada bidang penggalian data. Beberapa mata kuliah yang berkaitan dengan tugas akhir ini adalah Analitika Bisnis, Penggalian Data, dan Teknik Peramalan.



## **BAB II**

### **TINJAUAN PUSTAKA**

Bab ini menjelaskan tentang studi sebelumnya, yang mana penelitian sebelumnya yang terkait dengan tugas akhir ini. Tidak hanya itu, penjelasan dasar teori yang berisi tentang gambaran umum dari studi kasus dan teori mengenai metode yang akan digunakan dalam pengerjaan studi kasus tugas akhir ini.

#### **2.1 Studi Literatur**

Penelitian terdahulu ini menjadi salah satu acuan penulis sehingga penulis dapat memperkaya teori yang digunakan dalam melakukan pengkajian penelitian.

Pada penelitian yang berjudul *Dengue Outbreak Prediction: Hybrid Meta-heuristic Model* oleh Zuriani Mustaffa, Mohd Herwan Sulaiman, Ferda Ernawan, dan Yuhanis Yusof pada tahun 2018. Penelitian ini berisi tentang prediksi terhadap insiden demam berdarah yang ada di Yogyakarta menggunakan data bulanan dari tahun 2001 hingga 2013. Metode yang digunakan pada penelitian ini adalah *Least Squares Support Vector Machines* (LSSVM) dan *Flower Pollination Algorithm* (FPA). Variabel yang digunakan untuk memprediksi adalah nilai kumulatif curah hujan (mm), rata-rata temperatur (oC), rata-rata nilai kelembaban (%), dan jumlah kasus demam berdarah. Penelitian ini dapat menjadi referensi penelitian penulis terkait prediksi kejadian luar biasa kasus demam berdarah dengan metode SVM dan FPA [10].

Pada penelitian yang berjudul *Morbidity Rate Prediction of Dengue Hemorrhagic Fever (DHF) Using the Support Vector Machine and the Aedes aegypti Infection Rate in Similar Climates and Geographical Areas* oleh Kraissak Kesorn, Phatsavee Ongruk, Jakkrawarn Chomposri, Atchara Phumee, Usavadee Thavara, Apiwat Tawatsin, dan Padet Siriyasatien pada tahun 2015. Penelitian ini berisi tentang prediksi

morbiditas demam berdarah dengan memanfaatkan data tingkat infeksi demam berdarah menggunakan metode SVM. Tujuan dari penelitian ini adalah untuk meningkatkan pengawasan terhadap demam berdarah di daerah dengan iklim serupa. Penelitian ini juga membandingkan SVM dengan beberapa metode lainnya seperti *neural network*, *decision tree*, KNN, SVM-L, SVM-P, dan SVM-R [8].

Pada penelitian yang berjudul *Using Public Open Data to Predict Dengue Epidemic Assessment of Weather Variability, Population Density, and Land use as Predictor Variables for Dengue Outbreak Prediction using Support Vector Machine* oleh Chiung Ching Ho, Choo-Yee Ting, dan Dhesi Baha Raja pada tahun 2018. Penelitian ini berisi tentang prediksi wabah demam berdarah dengan menggunakan data variabilitas, populasi, penggunaan lahan dan letak geografis di Malaysia. Metode yang digunakan dalam melakukan peramalan adalah SVM. Hasil dari penelitian menunjukkan bahwa populasi memiliki korelasi paling besar dengan kejadian luar biasa dari demam berdarah [11].

## **2.2 Dasar Teori**

Pada sub bab ini akan dibahas mengenai dasar teori yang berhubungan dengan tugas akhir ini.

### **2.2.1 Demam Berdarah Dengue**

Demam dengue atau biasa disebut demam berdarah merupakan ancaman global terhadap kesehatan masyarakat, menyebabkan banyak beban besar terutama negara-negara yang berada di subtropis. Penyakit demam berdarah disebabkan oleh virus dengue (DENV), anggota *genus flavivirus*, yang ditularkan melalui gigitan nyamuk betina (*Aedes aegypti* dan *Aedes albopictus*). Penyakit ini dapat menyerang orang dewasa maupun anak-anak, bahkan dapat menyerang wanita hamil yang juga dapat berakibat fatal pada kandungannya [12].

*World Health Organization* (WHO) melaporkan bahwa setelah Kejadian Luar Biasa (KLB) dengue yang terjadi di Filipina pada tahun 1955 mengakibatkan peningkatan kejadian demam berdarah. Berdasarkan hasil estimasi kasus demam berdarah yang dilakukan oleh WHO, terdapat sekitar 390 juta kasus demam berdarah setiap tahunnya [1].

*Aedes Aegypti*, *Aedes Albopictus*, dan *Aedes Scutellaris* merupakan nyamuk penular penyakit demam berdarah di Indonesia, tetapi hingga saat ini yang menjadi nyamuk penular penyakit demam berdarah yang paling dominan adalah *Aedes Aegypti*. Penyakit demam berdarah di Indonesia awalnya ditemukan pada tahun 1968 di Surabaya dengan kasus 58 anak dan 24 diantaranya meninggal dunia. Setelah kejadian tersebut, jumlah kasus demam berdarah cenderung terus meningkat dan terus menyebar di berbagai daerah lainnya [13]. Berdasarkan Pusat Data dan Informasi Kementerian Kesehatan RI, pada tahun 2016 terdapat 204.171 kasus demam berdarah serta terdapat 1.500 kasus kematian yang dilaporkan [3]. Pada tahun 2016, di Kota Malang terdapat 464 warga yang menderita demam berdarah [4]. Peningkatan kasus demam berdarah ini sangat erat kaitannya terhadap iklim di daerah subtropis, urbanisasi yang tidak terkendali, peningkatan populasi penduduk yang signifikan, serta keganasan virus dengue.

Penderita penyakit demam berdarah dapat diidentifikasi dari beberapa gejala yang muncul. Gejala tersebut ditandai dengan demam tinggi selama 2 hingga 7 hari, nyeri ulu hati, badan berasa lemas, serta bintik-bintik merah dan ruam pada kulit. Pasien juga dapat kehilangan kesadaran secara tiba-tiba yang diakibatkan oleh kebocoran plasma yang dapat berakibat fatal [14].

### **2.2.2 Kejadian Luar Biasa (Outbreak)**

Kejadian luar biasa adalah timbulnya atau meningkatnya kejadian kesakitan dan atau kematian yang bermakna secara

epidemiologi pada suatu daerah dalam kurun waktu tertentu. Kejadian luar biasa penyakit dapat mengakibatkan terjadinya peningkatan kematian yang besar, yang juga dapat berdampak pada pariwisata, ekonomi, dan sosial. Kejadian luar biasa penyakit membutuhkan perhatian dan penanganan oleh semua pihak terkait dan perlu dideteksi sejak dini yang diikuti dengan tindakan penanganan yang cepat dan tepat. Penanggulangan kejadian luar biasa penyakit diatur dalam UU No. 4 tahun 1984 tentang Wabah Penyakit Menular, PP No. 40 tahun 1991 tentang Penanggulangan Wabah Penyakit Menular, Peraturan Menteri Kesehatan No. 560 tentang Jenis Penyakit Tertentu Yang Dapat Menimbulkan Wabah, dan lainnya. Aturan yang digunakan dalam menentukan suatu kecamatan mengalami kejadian luar biasa atau tidak adalah sebagai berikut [14]:

- a. Jumlah penderita demam berdarah baru dalam periode 1 bulan menunjukkan kenaikan dua kali atau lebih dibandingkan dengan angka rata-rata perbulan dalam periode dalam tahun sebelumnya.
- b. Jumlah penderita demam berdarah mengalami peningkatan dua kali atau lebih dibandingkan dengan periode sebelumnya.

### **2.2.3 Letak Geografis**

Kabupaten Malang merupakan daerah terbesar kedua di Jawa Timur dengan luas mencapai 3.534,86 km<sup>2</sup> atau 383.486 ha. Kabupaten Malang merupakan kabupaten terluas kedua di Jawa Timur dan merupakan kabupaten dengan populasi terbesar di Jawa Timur. Dengan keberagaman kondisi geografis menjadikan Kawasan ini memiliki potensi wisata yang cukup besar. Namun, beberapa daerah pada kabupaten Malang yang terletak di lereng gunung, pantai, dan laut dapat membuat nyamuk demam berdarah tumbuh dan berkembang dengan cepat. Kabupaten Malang dikabarkan menjadi daerah endemis dengue tertinggi di Jawa Timur. Letak geografis pada

kabupaten Malang sendiri terbagi menjadi 3 bagian, yaitu dataran rendah, dataran menengah, dan dataran tinggi [15].

### 2.2.4 Praproses Data

Praproses merupakan tahapan untuk mempersiapkan data mentah sebelum melakukan tahapan lain. Pada umumnya, tahapan ini dilakukan dengan cara mengeliminasi data yang tidak sesuai atau mengubah data menjadi bentuk yang lebih mudah diproses oleh komputer. Tahap praproses ini terdiri dari [16]:

#### a. Menangani Data Yang Hilang

Dikatakan data yang hilang (*missing value*) apabila jika terdapat nilai yang hilang pada salah satu atribut atau *record*. Penanganan data yang hilang dapat dilakukan dengan beberapa pendekatan, seperti menghapus atribut maupun record yang memiliki missing value atau mengisi missing value tersebut sebelum melakukan tahapan lainnya. Berikut adalah beberapa cara untuk mengisi data yang hilang.

1) Metode *most common* (MC), data yang hilang diganti dengan nilai rata-rata untuk atribut dengan tipe numerik dan modus untuk atribut dengan tipe nominal.

2) Konsep *most common* (CMC), cara ini hampir sama dengan metode *most common* dalam pengisian data yang hilang tetapi melakukan pertimbangan pada atribut target atau kelas yang sama.

3) *K-Nearest Neighbor* (KNNI), cara ini dilakukan dengan membuat kluster dari data, lalu melakukan metode *most common* diantara semua kluster yang terbentuk.

Pendekatan untuk mengisi data yang hilang menggunakan metode *most common* yang digunakan untuk mengatasi data yang hilang dalam pengerjaan tugas akhir ini.

#### b. Mengatasi Data Yang Tidak Berimbang

Terdapat dua kondisi pada suatu *dataset* yaitu *balanced data* dan *imbalanced data*. *Imbalanced data* atau data yang tidak berimbang merupakan kondisi dimana pada *dataset* terdapat satu kelas yang memiliki jumlah *instance* yang kecil dibandingkan dengan kelas lainnya. Kondisi ini dapat mengakibatkan hasil prediksi yang tidak optimal. Terdapat dua pendekatan yang dapat diterapkan untuk mengatasi permasalahan tersebut yaitu pendekatan level data (*sampling technique*) dan level algoritma. Pendekatan level data terdiri dari 2 metode yaitu *over-sampling* kelas terkecil dan *under-sampling* kelas terbesar. *The Synthetic Minority Oversampling Technique* (SMOTE) merupakan salah satu algoritma dari metode *over-sampling* kelas terkecil [17]. Algoritma tersebut digunakan dalam pengerjaan tugas akhir ini untuk mengatasi data yang tidak berimbang.

c. Pembagian Data Latih Dan Uji

Pembagian data dilakukan menggunakan metode *holdout* yaitu data dibagi secara acak menjadi data latih dan data uji. Pembagian data pada skenario tersebut adalah 80% data latih dan 20% data validasi.

Variabel target atau independent diprediksi menggunakan variabel dependen satu periode sebelumnya sehingga untuk memprediksi KLB atau tidak pada bulan Februari dibutuhkan data suhu, kelembaban, curah hujan, dll pada bulan Januari.

### 2.2.5 Penggalian Data

Penggalian data merupakan proses menggunakan penggabungan metode analisis tradisional dan algoritma semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* agar dapat mengekstraksi dan mengidentifikasi informasi pengetahuan

baru. Penggalian data memiliki tugas yang dikategorikan menjadi dua kategori utama, yaitu [18]:

a. Tugas Prediksi

Tugas ini bertujuan untuk memprediksi nilai dari atribut tertentu berdasarkan nilai dari atribut lainnya. Atribut yang digunakan untuk membuat prediksi disebut variabel independent sedangkan atribut yang diprediksi dikenal dengan atribut target atau variabel dependen.

b. Tugas Deskriptif

Tugas ini bertujuan untuk mencari pola yang dapat menyimpulkan hubungan dalam suatu data. Tugas deskriptif terdiri dari beberapa model diantaranya adalah analisa asosiasi yang digunakan untuk menentukan pola yang mendeskripsikan fitur-fitur data yang saling berhubungan dan analisa kluster yang merupakan proses untuk mencari kelompok-kelompok data berdasarkan kemiripan antar atribut.

Pengerjaan tugas akhir ini fokus terhadap prediksi dibandingkan peramalan karena didasarkan pada faktor-faktor yang mempengaruhi variabel dependen (KLB). Pengerjaan tugas akhir ini diharapkan dapat menghasilkan model prediksi yang akurat serta dapat menunjukan variabel-variabel yang signifikan mempengaruhi terjadinya KLB.

### **2.2.6 Flower Pollination Algorithm**

*Flower Pollination Algorithm* (FPA) dibuat oleh Yang pada tahun 2012 yang terinspirasi dari proses penyerbukan bunga. Tujuan utama bunga adalah berkembang biak melalui penyerbukan. Penyerbukan bunga dapat terjadi apabila terdapat serbuk sari yang sering dikaitkan dengan penyerbuk. Penyerbukan ini terdiri atas dua bentuk, yaitu abiotik dan biotik [19]. Kebanyakan dari bunga bergantung pada penyerbukan biotik dimana serbuk sari ditransfer ke kepala putik saat organisme menyentuhnya. Angin dan air juga dapat membantu

penyerbukan tersebut secara abiotik. Penyerbukan dapat terjadi dengan dua cara yaitu, *self-pollination* dan *cross-pollination*. *Self-pollination* merupakan penyerbukan pada satu bunga yang dilakukan oleh serbuk sari dari bunga itu sendiri. *Cross-pollination* merupakan penyerbukan yang dilakukan serbuk sari dari bunga yang berbeda. Tujuan dari penyerbukan adalah untuk bertahan hidup dan mengoptimalkan alat reproduksi dari tumbuhan tersebut.

Terdapat empat kaidah penting yang digunakan, antara lain [19];

- a. Kaidah satu, merupakan *global pollination* yang berupa biotik dan *cross-pollination*. Penyerbukan ini mengikuti aturan dari *levy distribution* dengan melompat atau terbang dengan jarak yang jauh.
- b. Kaidah dua, merupakan *local pollination* yang berupa abiotik dan *self-pollination*.
- c. Kaidah tiga, merupakan *flower constancy* yang dikembangkan oleh serangga yang mana senilai dengan probabilitas reproduksi dan kemiripan dua bunga yang terlibat.
- d. Kaidah empat, merupakan interaksi *local* dan *global pollination* yang dikendalikan oleh probabilitas  $p \in [0,1]$ , disebut juga sebagai *switch probability*.

Pada FPA, terdapat dua tahapan kunci yaitu penyerbukan global dan lokal. Pada penyerbukan global, kaidah satu dan tiga digunakan untuk menemukan solusi pada langkah selanjutnya ( $x_i^{t+1}$ ) menggunakan nilai dari langkah sebelumnya ( $x_i^t$ ). Formulasi penyerbukan global adalah sebagai berikut.

$$x_i^{t+1} = x_i^t + L(x_i^t - g^*) \quad (2.1)$$

Penyerbukan lokal menggunakan kaidah ketiga tentang *flower constancy* yang ditunjukkan pada persamaan berikut.

$$x_i^{t+1} = x_i^t + \varepsilon(x_j^t - x_k^t) \quad (2.2)$$



Subscript  $i$  menunjukkan serbuk ke- $i$  (atau bunga) dan persamaan (2.1) diimplementasikan untuk serbuk pada bunga.  $g^*$  merupakan solusi terbaik pada saat itu.  $L$  merupakan jarak terbang yang didapatkan dari *levy distribution*.  $x_j^t$  dan  $x_k^t$  merupakan solusi dari tanaman yang berbeda. Berdasarkan keempat kaidah tersebut, *switch probability* ( $p$ ) digunakan untuk memilih tipe penyerbukan yang akan mengontrol proses optimasi pada iterasi [19].

Berikut adalah *pseudo-code* acuan yang digunakan dalam pengerjaan tugas akhir ini [19].

---

**Algorithm 1:** Flower Pollination Algorithm

---

```

Definisikan fungsi objektif;
Inisialisasi jumlah populasi dengan solusi acak;
Definisikan switch probability ( $p$ );
Definisikan d-dimensi dari fungsi objektif;
Temukan solusi terbaik ( $g^*$ ) dari inisial populasi;
while  $t < MaxIterasi$  do
  for  $i \leftarrow 0$  to  $n$  (semua  $n$  solusi dalam populasi) do
    if  $rand < p$  then
      | Global pollination menggunakan persamaan (2.1);
    else
      | Local pollination menggunakan persamaan (2.2);
    end
    Evaluasi solusi baru;
    Jika solusi baru lebih baik, update solusi tersebut kedalam populasi;
  end
  tentukan solusi terbaik saat ini( $g^*$ );
   $t = t + 1$ ;
end

```

---

**Gambar 2.1 Pseudo-code Flower Pollination Algorithm**

## 2.2.7 Support Vector Machine

*Support vector machines* (SVM) merupakan salah satu bidang metode *supervised machine learning* yang awalnya dikembangkan oleh Vapnik. Latar belakang dari metode ini adalah kombinasi dari teori statistik dan optimalisasi pada matematika. Metode ini telah diterapkan pada berbagai bidang, seperti peramalan *time-series*, analisa regresi, pengenalan pola, dll.

Konsep dari SVM adalah mencari *hyperplane* terbaik yang merupakan pemisah dua buah kelas pada *input space*.

*Hyperplane* pemisah terbaik antara dua kelas didapatkan dari pengukuran *margin*. *Margin* yang dimaksud merupakan jarak terdekat antara data dan *hyperplane* tersebut. SVM pada umumnya merupakan *linear classifier*, tetapi juga mampu menyelesaikan permasalahan *non linear* dengan menggunakan dua tahap. Tahapan pertama, data pelatihan diproyeksikan ke vektor baru berdimensi tinggi sehingga data itu dapat terpisah secara linier. Tahap selanjutnya, SVM mencari *hyperplane* optimal dengan cara yang sama seperti yang telah disebutkan sebelumnya. Pembagian dua kelas dari *hyperplane* tersebut dapat dibagi kedalam kelas +1 dan -1. Jika kedua kelas +1 dan -1 dan *hyperplane* berdimensi  $d$  maka dapat didefinisikan sebagai [20]:

$$\vec{w} \cdot \vec{x} + b = 0 \quad (2.3)$$

Pola  $\vec{x}_1$  yang masuk pada kelas -1 dapat memenuhi pertidaksamaan berikut:

$$\vec{w} \cdot \vec{x} + b \leq -1 \quad (2.4)$$

Pola  $\vec{x}_i$  yang masuk pada kelas +1 dapat memenuhi pertidaksamaan berikut:

$$\vec{w} \cdot \vec{x} + b \geq -1 \quad (2.5)$$

*Margin* terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara *hyperplane* dan titik terdekatnya, yaitu  $\frac{1}{\|\vec{w}\|}$ .

Kernal SVM yang digunakan dalam tugas akhir ini adalah *radial basis function kernel*. Terdapat 2 parameter SVM dengan kernel *radial* yaitu nilai *cost* (C) dan *gamma* ( $\gamma$ ). Pembagian data dilakukan menggunakan metode *holdout* yaitu dengan membagi data menjadi 80% data pelatihan dan 20% data validasi. Adapun langkah-langkah untuk melakukan prediksi adalah sebagai berikut.

- a. Pencarian parameter optimal menggunakan algoritma penyerbukan bunga yang terdapat pada sub bab 2.2.6.
- b. Melakukan pelatihan model SVM dengan parameter optimal menggunakan data pelatihan.

- c. Melakukan pengujian pada data validasi
- d. Melakukan prediksi

Parameter yang terdapat pada SVM dengan kernel *Radial Basis Function* adalah parameter *cost* (  $C$  ) dan  $\gamma$ . Rentang nilai dari parameter  $C$  yang baik adalah  $0.03125 - 32768$  sedangkan rentang nilai dari parameter  $\gamma$  yang baik adalah  $0.00003 - 8$ . FPA menginisialisasi populasi acak seragam ( $n$ ) menggunakan *levy random walk*. SVM dilatih dan diuji  $n$  kali dan akurasi terbaik dipilih, dimana parameter  $C$  dan  $\gamma$  disimpan sebagai parameter optimal, sehingga algoritma ini dapat menentukan kombinasi parameter dengan bilangan riil terbaik untuk digunakan oleh SVM [9].

### 2.2.8 Evaluasi Kinerja

Evaluasi kinerja dilakukan dengan tujuan untuk menilai apakah suatu model prediksi dapat memprediksi dengan akurat. Evaluasi model prediksi umumnya dimulai dengan membuat *confusion matrix*. Tabel 2.1 merupakan contoh dari *confusion matrix*. Akurasi, *recall* dan *precision* dapat diukur melalui *confusion matrix*. *Accuracy* digunakan untuk mengetahui seberapa besar ketepatan prediksi model SVM, *recall* digunakan untuk mengetahui prosentase 'True Positive' yang benar, lalu untuk *precision* digunakan untuk mengetahui persentase 'True Negative' yang dapat diketahui dengan benar.

Tabel 2.1 Confusion Matrix

		<i>Actual Class</i>	
		Actual Negative	Actual Positive
Predicted Class	Predicted Negative	TN	FN
	Predicted Positive	FP	TP

Rumus perhitungan yang digunakan adalah sebagai berikut [21].

$$\text{Akurasi (\%)} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (2.6)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2.7)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2.8)$$

Namun, tolok ukur performa model prediksi yang dilakukan pada data tidak berimbang tidak dapat diinterpretasi langsung dari nilai masing-masing metrik pengukuran seperti akurasi, *precision*, dan *recall* tetapi diperlukan interpretasi berupa *Precision-Recall Curve* (PRC) [22]. PRC dibuat berdasarkan ambang batas dalam penentuan kelas. Nilai ambang batas pada *library* sklearn terdiri atas 2 jenis, yaitu '*predict\_proba*' dan '*decision\_function*'. *Predict\_proba* memiliki nilai 0 sampai 1 dimana apabila ambang batas = 0,1 maka apabila model memprediksi suatu entitas (aktual positif) memiliki probabilitas 0,2 sebagai kelas positif maka model tersebut sukses memprediksi. Disisi lain, apabila ambang batas = 0,5 dan model memprediksi suatu entitas (aktual positif) memiliki probabilitas 0,2 sebagai kelas positif maka model tersebut akan salah memprediksi. *Decision\_function* merupakan fungsi yang menerima dataset sebagai masukan dan keputusan sebagai luaran. Jumlah maksimal batas ambang pada jenis ini didasarkan pada jumlah data. Luaran dapat berupa angka yang tidak terbatas. Setiap nilai ambang batas membentuk titik ( $x = \text{recall}$ ,  $y = \text{precision}$ ) sehingga dari kumpulan titik tersebut dapat dibentuk suatu garis. Skor *Average Precision* (AP) dapat dianalisa setelah nilai *precision* dan *recall* dari setiap ambang batas didapatkan. Skor AP merupakan rata-rata terbobot dari *precision* pada setiap ambang probabilitas, diikuti dengan

peningkatan *recall* pada ambang batas sebelumnya atau dapat diformulasikan sebagai berikut.

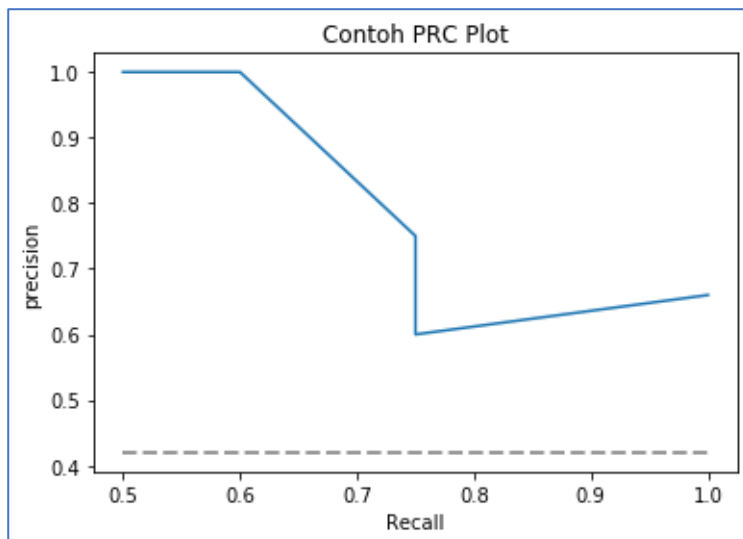
$$AP = \sum_n (R_n - R_{n-1}) P_n \quad (2.9)$$

Dimana  $P_n$  dan  $R_n$  secara berturut-turut adalah nilai *precision* dan *recall* pada  $n$  ambang batas. Berikut adalah contoh sederhana dalam melakukan perhitungan PRC. Pada contoh berikut, dibuat model prediksi untuk menentukan apakah seekor tikus dikatakan obesitas (kelebihan lemak, lebih berat belum tentu masuk kategori obesitas) berdasarkan berat dari tikus tersebut. Tabel 2.2 berisi data berat tikus, kelas obesitas,  $p$  (probabilitas tikus dikatakan obesitas berdasarkan model prediksi), dan  $T$  yang merupakan ambang batas. Apabila ambang batas = 0,2; maka tikus yang memiliki hasil prediksi dengan probabilitas obesitas  $> 0,2$  diprediksi obesitas dan  $\leq 0,2$  diprediksi tidak obesitas. Apabila ambang batas = 0,4; maka tikus yang memiliki hasil prediksi dengan probabilitas obesitas  $> 0,4$  diprediksi obesitas dan  $\leq 0,4$  diprediksi tidak obesitas. Berdasarkan hal tersebut, dibentuk *confusion matrix* sederhana yang dapat menghasilkan nilai *recall* dan *precision* seperti pada Tabel 2.2.

**Tabel 2.2 Data Berat Tikus**

Berat	Aktual	$p$	T=0	T=0,2	T=0,4	T=0,6	T=0,8	T=1
3	0	0	TN	TN	TN	TN	TN	TN
5	0	0	TN	TN	TN	TN	TN	TN
9	1	0.11	TP	FN	FN	FN	FN	FN
11	0	0.25	FP	FP	TN	TN	TN	TN
16	0	0.9	FP	FP	FP	FN	FN	TN
18	1	0.95	TP	TP	TP	TP	TP	FN
19	1	1	TP	TP	TP	TP	TP	TP
20	1	1	TP	TP	TP	TP	TP	TP
<b>RECALL</b>			1	0.75	0.75	0.6	0.6	0.5
<b>PRECISION</b>			0.66	0.6	0.75	1	1	1

Nilai recall dan precision tersebut membentuk titik (recall, precision) dengan hasil visualisasi seperti pada Gambar 2.2.

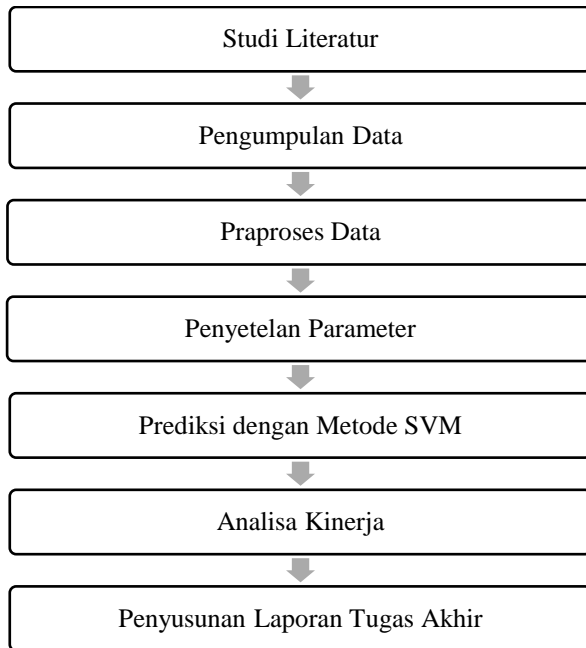


**Gambar 2.2** Contoh Plot PRC

PRC juga memiliki garis basis, dimana garis basis tersebut merupakan rasio dari kelas minoritas. Nilai *precision* dan *recall* yang baik berada diatas garis basis tersebut.

## **BAB III METODOLOGI**

Pada bab ini akan dijelaskan mengenai tahapan – tahapan yang dikerjakan dalam pelaksanaan tugas akhir sebagai panduan pengerjaan yang sistematis. Adapun metodologi yang digunakan pada tugas akhir ini disajikan pada Gambar 3.1.



**Gambar 3.1 Tahapan Pengerjaan**

### **3.1 Studi Literatur**

Pada tahap ini, setelah mengetahui permasalahan, kemudian menyusun rumusan masalah, batasan masalah, dan menetapkan tujuan. Studi literatur dilakukan dengan mengumpulkan paper dari jurnal dan buku. Selain melakukan pencarian literatur tentang penyakit demam berdarah, diperlukan untuk mencari literatur terkait metode yang dipilih. Hal ini dilakukan untuk menambah pemahaman terkait demam berdarah, konsep

metode yang dipilih, dan cara mengimplementasikan metode tersebut.

### 3.2 Pengumpulan Data

Pada tahap ini dilakukan pencarian dan pemilihan data yang akan diteliti sehingga dapat mendukung dalam pengerjaan tugas akhir ini. Data yang digunakan merupakan data resmi dari penderita demam berdarah yang didapatkan dari Dinas Kesehatan Kabupaten Malang dengan periode bulanan dan data suhu, kelembaban, curah hujan, kecepatan angin yang didapatkan dari BMKG. Dua dataset tersebut akan diolah dengan cara menggabungkan kedua dataset tersebut pada periode yang sama. Terdapat 10 atribut data yang digunakan dalam pengerjaan tugas akhir ini. Atribut yang digunakan dalam pengerjaan Tugas Akhir ini beserta keterangannya terdapat pada Tabel 3.1

**Tabel 3.1 Deskripsi Umum Atribut Yang Digunakan**

No	Nama Atribut	Tipe Atribut	Keterangan
1.	SU	Interval	Suhu udara (derajat celcius)
2.	KU	Rasio	Kelembaban (%)
3.	CH	Rasio	Curah hujan (mm)
4.	KA	Rasio	Kecepatan angin(m/s)
5.	KP	Rasio	Kepadatan penduduk
6.	ABJ	Rasio	Angka bebas jentik(%)
7.	KDB	Rasio	Jumlah kejadian demam berdarah
8.	Periode	Nominal	Periode dalam bulanan (Januari - Desember)
9.	Kecamatan	Nominal	Nama kecamatan
10.	KLB	Nominal	Merupakan atribut target yang menunjukkan apakah merupakan kejadian luar biasa atau tidak (0/1).

### 3.3 Praproses Data

Data yang didapatkan dari tahap sebelumnya merupakan data yang masih mentah berupa data periode bulanan yang masih belum teratur, kemudian dilakukan pengolahan data. Pengolahan data ini dilakukan dengan mengatasi missing value dan duplikasi pada data. Tahapan dan metode untuk mengatasi



missing value dan duplikasi pada data terdapat dalam sub bab 2.2.4. Pada tahapan ini dilakukan normalisasi data serta dilakukan tahapan untuk mengatasi data yang tidak berimbang menggunakan algoritma SMOTE serta dilakukan pemisahan data menjadi data latih dan data validasi.

### 3.4 Penyetelan Parameter

Pada tahap ini dilakukan pencarian parameter SVM yang paling optimal dengan menggunakan *flower pollination algorithm* (FPA). Parameter SVM yang dicari menggunakan metode FPA adalah nilai *cost* ( $C$ ) dan  $\gamma$ . Optimasi parameter SVM ini dibutuhkan karena *search space* dari kombinasi parameter  $C$  dan  $\gamma$  tidak memiliki kisaran rentang tepat (tidak terbatas). Penyetelan parameter yang tidak tepat menyebabkan hasil prediksi yang buruk. Nilai parameter  $C$  mempengaruhi besar margin minimum pada hasil prediksi, dimana semakin besar nilai  $C$  maka semakin kecil nilai margin. Nilai parameter  $\gamma$  menentukan seberapa jauh pengaruh dari satu sampel *training dataset* dengan nilai rendah berarti “jauh”, dan nilai tinggi berarti “dekat”. Dengan  $\gamma$  yang rendah, titik yang berada jauh dari garis pemisah yang masuk akal dipertimbangkan dalam perhitungan untuk garis pemisah. Ketika  $\gamma$  tinggi berarti titik – titik berada di sekitar garis yang masuk akal akan dipertimbangkan dalam perhitungan. Oleh karena itu, pencarian nilai parameter  $C$  dan  $\gamma$  (bersifat unik) dilakukan menggunakan FPA dan tidak menggunakan parameter dari studi sebelumnya.

Metode FPA ini akan mencari solusi yang paling optimal dari parameter tersebut berdasarkan penyerbukan bunga oleh serangga. Parameter yang didapatkan dari metode ini akan digunakan sebagai parameter SVM pada tahap selanjutnya. Adapun detail langkah-langkah pencarian parameter menggunakan metode FPA terdapat dalam sub bab 2.2.6.

### **3.5 Prediksi Menggunakan Support Vector Machine**

Pada tahap ini dilakukan prediksi data demam berdarah yang didapatkan dari Dinas Kesehatan Kabupaten Malang menggunakan metode SVM. Prediksi menggunakan metode SVM akan dilakukan dua kali. Prediksi pertama menggunakan metode SVM pada data pelatihan menggunakan parameter acak dan prediksi kedua menggunakan metode SVM pada data pelatihan menggunakan parameter optimal yang didapatkan dari metode FPA.

Prediksi dilakukan dengan masukan data pelatihan serta parameter yang telah ditetapkan pada tahap praproses data. Selanjutnya model prediksi diuji menggunakan data validasi. Apabila model prediksi tersebut menghasilkan kinerja yang baik, maka model tersebut akan digunakan untuk memprediksi kejadian luar biasa demam berdarah masa mendatang. Tahapan dalam melakukan prediksi menggunakan SVM terdapat dalam sub bab 2.2.7.

### **3.6 Analisa Kinerja**

Pada tahap ini dilakukan analisa terhadap performa model SVM dengan menghitung nilai rata-rata akurasi, presisi, dan *recall* dari setiap model prediksi menggunakan confusion matriks yang terdapat pada sub bab 2.2.8. Adapun data yang diprediksi dan di evaluasi adalah data validasi untuk mendapatkan model terbaik. Hasil dari analisa ini akan menjadi acuan penarikan kesimpulan dari tugas akhir ini, sehingga dapat dijadikan sumber untuk pengambilan keputusan di masa mendatang.

### **3.7 Penyusunan Tugas Akhir**

Pada Tahapan terakhir dalam pengerjaan tugas akhir ini adalah menggabungkan seluruh kegiatan untuk dijadikan laporan tugas akhir. Tahapan penulisan laporan tugas akhir ini dijelaskan sebagai berikut.

## **1. BAB I PENDAHULUAN**

Bab ini akan dijelaskan mengenai latar belakang tugas akhir, rumusan dan batasan masalah, serta tujuan dan manfaat tugas akhir ini.

2. **BAB II TINJAUAN PUSTAKA**

Bab ini akan dijelaskan mengenai penelitian serupa yang telah dilakukan dan juga dasar-dasar teori yang digunakan penulis untuk mendukung pengerjaan tugas akhir ini.

3. **BAB III METODOLOGI**

Bab ini akan dijelaskan mengenai tahapan dalam pengerjaan tugas akhir ini.

4. **BAB IV PERANCANGAN**

Bab ini akan dijelaskan detail rancangan yang akan dilakukan untuk mengimplementasikan metode yang digunakan.

5. **BAB V IMPLEMENTASI**

Bab ini akan berisi mengenai implementasi dalam menyelesaikan tujuan tentang permasalahan yang dibahas pada tugas akhir ini.

6. **BAB VI HASIL DAN PEMBAHASAN**

Bab ini berisi tentang hasil pembahasan dalam penyelesaian masalah yang diangkat pada tugas akhir ini.

7. **BAB VII KESIMPULAN DAN SARAN**

Bab ini berisi kesimpulan dan saran sebagai pelengkap tugas akhir ini.

*Halaman ini sengaja dikosongkan*

## BAB IV PERANCANGAN

Pada bab ini akan menjelaskan mengenai bagaimana rancangan dari penelitain tugas akhir yang terdiri dari mendefinisikan ruang lingkup penelitian dan bagaimana penelitain ini akan dilakukan.

### 4.1 Pengumpulan Data dan Eksplorasi

Pada tahap ini dilakukan proses pengumpulan data hingga data dapat digunakan untuk pengerjaan tugas akhir ini. Variabel-variabel yang dikumpulkan beserta sumbernya dijelaskan pada Tabel 4.1.

**Tabel 4.1 Sumber Data yang Digunakan**

No	Variabel	Sumber
1.	Suhu Udara	Situs pusat data BMKG
2.	Kelembaban Udara	Situs pusat data BMKG
3.	Curah Hujan	Situs pusat data BMKG
4.	Kecepatan Angin	Situs pusat data BMKG
5.	Kepadatan Penduduk	BPS
6.	Angka bebas Jentik	Dinas Kesehatan Kabupaten Malang
7.	Kejadian Demam Berdarah	Dinas Kesehatan Kabupaten Malang

Eksplorasi dan praproses data akan dilakukan terhadap setiap variabel untuk mendapatkan wawasan awal menggunakan *library* matplotlib.

### 4.2 Praproses Data

Tahap praproses dilakukan agar model prediksi dapat memprediksi KLB dengan akurat. Terdapat beberapa langkah yang dilakukan pada tahap ini, yaitu mengubah data harian variabel iklim, menangani data yang hilang, normalisasi data, oversampling data, dan membagi data latih dan data validasi.

### **4.2.1 Mengubah Data Harian Variabel Iklim**

Data iklim (suhu udara, kelembaban udara, curah hujan, dan kecepatan angin) yang didapatkan dari situs pusat data Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) memiliki rentang waktu harian. Oleh karena itu, dibutuhkan perubahan rentang waktu menjadi bulanan agar dapat diproses ke tahap selanjutnya. Perubahan rentang waktu harian menjadi bulanan dilakukan dengan mengambil nilai rata-rata dari setiap variabel selama sebulan.

### **4.2.2 Menangani Data Yang Hilang**

Variabel yang didapatkan terdapat beberapa data yang hilang terutama pada data iklim dan angka bebas jentik. Pada penelitian ini, data yang hilang pada data iklim (suhu udara, kelembaban udara, curah hujan, dan kecepatan angin) dilakukan pengisian dengan nilai rata-rata pada bulan yang mengalami data yang hilang karena mempertimbangkan data iklim memiliki sifat musiman. Data yang hilang pada variabel angka bebas jentik dilakukan pengisian dengan nilai rata-rata dari seluruh sampel berdasarkan kecamatan.

### **4.2.3 Normalisasi Data**

Variabel yang didapatkan memiliki skala nilai dan satuan yang berbeda-beda, sehingga perlu dilakukan normalisasi untuk mengurangi dampak dari skala nilai yang berbeda-beda tersebut. Data iklim (suhu udara, kelembaban udara, curah hujan, dan kecepatan angin) memiliki skala nilai satuan dan puluhan, sedangkan kepadatan penduduk memiliki skala nilai ratusan hingga ribuan.

Pada penelitian ini, metode *standard normalization* (z-score) akan digunakan sebagai metode normalisasi data. Hasil dari normalisasi menggunakan metode ini adalah setiap variabel memiliki nilai rata-rata 0 dan nilai standar deviasi 1. Nilai variabel baru yang dihasilkan dari metode *standard normalization* (z-score) didapatkan dari perhitungan dengan Persamaan 4.1

dimana secara berturut-turut  $x_i$ ,  $\bar{x}$ ,  $s$  adalah data point, rata-rata sampel, dan standar deviasi sampel.

$$X_{baru} = \frac{x_i - \bar{x}}{s} \quad (4.1)$$

#### 4.2.4 Oversampling Data

Variabel target (Kejadian Luar Biasa/*outbreak*) memiliki jumlah kelas yang berbeda jauh. Jumlah data dengan kelas KLB terdapat 20% dari keseluruhan sampel yang dimiliki, sehingga model akan mengalami kesulitan untuk memprediksi kelas KLB. Gambar 4.1 menunjukkan perbedaan jumlah yang signifikan terhadap variabel target.



**Gambar 4.1** Persebaran Data Variabel Target (KLB)

Pada penelitian ini, metode oversampling digunakan untuk mengatasi permasalahan tersebut. Metode ini akan membuat sampel baru dengan kelas KLB sehingga jumlah data dengan kelas KLB dan tidak KLB akan memiliki jumlah yang seimbang.

#### 4.2.5 Membagi Data Latih dan Uji

Pembagian data menjadi data latih dan data validasi dilakukan terlebih dahulu sebelum melakukan pembuatan model. Data latih akan digunakan untuk melatih model Support Vector

Machine, sedangkan data validasi untuk mengevaluasi seberapa akurat model melakukan prediksi. Pembagian data latih dan data validasi ini akan dilakukan menggunakan *library sklearn* dalam modul *train\_test\_split*. Pada tahap ini, jumlah kelas target (KLB dan bukan KLB) disamakan pada data latih maupun data uji menggunakan parameter *stratify* pada modul *train\_test\_split*. Pembagian data dilakukan dengan menggunakan metode *holdout* digunakan pada dataset tahun 2010-2017 dengan komposisi 80% data latih dan 20% data uji serta menggunakan metode *5-fold cross validation* untuk melakukan pengecekan apakah hasil prediksi mengalami *overfitting/underfitting* atau tidak. Kejadian Luar Biasa pada tahun 2018 diprediksi menggunakan model SVM yang telah dilatih dan dievaluasi kembali.

### 4.3 Penyetelan Parameter

Pencarian parameter yang digunakan sebagai masukan SVM adalah C (*cost*) dan gamma. Pencarian parameter ini menggunakan *Flower Pollination Algorithm* (FPA). Pada FPA terdapat beberapa parameter dengan detail dari setiap parameter yang terdapat pada Tabel 4.2.

Tabel 4.2 Parameter FPA

Parameter dalam FPA	Keterangan
<i>Population</i>	Populasi bunga/polen, berfungsi untuk menyimpan solusi parameter dalam hal ini C dan gamma
<i>Switch probability</i>	Probabilitas perpindahan serangga
<i>Dimension</i>	Posisi bunga, dimensi=2 jika posisi bunga ditentukan oleh x dan y
<i>Low, High</i>	Batas posisi serangga berpindah



Tabel 4.2 tersebut menjelaskan detail dari istilah yang terdapat pada FPA.

Banyak iterasi pada semua percobaan akan dilakukan sebanyak 10, 20, 50 dan 100 iterasi pada setiap daftar percobaan. Adapun dilakukan beberapa percobaan dengan nilai inialisasi populasi dan *switch probability* yang berbeda-beda seperti pada Tabel 4.3.

**Tabel 4.3 Daftar Percobaan Pada FPA**

No	Inialisasi Populasi	Switch Probability
1.	10	0,2; 0,4; 0,6; 0,8; 1
2.	20	0,2; 0,4; 0,6; 0,8; 1
3.	50	0,2; 0,4; 0,6; 0,8; 1
4.	100	0,2; 0,4; 0,6; 0,8; 1

Langkah-langkah yang dilakukan dalam tahapan ini akan disesuaikan dengan *pseudo-code* FPA pada sub-bab 2.2.6. Fungsi objektif yang digunakan adalah memaksimalkan nilai rataan dari akurasi, *recall*, dan *precision*.

#### **4.4 Prediksi Menggunakan Support Vector Machine**

Prediksi menggunakan model SVM ini akan dilakukan menggunakan inisiasi parameter *default* dan parameter yang didapatkan dari FPA. Parameter-parameter yang didapatkan dari beberapa skenario yang dilakukan akan dibandingkan pada data latih dan data validasi serta akan ditentukan model terbaik dari skenario tersebut. Model terbaik akan digunakan untuk memprediksi Kejadian Luar Biasa pada dataset DBD tahun 2018.

#### **4.5 Analisa Kinerja**

Klasifikasi *Support Vector Machine* tersebut akan diuji performanya apakah model yang sudah dilakukan itu baik atau tidak. Analisa kinerja ini akan menggunakan evaluasi akurasi, *recall* dan presisi.

*Halaman ini sengaja dikosongkan*

## BAB V IMPLEMENTASI

Bab ini menjelaskan tentang implementasi dari bab perancangan yang telah dilakukan sesuai dengan metode pengembangan yang dibuat.

### 5.1 Persiapan Implementasi

Penelitian ini menggunakan perangkat keras dengan spesifikasi yang dapat dilihat pada Tabel 5.1.

**Tabel 5.1 Spesifikasi Perangkat Keras Yang Digunakan**

Jenis Spesifikasi	Spesifikasi
Tipe Perangkat	Laptop Lenovo G40
Prosesor	Intel® Core™ i7-4510U CPU @ 2.00 GHz
Ram	8 GB
Sistem Operasi	Windows 10 Pro 64-bit

Penelitian ini dijalankan dengan menggunakan beberapa perangkat lunak seperti code editor, bahasa pemrograman, dan library yang dapat dilihat pada Tabel 5.2.

**Tabel 5.2 Perangkat Lunak Yang Digunakan**

Jenis Perangkat Lunak	Perangkat Lunak Yang Digunakan
Bahasa Pemrograman	Python
Code Editor	Jupyter Notebook
Library	<ul style="list-style-type: none"><li>• Pandas</li><li>• Numpy</li><li>• Matplotlib</li><li>• Sklearn</li><li>• Imblearn</li></ul>

### 5.2 Pengumpulan dan Eksplorasi Data

Data yang telah dikumpulkan sesuai dengan variabel pada Tabel 4.1 akan dilakukan eksplorasi menggunakan *library*

matplotlib. *Library* matplotlib merupakan *library plotting* yang berguna untuk membuat visualisasi dari data.

### 5.2.1 Ringkasan Statistik

Ringkasan statistik merupakan salah satu komponen penting sebelum melakukan prediksi. Ringkasan statistik ditampilkan kedalam bentuk tabel yang berisi jumlah baris data, rata-rata, nilai terkecil, nilai kuartil (Q1, Q2, Q3), dan nilai terbesar. Kode Program 5.1 merupakan baris program yang digunakan untuk mendapatkan ringkasan statistic pada variabel numerik.

```
columns = [
    'suhu_udara', 'kelembaban_udara',
    'curah_hujan', 'kecepatan_angin',
    'kepadatan_penduduk',
    'angka_bebas_jentik',
    'kejadian_demam_berdarah'
]
dbd[columns].describe().T
```

**Kode Program 5.1 Potongan Kode Program Ringkasan Statistik**

### 5.2.2 Distribusi dan Hubungan Antar Variabel

Distribusi data akan divisualisasikan menggunakan *scatter plot*, *box plot* dan *pie plot*. Kode Program 5.2 merupakan baris program yang digunakan untuk menampilkan visualisasi dari *scatter plot*, *box plot* dan *pie plot*.

```
# Membuat Pie Plot
var = dbd['nama_variabel'].value_counts()
plt.pie(var.values, labels = var.index,
autopct='%1.1f%%', shadow=True)

# Membuat Box Plot
sns.boxplot("nama_variabel", data=dbd)

# Membuat Scatter Plot
sns.scatterplot("nama_variabel1",
"nama_variabel2", hue='outbreak', data=dbd)
```

**Kode Program 5.2 Potongan Kode Visualisasi dan Eksplorasi**


## 5.3 Praproses Data

Tahapan ini dilakukan untuk mengatasi berbagai permasalahan yang terdapat pada data sehingga model prediksi yang dibentuk

dapat memprediksi Kejadian Luar Biasa DBD dengan akurat. Terdapat beberapa langkah pengerjaan pada tahapan ini, yaitu mengubah data harian variabel iklim, menangani data yang hilang, normalisasi data, oversampling data, dan membagi data latih dan data uji

### 5.3.1 Mengubah Data Harian Variabel Iklim

Data iklim (suhu udara, kelembaban udara, curah hujan, dan kecepatan angin) yang didapatkan dari situs pusat data Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) memiliki rentang waktu harian. Rataan dalam 1 bulan digunakan untuk mendapatkan data dengan periode bulanan. Berikut adalah bentuk dari data iklim yang didapatkan dari situs pusat data Badan Meteorologi, Klimatologi, dan Geofisika (BMKG).

1		ID WMO	: 96943		
2		Nama Stasiun	: Stasiun Klimatologi Malang		
3		Lintang	: -7.90080		
4		Bujur	: 112.59790		
5		Elevasi	: 590		
6					
7					
8					
9	Tanggal	Tavg	RH_avg	RR	ff_avg
10	01-01-2010	24.1	86	6	1
11	02-01-2010	24.9	81	34	2
12	03-01-2010	25.4	78	0	0
13	04-01-2010	24.8	85	1	2
14	05-01-2010	23.5	85	0	2
15	06-01-2010	24	83	39	1
16	07-01-2010	23.9	84	9	0
17	08-01-2010	23.7	83	0	1
18	09-01-2010	22.7	90	4	2
19	10-01-2010	23.6	87	31	1
20	11-01-2010	23.5	88	0	1
21	12-01-2010	24	84	9	1
22	13-01-2010	24.1	85	3	2
23	14-01-2010	24.7	76	9	2
24	15-01-2010	23.7	78	3	2
25	16-01-2010	25.4	68	4	4
26	17-01-2010	25.8	68	23	3
27	18-01-2010	26.1	65	4	4
28	19-01-2010	24.5	82	2	0
29	20-01-2010	22.7	87	4	0
30	21-01-2010	23.2	88	12	1
31	22-01-2010	23.7	85	14	1
32	23-01-2010	23.4	90	9	0
33	24-01-2010	23.3	86	14	0
34	25-01-2010	22.9	90	30	0
35	26-01-2010	22.5	88	59	0
36	27-01-2010	21.9	88	9	1
37	28-01-2010	23.1	85	13	1
38	29-01-2010	23.6	78	0	0
39	30-01-2010	23.9	82	0	1
40	31-01-2010	22.6	92	6	0
41		23.85	83.06	11.32	1.16
42					
43					
44	Keterangan :				
45	8888: data tidak terukur				
46	9999: Tidak Ada Data (tidak dilakukan pengukuran)				
47	Tavg: Temperatur rata-rata (°C)				
48	RH_avg: Kelembapan rata-rata (%)				
49	RR: Curah hujan (mm)				
50	ff_avg: Kecepatan angin rata-rata (m/s)				

Gambar 5.1 Data Iklim

Gambar 5.1 merupakan contoh dari data iklim yang didapatkan dari situs BMKG. Untuk mendapatkan nilai rata-rata tersebut digunakan fitur formula pada Microsoft Excel. Tabel 5.3 menjelaskan formula yang digunakan untuk mendapatkan rata-rata periode bulanan.

**Tabel 5.3 Formula Mengubah Data Harian Variabel Iklim**

Variabel	Formula
Temperatur rata-rata	=AVERAGE(B10:B40)
Kelembaban rata-rata	=AVERAGE(C10:C40)
Curah hujan	=AVERAGE(D10:D40)
Kecepatan angin rata-rata	=AVERAGE(E10:E40)

### 5.3.2 Menangani Data Yang Hilang

Data yang hilang pada data iklim ditangani dengan cara melakukan pengisian menggunakan nilai rata-rata pada bulan yang mengalami kehilangan data tersebut. Tabel 5.4 merupakan contoh penanganan data yang hilang pada data suhu di Turen pada Januari 2014.

**Tabel 5.4 Contoh Penangan Data Yang Hilang Pada Data Iklim**

Periode	Sebelum Imputasi	Setelah Imputasi
Januari 2010	26.02	26.02
Januari 2011	25.89	25.89
Januari 2012	25.47	25.47
Januari 2013	25.92	25.92
Januari 2014	NA	25.91
Januari 2015	26.06	26.06
Januari 2016	26.92	26.92
Januari 2017	25.72	25.72
Januari 2018	25.28	25.28

Nilai 25.91 tersebut didapatkan dengan menghitung nilai rata-rata pada periode Januari 2010, Januari 2011, Januari 2012, Januari 2013, Januari 2015, Januari 2016, Januari 2017, dan Januari 2018.

Data yang hilang pada variabel Angka Bebas Jentik (ABJ) ditangani dengan cara melakukan pengisian rata-rata pada seluruh

data sampel berdasarkan kecamatan. Kode Program 5.3 merupakan baris program yang digunakan untuk melakukan pengisian rata-rata pada data ABJ yang hilang.

```
missing_value_vars = [
    'Kepanjen', 'Donomulyo', 'Sumbermanjing Wetan',
    'Wajak', 'Lawang', 'Karangploso', 'Tumpang'
]
for var in missing_value_vars:
    abj[var].fillna(abj[var].mean(), inplace=True)
```

**Kode Program 5.3 Penanganan Missing Value Pada Data ABJ**

### 5.3.3 Normalisasi Data

Pada penelitian ini, metode *standard normalization* (z-score) digunakan sebagai metode normalisasi data. Hasil dari normalisasi menggunakan metode ini adalah setiap variabel memiliki nilai rata-rata 0 dan nilai standar deviasi 1. Metode *standard normalization* (z-score) dijalankan menggunakan *library sklearn*. Kode Program 5.4 merupakan baris program yang digunakan untuk melakukan normalisasi data.

```
from sklearn.preprocessing import StandardScaler

train = dbd.drop("outbreak", axis=1)
scaler = StandardScaler()
train = scaler.fit_transform(train)
```

**Kode Program 5.4 Potongan Kode Normalisasi Data**

### 5.3.4 Oversampling Data

*Oversampling* data dilakukan karena kelas target memiliki jumlah nilai yang tidak seimbang. Kelas target pada penelitian ini adalah Kejadian Luar Biasa (*outbreak*) yang berisi 80% data tidak terjadi Kejadian Luar Biasa. *Oversampling* data pada tahapan ini dilakukan menggunakan *library imblearn* dengan metode SMOTE. Kode Program 5.5 merupakan baris program yang digunakan untuk melakukan normalisasi data.

```
from imblearn.over_sampling import SMOTE

smote = SMOTE(sampling_strategy='minority')
train, target = smote.fit_sample(train, target)
```

### Kode Program 5.5 Potongan Kode Oversampling Data

#### 5.3.5 Pembagian Data Latih dan Uji

Pembagian data dilakukan dengan menggunakan metode *holdout* dengan pembagian 80% data latih dan 20% data uji serta melakukan pembagian data menggunakan metode *5-fold cross validation* untuk menguji apakah hasil prediksi mengalami *overfitting/underfitting*. Kode Program 5.6 merupakan baris program yang digunakan untuk melakukan pembagian data latih dan uji.

```

from sklearn.model_selection import
train_test_split, StratifiedKFold, cross_validate

# Metode Holdout (80% data latih, 20% data uji)
X_train, X_test, y_train, y_test =
train_test_split(train, target, test_size=0.2,
random_state = 123, stratify=target)

# Metode 5-cross-validation
kfold = StratifiedKFold(n_splits=5,
random_state=123)

```

### Kode Program 5.6 Potongan Kode Pembagian Data Latih dan Uji

Parameter *stratify* digunakan untuk menyeimbangkan jumlah variabel target pada data latih dan data uji.

## 5.4 Penyetelan Parameter

Parameter *C* dan *gamma* optimal pada SVM dicari menggunakan metode FPA yang mana dapat memaksimalkan tingkat akurasi model prediksi. Adapun langkah-langkah yang digunakan pada pencarian parameter tersebut adalah sebagai berikut.

### 5.4.1 Konfigurasi Parameter FPA

Penetapan nilai inisiasi parameter pada FPA dilakukan sesuai dengan skenario yang telah dijelaskan pada sub bab 4.3 yaitu dengan percobaan inisiasi populasi sebesar 10, 20, 50 dan 100 dan switch probability dengan rentan percobaan sebesar 0.1 –



1. Skenario tersebut digunakan untuk mendapatkan akurasi prediksi paling optimal.

### 5.4.2 Pencarian Parameter dengan FPA

Pencarian parameter yang digunakan sebagai masukan SVM adalah  $C$  (*cost*) dan  $\gamma$ . Pencarian parameter ini menggunakan *Flower Pollination Algorithm* (FPA).

```
def _phi(alpha, beta):
    return beta * np.tan(np.pi * alpha / 2.0)

def change_par(alpha, beta, mu, sigma, par_input,
par_output):
    if par_input == par_output:
        return mu
    elif (par_input == 0) and (par_output == 1):
        return mu - sigma * _phi(alpha, beta)
    elif (par_input == 1) and (par_output == 0):
        return mu + sigma * _phi(alpha, beta)

def random_levy(alpha, beta, mu=0.0, sigma=1.0,
shape=(), par=0):
    loc = change_par(alpha, beta, mu, sigma, par, 0)
    if alpha == 2:
        return np.random.standard_normal(shape) *
np.sqrt(2.0)

    radius = 1e-15
    if np.absolute(alpha - 1.0) < radius:
        alpha = 1.0 + radius

    r1 = np.random.random(shape)
    r2 = np.random.random(shape)
    pi = np.pi

    a = 1.0 - alpha
    b = r1 - 0.5
    c = a * b * pi
    e = _phi(alpha, beta)
    f = (-np.cos(c) + e * np.sin(c)) / (np.log(r2) *
np.cos(b * pi)) ** (a / alpha)
    g = np.tan(pi * b / 2.0)
    h = np.tan(c / 2.0)
    i = 1.0 - g ** 2.0
    j = f * (2.0 * (g - h) * (g * h + 1.0) - (h * i -
2.0 * g) * e * 2.0 * h)
    k = j / (i * (h ** 2.0 + 1.0)) + e * (f - 1.0)
    return loc + sigma * k

def get_levy_flight_array(dim=10):
```

```

return np.array([abs(random_levy(1.5, 0)) for _ in
range(dim)])

```

### Kode Program 5.7 Potongan Kode Random Levy 2 Dimension

Kode Program 5.7 diatas merupakan fungsi-fungsi dasar yang dibutuhkan oleh metode FPA yang mana menghasilkan 2 dimensi *random levy* (L).

```

class FlowerPollinationAlgorithm:
    def init (self, obj_fun, iterations_num,
pop_size, p=0.5, dim = 2, low=0, high=1):
        self.obj_fun = obj_fun
        self.iterations_num = iterations_num
        self.pop_size = pop_size
        self.population = None
        self.best = (0, None)
        self.p = p
        self.obj_vals = np.zeros(pop_size)
        self.low = low
        self.high = high
        self.dim = dim

    def initialize_population(self):
        self.population =
[ np.random.uniform(low=self.low,high=self.high,
size=self.dim) for _ in range(self.pop_size) ]

    def local_pollination(self, i):
        e = np.random.rand(self.dim)
        indexes = sample(range(0, self.pop_size), 2)
        sol1 = self.population[indexes[0]]
        sol2 = self.population[indexes[1]]
        tmp = e * (sol1 - sol2)
        new_solution = self.population[i] + tmp
        if new_solution[0] < 0:
            new_solution[0] = new_solution[0] * -1
        if new_solution[1] < 0:
            new_solution[1] = new_solution[1] * -1
        return new_solution

    def global_pollination(self, i):
        levy_vector = get_levy_flight_array(self.dim)
        tmp = levy_vector * (self.best[1] -
self.population[i])
        new_solution = self.population[i] + tmp
        if new_solution[0] < 0:
            new_solution[0] = new_solution[0] * -1
        if new_solution[1] < 0:
            new_solution[1] = new_solution[1] * -1
        return new_solution

```

```

def calculate_obj(self,x_train, y_train):
    for i, solution in enumerate(self.population):
        self.obj_vals[i] = self.obj_fun(solution,
x_train, y_train)

def find_best(self):
    if max(self.obj_vals) > self.best[0]:
        ind = np.argmax(self.obj_vals)
        self.best = (self.obj_vals[ind],
self.population[ind])

def find_solution(self, x_train, y_train):
    self.initialize_population()
    self.calculate_obj(x_train, y_train)
    self.avg = []
    self.bests = []
    self.flower = []
    for ii in range(self.iterations_num):
        self.find_best()
        for i in range(self.pop_size):
            if random() < self.p:
                new_flower =
self.global_pollination(i)
            else:
                new_flower =
self.local_pollination(i)
            new_obj_val = self.obj_fun(new_flower,
x_train, y_train)
            if new_obj_val > self.obj_vals[i]:
                self.population[i] = new_flower
                self.obj_vals[i] = new_obj_val

        self.avg.append(np.mean(self.obj_vals))
        self.bests.append(max(self.obj_vals))
        self.find_best()

```

#### Kode Program 5.8 Potongan Kode Flower Pollination Algorithm

Kode Program 5.8 diatas merupakan fungsi-fungsi yang terdapat pada kelas FPA yang digunakan untuk menjalankan metode FPA. `def __init__(self, obj_fun, iterations_num, pop_size, p=0.5, dim = 2, low=0, high=1)` merupakan fungsi untuk menginisiasi metode FPA dengan parameter-parameter yang dibutuhkannya. `def initialize_population(self)` merupakan fungsi untuk menginisiasi isi dari populasi. `def local_pollination(self, i)` dan `def global_pollination(self, i)` berturut-turut merupakan fungsi untuk melakukan penyerbukan secara lokal dan penyerbukan secara global dengan formula mengacu pada sub bab 2.2.6. `def`

`calculate_obj(self,x_train, y_train)` merupakan fungsi untuk menghitung skor akurasi dari parameter yang didapatkan. `def find_best(self dan def find_solution(self, x_train, y_train)` merupakan fungsi yang digunakan untuk mencari parameter optimalnya.

```
def fitness_acc(f, X_train, y_train):
    clf = SVC(gamma=f[0], C=f[1])
    clf.fit(X_train, y_train)
    y_pred = clf.predict(X_test)
    acc = accuracy_score(y_test, y_pred)
    precision = precision_score(y_test, y_pred)
    recall = recall_score(y_test, y_pred)
    return (acc + precision + recall)/3
```

### Kode Program 5.9 Potongan Kode Fungsi Objektif

Kode Program 5.9 diatas merupakan fungsi untuk menghitung nilai akurasi sehingga metode FPA dapat membandingkan nilai akurasi dari parameter-parameter yang dihasilkan.

1. low, high = 0, 10
2. pop\_init = [10, 20, 50, 100]
3. switch\_prob = [0.2, 0.4, 0.6, 0.8, 1]

### Kode Program 5.10 Inisiasi Variabel Skenario Pencarian Parameter

Kode Program 5.10 diatas merupakan skenario inisiasi parameter pada FPA. Pada baris 1 berisi batasan untuk nilai inisiasi isi dari populasi awal, baris 2 berisi beberapa skenario inisiasi populasi, dan pada baris 3 berisi beberapa skenario inisiasi *switch probability*.

```
C_value = []
gamma_value = []
population = []
switch_p = []
accuracy = []
precision = []
recall = []
rataaan = []

for pop in pop_init:
    for p in switch_prob:
        fpa = FlowerPollinationAlgorithm(fitness_acc,
10, pop, p=p, low=low, high=high)
        fpa.find_solution(X_train,y_train)
        best_fitness = fpa.best[0]
```

```

solution = fpa.best[1]
nilai_c = solution[1]
nilai_gamma = solution[0]

population.append(pop)
switch_p.append(p)
C_value.append(nilai_c)
gamma_value.append(nilai_gamma)
rataaan.append(best_fitness)

clf1 = SVC(C=nilai_c, gamma=nilai_gamma)
clf1.fit(X_train, y_train)
y_pred = clf1.predict(X_test)
accu = accuracy_score(y_test, y_pred)
pre = precision_score(y_test, y_pred)
rec = recall_score(y_test, y_pred)
precision.append(pre)
recall.append(rec)
accuracy.append(accu)

df1 = pd.DataFrame(list(zip(population, switch_p,
C_value, gamma_value, accuracy, precision, recall,
rataaan)),
                    columns=['Population', 'Switch_Prob',
' C', 'Gamma', 'Accuracy', 'Precision', 'Recall', 'Nilai
Fungsi Terbaik'])

df1.sort_values(by='Nilai Fungsi Terbaik',
ascending=False)

```

#### Kode Program 5.11 Kode Program Skenario Pencarian Parameter

Kode Program 5.11 merupakan baris kode yang digunakan untuk menjalankan skenario-skenario percobaan yang dapat dilihat pada sub bab 4.3.

### 5.5 Prediksi Support Vector Machine

Pembuatan model SVM dilakukan menggunakan bahasa python dengan beberapa *library* antara lain Pandas dan Numpy untuk pengolahan data, sklearn digunakan untuk normalisasi, mengubah variabel kategorik menjadi numerik, membagi dataset menjadi data latih dan uji, serta membuat pemodelan SVM dan evaluasi performa, dan imblearn digunakan untuk menjalankan skenario *oversampling*. Semua *library* yang digunakan dipanggil seperti yang terdapat pada Kode Program 5.12.

```

1. import pandas as pd
2. import numpy as np
3. from imblearn.over_sampling import SMOTE
4. from sklearn.model_selection import train_test_split
   from sklearn.preprocessing import StandardScaler, OneHotEncoder
   from sklearn.svm import LinearSVC
   from sklearn.svm import SVC
   from sklearn.metrics import confusion_matrix, accuracy_score,
   precision_score, recall_score, roc_auc_score

```

**Kode Program 5.12 Potongan Kode Import Library Yang Dibutuhkan**

### 5.5.1 Load Data

Sebelum memasuki proses prediksi menggunakan model SVM, maka perlu untuk memasukkan data yang akan digunakan sebagai data latih dan uji. *Library* yang digunakan dalam tahap ini adalah Pandas. Kode Program 5.13 merupakan baris program yang digunakan untuk memasukan data kedalam objek *DataFrame*.

```

dbd = pd.read_csv('D:/Dataset/tugas_akhir (dbd
kabupaten malang)/dbd_lag1.csv')

```

**Kode Program 5.13 Load Data**

### 5.5.2 Prediksi SVM

Prediksi SVM dilakukan dengan menggunakan *library sklearn* dimana default parameter dari SVM terdapat pada Tabel 5.5.

**Tabel 5.5 Default Parameter SVM**

<b>Kernel</b>	Radial Basis Function (RBF)
<b>C</b>	1.0
<b>gamma</b>	$1 / (n\_features * X.var())$

Prediksi SVM dilakukan dengan menggunakan percobaan parameter *default* dan parameter hasil pencarian menggunakan algoritma FPA. Kode Program 5.14 merupakan baris program untuk membuat model SVM dengan parameter standar yang tertera pada Tabel 5.5.

```

clf = SVC()
clf.fit(X_train, y_train)
y_pred = clf.predict(X_test)

```

### Kode Program 5.14 Prediksi SVM dengan Default Parameter

Kode Program 5.15 merupakan baris program untuk membuat model SVM dengan parameter yang didapatkan pada proses pencarian parameter menggunakan *Flower Pollination Algorithm* (FPA).

```
clf = SVC(C=fpa_best_C, gamma=fba_best_gamma)
clf.fit(X_train, y_train)
y_pred = clf.predict(X_test)
```

### Kode Program 5.15 Prediksi SVM dengan Parameter Hasil FPA

Model yang dihasilkan dari kedua percobaan diatas akan dilakukan *cross validation* dengan jumlah lipatan sebanyak 5 lipatan. Kode Program 5.16 merupakan baris program untuk melakukan *5-fold cross validation* berdasarkan pembagian data yang terdapat pada sub bab 5.3.5.

```
cross_validate(estimator=clf, X=train, y=target,
cv=kfold, scoring=['accuracy', 'precision', 'recall'])
```

### Kode Program 5.16 Prediksi SVM pada 5-fold cross validation

## 5.6 Analisa Kinerja

Hasil prediksi model dengan menggunakan percobaan parameter default dan parameter hasil pencarian algoritma FPA akan dianalisis dengan menghitung nilai akurasi, presisi, dan *recall*.

```
print(accuracy_score(y_test, y_pred),
precision_score(y_test, y_pred), recall_score(y_test,
y_pred))

tn, fp, fn, tp = confusion_matrix(y_true=y_test,
y_pred=y_pred).ravel()
print('TN:', tn, '\n FP:', fp, '\n FN:', fn, '\n TP:',
tp)
```

### Kode Program 5.17 Evaluasi Kinerja Model Prediksi

Kode Program 5.17 digunakan untuk menghitung nilai akurasi, presisi, dan *recall* dari model yang dimiliki.

*Halaman ini sengaja dikosongkan*



## BAB VI HASIL DAN PEMBAHASAN

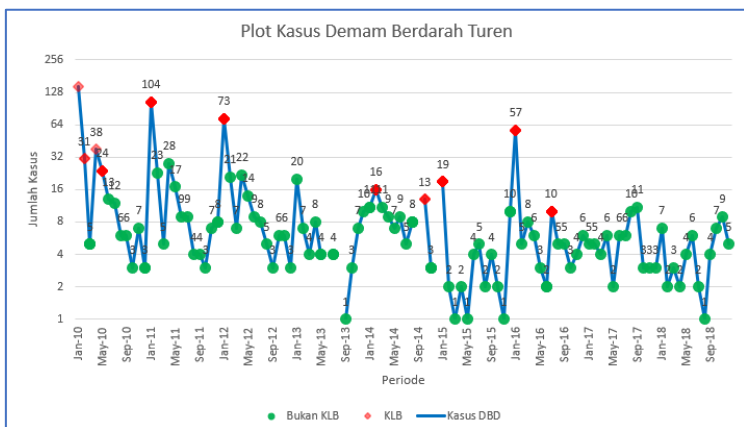
Pada bab ini akan berisi tentang hasil dari penelitian beserta pembahasan dari implementasi yang sudah dilakukan sesuai dengan metode penelitian. Hasil yang akan dibahas di bab ini adalah mengenai praproses data, penentuan parameter SVM, performa prediksi menggunakan nilai rata-rata akurasi, sensitivitas dan presisi.

### 6.1 Hasil Eksplorasi Data

Eksplorasi data dilakukan untuk mendapatkan wawasan mengenai data yang dimiliki. Tahap ini terdiri atas ringkasan statistik, distribusi data dan hubungan antar variabel.

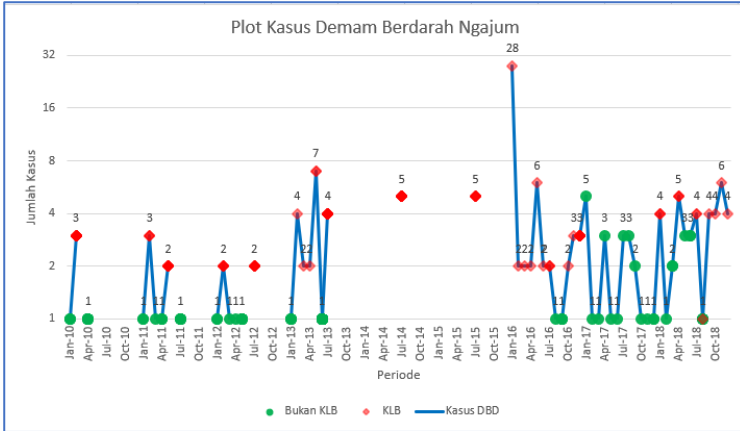
#### 6.1.1 Visualisasi Jumlah Kasus Demam Berdarah

Data kasus demam berdarah yang dimiliki terdiri atas beberapa kecamatan dengan jumlah kasus demam berdarah yang berbeda-beda. Visualisasi dilakukan menggunakan skala logaritma dengan dasar 2.



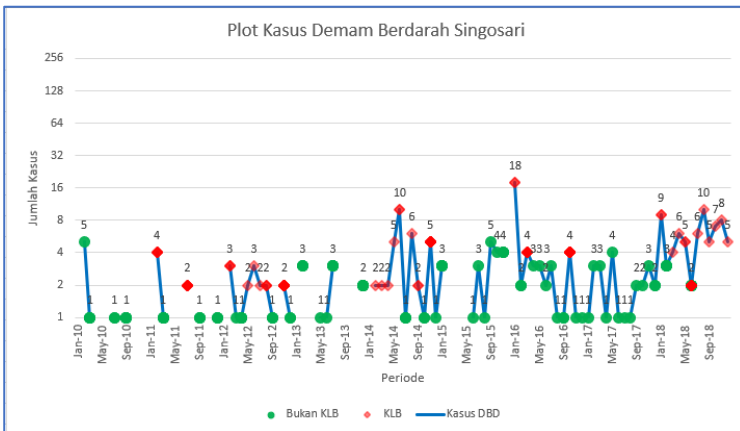
Gambar 6.1 Plot Kasus Demam Berdarah di Kecamatan Turen

Gambar 6.1 menunjukkan titik-titik periode yang merupakan Kejadian Luar Biasa (KLB) di Kecamatan Turen.



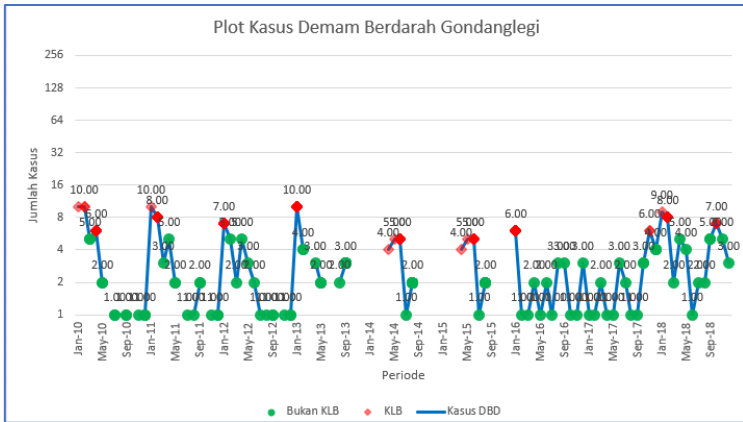
**Gambar 6.2** Plot Kasus Demam Berdarah di Kecamatan Ngajum

Gambar 6.2 menunjukkan titik-titik periode yang merupakan Kejadian Luar Biasa (KLB) di Kecamatan Ngajum. Terdapat banyak periode yang mendapatkan status KLB yang diakibatkan oleh jumlah kasus DBD yang normal di daerah tersebut berada pada kisaran 0 dan 1.



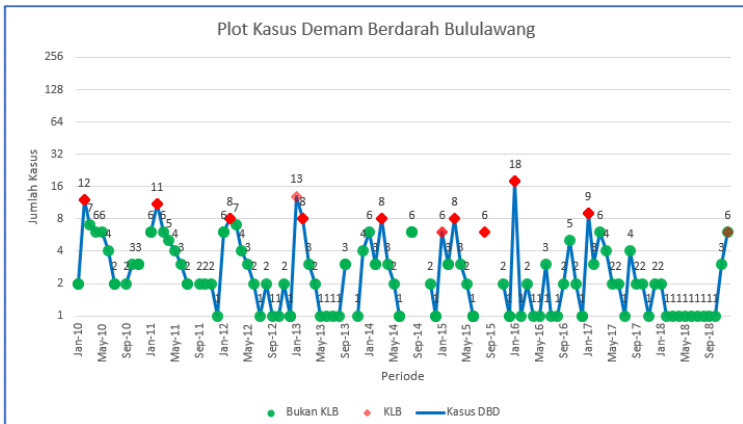
**Gambar 6.3** Plot Kasus Demam Berdarah di Kecamatan Singosari

Gambar 6.3 menunjukkan titik-titik periode yang merupakan Kejadian Luar Biasa (KLB) di Kecamatan Singosari.



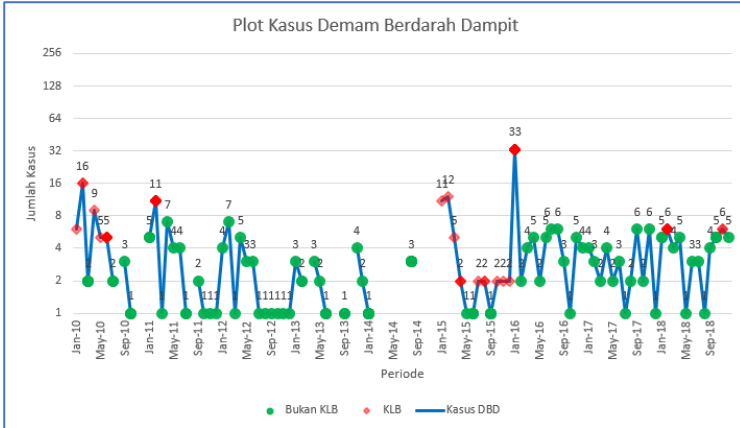
**Gambar 6.4** Plot Kasus Demam Berdarah di Kecamatan Gondanglegi

Gambar 6.4 menunjukkan titik-titik periode yang merupakan Kejadian Luar Biasa (KLB) di Kecamatan Gondanglegi.



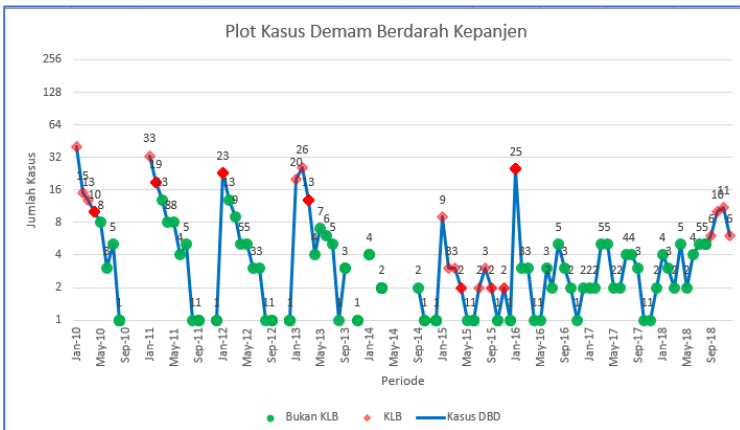
**Gambar 6.5** Plot Kasus Demam Berdarah di Kecamatan Bululawang

Gambar 6.5 menunjukkan titik-titik periode yang merupakan Kejadian Luar Biasa (KLB) di Kecamatan Bululawang.



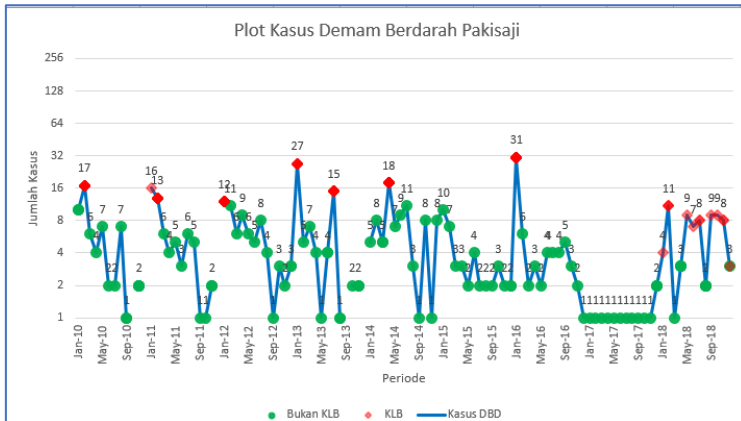
**Gambar 6.6 Plot Kasus Demam Berdarah di Kecamatan Dampit**

Gambar 6.6 menunjukkan titik-titik periode yang merupakan Kejadian Luar Biasa (KLB) di Kecamatan Dampit.



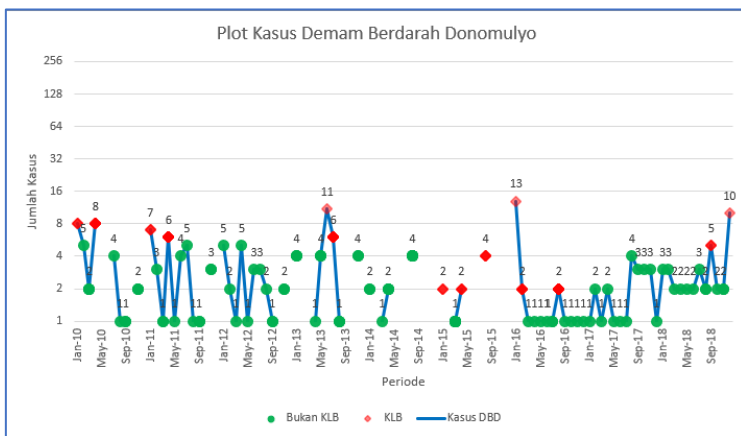
**Gambar 6.7 Plot Kasus Demam Berdarah di Kecamatan Kapanjen**

Gambar 6.7 menunjukkan titik-titik periode yang merupakan Kejadian Luar Biasa (KLB) di Kecamatan Kapanjen.



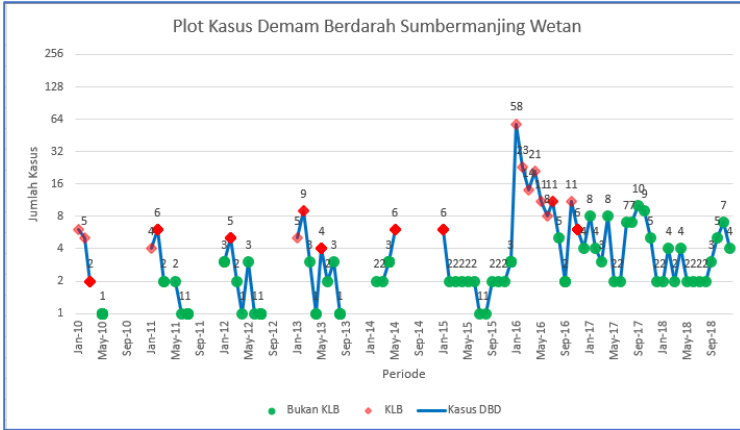
**Gambar 6.8 Plot Kasus Demam Berdarah di Kecamatan Pakisaji**

Gambar 6.8 menunjukkan titik-titik periode yang merupakan Kejadian Luar Biasa (KLB) di Kecamatan Pakisaji.



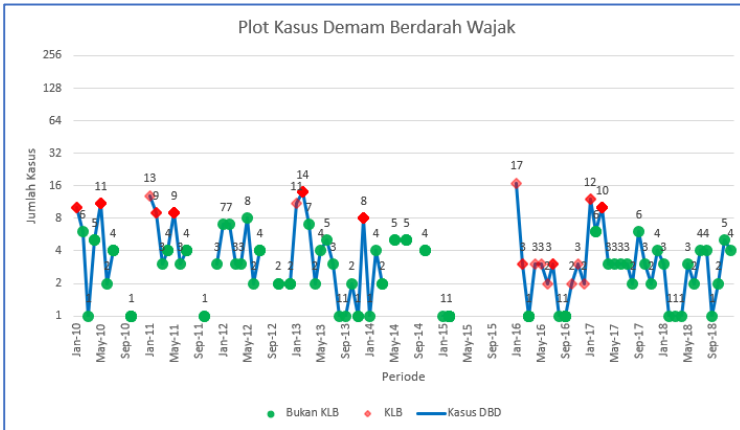
**Gambar 6.9 Plot Kasus Demam Berdarah di Kecamatan Donomulyo**

Gambar 6.9 menunjukkan titik-titik periode yang merupakan Kejadian Luar Biasa (KLB) di Kecamatan Donomulyo.



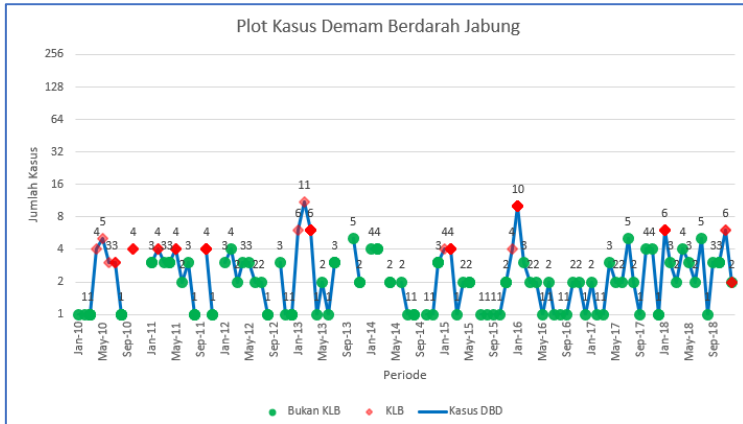
**Gambar 6.10** Plot Kasus Demam Berdarah di Kecamatan Sumbermanjing Wetan

Gambar 6.10 menunjukkan titik-titik periode yang merupakan Kejadian Luar Biasa (KLB) di Kecamatan Sumbermanjing Wetang.



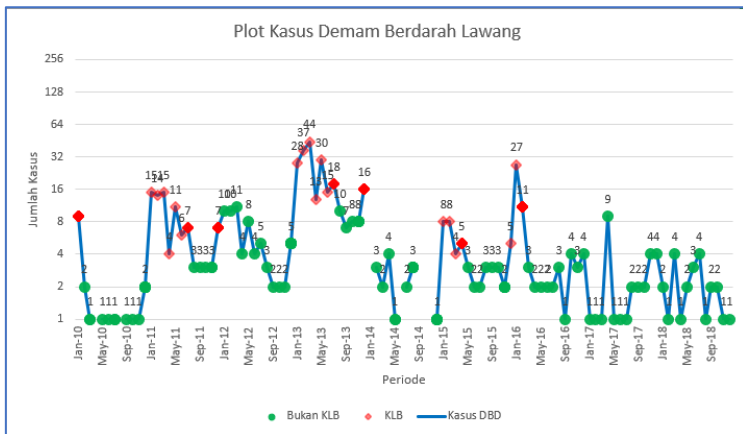
**Gambar 6.11** Plot Kasus Demam Berdarah di Kecamatan Wajak

Gambar 6.11 menunjukkan titik-titik periode yang merupakan Kejadian Luar Biasa (KLB) di Kecamatan Wajak.



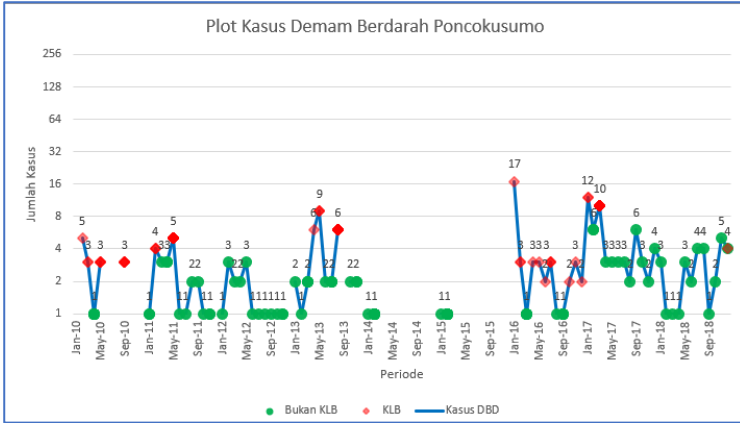
**Gambar 6.12** Plot Kasus Demam Berdarah di Kecamatan Jabung

Gambar 6.12 menunjukkan titik-titik periode yang merupakan Kejadian Luar Biasa (KLB) di Kecamatan Jabung.



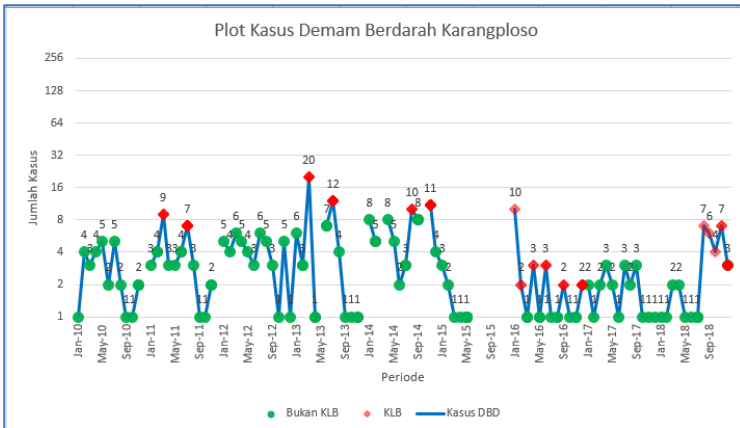
**Gambar 6.13** Plot Kasus Demam Berdarah di Kecamatan Lawang

Gambar 6.13 menunjukkan titik-titik periode yang merupakan Kejadian Luar Biasa (KLB) di Kecamatan Lawang.



**Gambar 6.14 Plot Kasus Demam Berdarah di Kecamatan Poncokusumo**

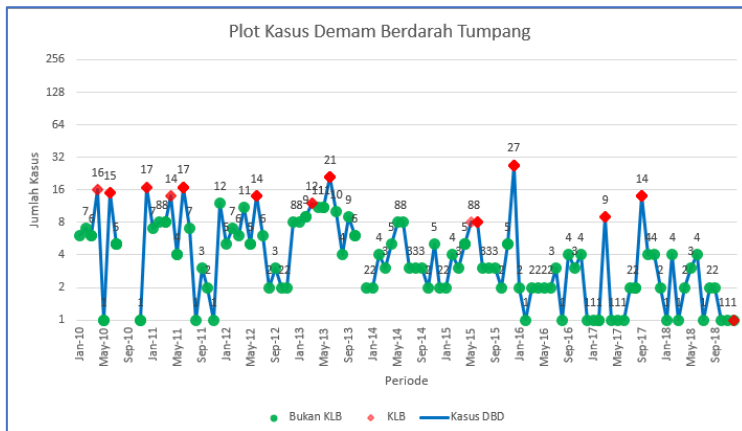
Gambar 6.14 menunjukkan titik-titik periode yang merupakan Kejadian Luar Biasa (KLB) di Kecamatan Poncokusumo.



**Gambar 6.15 Plot Kasus Demam Berdarah di Kecamatan Karangploso**

Gambar 6.15 menunjukkan titik-titik periode yang merupakan Kejadian Luar Biasa (KLB) di Kecamatan Karangploso.





**Gambar 6.16** Plot Kasus Demam Berdarah di Kecamatan Tumpang

Gambar 6.16 menunjukkan titik-titik periode yang merupakan Kejadian Luar Biasa (KLB) di Kecamatan Tumpang.

### 6.1.2 Ringkasan Statistik

Data yang dimiliki terdiri atas variabel independen dan variabel dependen. Semua variabel memiliki jumlah baris yang sama yaitu 1712 baris. Tabel 6.1 merupakan ringkasan statistik dari variabel numerik yang dimiliki.

**Tabel 6.1** Ringkasan Statistik Variabel Numerik

Variabel	STD	Mean	Min	Q1	Q2	Q3	Max
Suhu Udara	1.3	24.9	21.66	23.91	25	26	27.8
Kelembaban Udara	4.72	80.58	62.03	77.8	81.7	84.25	88.84
Curah Hujan	5.81	7.17	0	1.65	6.74	11	26.54
Kecepatan Angin	0.7	1.41	0.1	0.84	1.39	1.94	3.42
Angka Bebas Jentik	8.95	86.31	0	84.33	87.7	91.52	100
Kejadian Demam Berdarah	6.7	3.72	0	1	2	4	146

### 6.1.3 Distribusi Kelas Target

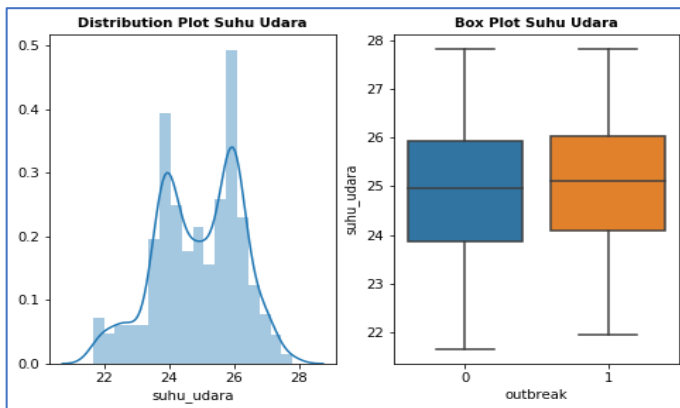
Variabel dependen atau target yang dimiliki tergolong kedalam distribusi yang tidak berimbang dimana data terdiri atas 80% bukan Kejadian Luar Biasa seperti yang tertera pada Gambar 6.17. Oleh karena kelas target yang tidak seimbang ini, model prediksi akan sangat kesulitan untuk memprediksi terjadinya KLB sehingga dibutuhkan penanganan. Pada penelitian ini, metode *oversampling* akan digunakan untuk menangani permasalahan tersebut.



Gambar 6.17 Distribusi Kelas Target

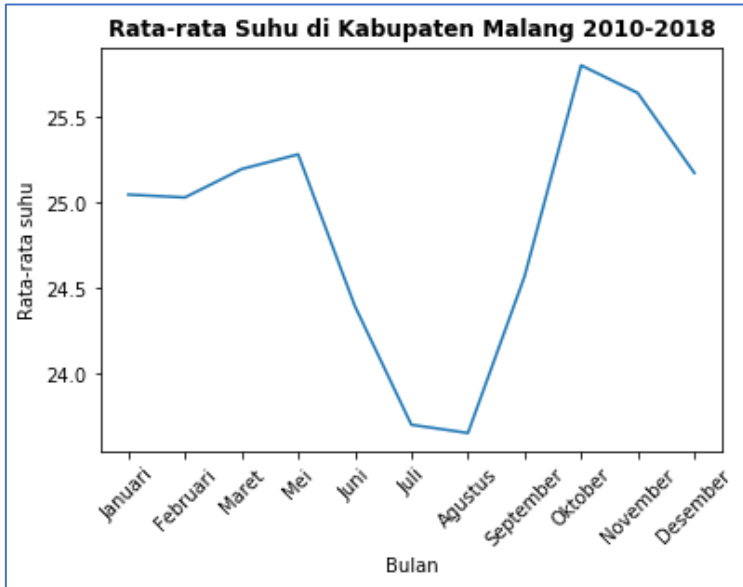
### 6.1.4 Distribusi Tiap Variabel

Distribusi setiap variabel sangat berguna untuk mengetahui bagaimana karakteristik data yang dimiliki.



Gambar 6.18 Distribusi Variabel Suhu

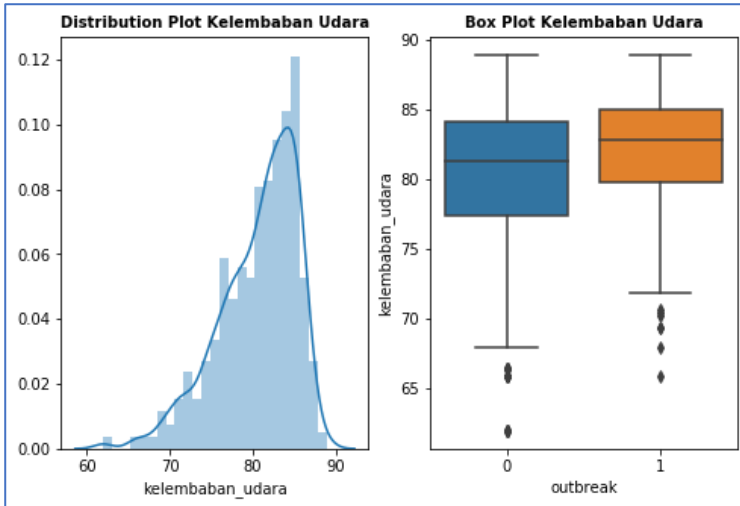
Berdasarkan Gambar 6.18, variabel suhu terlihat memiliki distribusi yang normal. Distribusi suhu pada setiap kelas target juga tidak terlihat perbedaan yang signifikan.



**Gambar 6.19 Rata-rata Suhu di Malang pada Tahun 2010-2018**

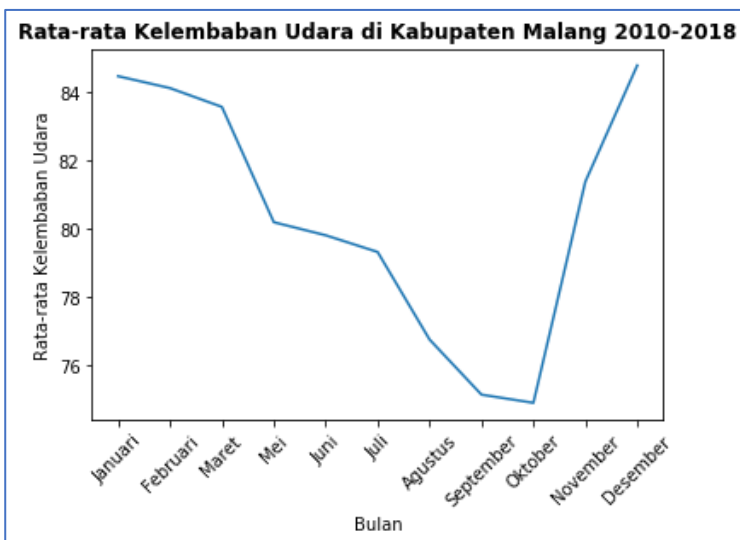
Berdasarkan Gambar 6.19, rata-rata suhu di Malang tidak mengalami perubahan yang signifikan setiap bulannya kecuali yang terjadi pada bulan Juli dan Agustus. Pada bulan Juli dan Agustus terlihat bahwa suhu mengalami penurunan yang signifikan dibandingkan bulan-bulan sebelumnya.

Berdasarkan Gambar 6.20, variabel kelembaban udara memiliki distribusi yang sedikit *left-skewed* (*negative skewness*). Terlihat bahwa kelembaban udara yang berada dibawah 70 tidak mengalami Kejadian Luar Biasa. Hal tersebut sangat sesuai dengan hipotesis awal dimana nyamuk *Aedes Aegypti* dapat berkembang biak dengan mudah pada lingkungan dengan kelembaban udara yang tinggi.



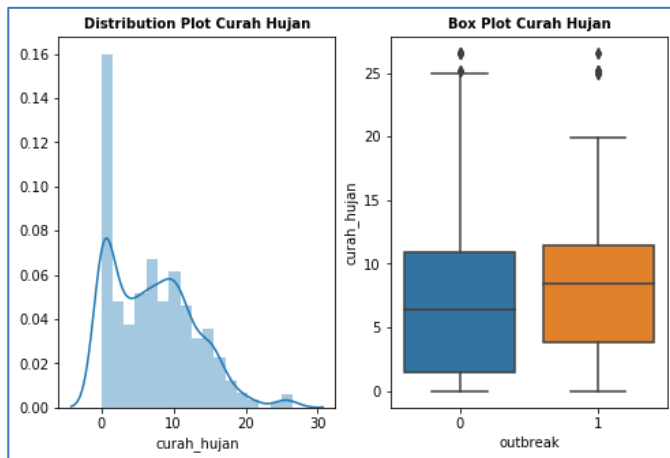
**Gambar 6.20 Distribusi Kelembaban Udara**

Berdasarkan Gambar 6.21, perubahan kelembaban udara setiap bulannya terjadi perubahan signifikan terutama pada saat pergantian musim.



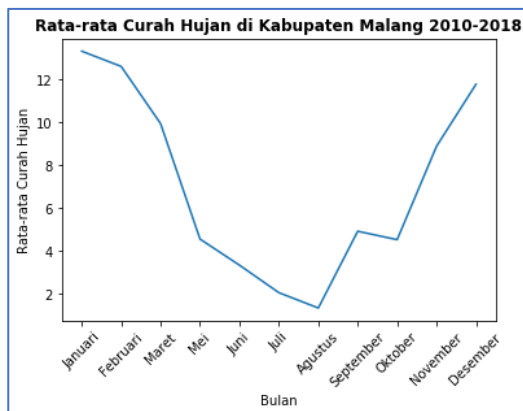
**Gambar 6.21 Rata-Rata Kelembaban Udara di Kabupaten Malang pada Tahun 2010-2018**

Perubahan signifikan terjadi ketika memasuki bulan Oktober-November yang mulai mengalami peningkatan kelembaban yang dipicu oleh musim hujan.



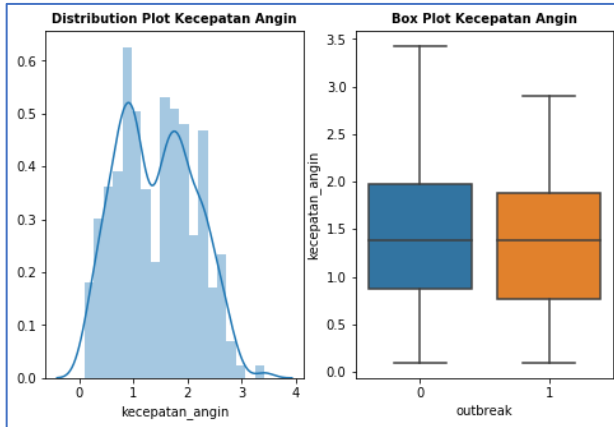
**Gambar 6.22 Distribusi Variabel Curah Hujan**

Berdasarkan Gambar 6.22, variabel curah hujan memiliki distribusi *right-skewed* (*positif skewed*). Hal ini terjadi karena data curah hujan memiliki banyak nilai 0 (tidak mengalami hujan pada bulan tersebut).



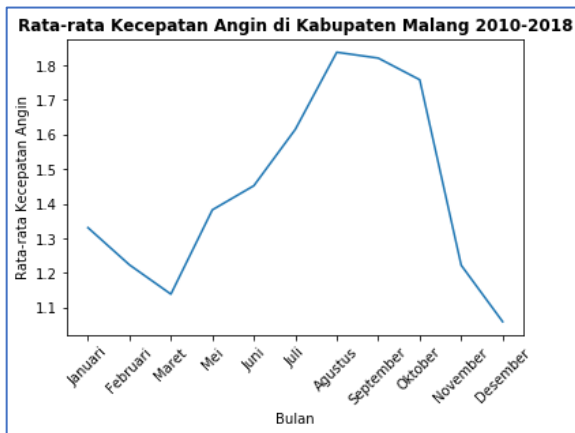
**Gambar 6.23 Rata-rata Curah Hujan pada Tahun 2010-2018**

Berdasarkan Gambar 6.23, curah hujan mengalami peningkatan pada bulan September – Februari. Hal ini sangat berkorelasi dengan peningkatan yang terjadi pada variabel kelembaban udara yang dapat dilihat pada Gambar 6.21.



**Gambar 6.24 Distribusi Variabel Kecepatan Angin**

Berdasarkan Gambar 6.24, variabel kecepatan angin memiliki distribusi yang normal dan memiliki distribusi antar kelas target yang berimbang.



**Gambar 6.25 Rata-rata Kecepatan Angin pada Tahun 2010-2018**

Berdasarkan Gambar 6.25, kecepatan angin mengalami peningkatan pada bulan April hingga Agustus dan mengalami penurunan dari bulan September hingga Maret.

Kesimpulan sementara yang dapat diambil dari tahap ini adalah setiap variabel numerik harus dilakukan normalisasi agar menjadi distribusi yang normal. Semua variabel diatas juga terlihat sangat berkorelasi terhadap Kejadian Demam Berdarah di Kabupaten Malang pada tahun 2010-2018.

## 6.2 Hasil Mengubah Data Harian Variabel Iklim

Data iklim yang tersedia pada situs pusat data Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) masih memiliki rentang waktu harian. Untuk mendapatkan data iklim dengan periode bulanan, maka data harian tersebut dirata-ratakan. Tabel 6.2 berikut menunjukkan hasil menghitung nilai rata-rata variabel iklim di Stasiun Geofisika Karang Kates pada tahun 2010.

Tabel 6.2 Hasil Mengubah Data Harian Variabel Iklim

Periode	Suhu Udara	Kelembaban Udara	Curah Hujan	Kecepatan Angin
Januari	26.02	84.74	10.52	0.84
Februari	26.32	84.64	12.04	0.86
Maret	26.39	85	8.65	0.55
April	26.29	86.17	10.37	0.3
Mei	26.57	85.61	7.35	0.58
Juni	25.62	84.1	9.23	1.59
Juli	25.35	83.52	2.16	1.77
Agustus	25.35	81.71	3.94	1.68
September	25.73	83.93	9.4	1.37
Oktober	26.19	82.03	12.13	1.71
November	26	84.53	17.53	1.23
Desember	25.83	83.39	8.45	1.77

Setelah memastikan semua variabel memiliki periode yang sama, maka tahap selanjutnya adalah melakukan penanganan terhadap data yang hilang.

### 6.3 Hasil Pengisian Data Yang Hilang

Praproses data berupa pengisian data yang hilang dilakukan pada variabel suhu rata-rata perbulan, rata-rata persentase kelembaban perbulan, rata-rata curah hujan perbulan, rata-rata kecepatan angin perbulan dan angka bebas jentik.

**Tabel 6.3 Hasil Pengisian Data Yang Hilang Pada Data Iklim**

<b>Variabel</b>	<b>Periode</b>	<b>Nilai Pengisian</b>
<b>Suhu Udara</b>	Januari	25.93
	Februari	25.94
	Maret	26.07
	April	26.26
	Mei	26.11
	Juni	25.16
	Agustus	24.55
	Desember	26.13
<b>Kelembaban Udara</b>	Januari	85.28
	Februari	85,22
	Maret	84.95
	April	84.25
	Mei	81.93
	Juni	81.63
	Agustus	78.78
	Desember	85.14
<b>Curah Hujan</b>	Februari	14.3
	Maret	11.45
	April	11.03
	Mei	5.15
	Juni	4.84
	Juli	2.79
	Agustus	1.56
	September	7.33
	Oktober	5.32
	November	8.55
<b>Kecepatan Angin</b>	Februari	0.81
	Maret	0.74



	April	0.68
	Mei	0.9
	Juni	1.1

Tabel 6.3 menunjukkan pengisian data yang hilang pada periode-periode yang disebutkan. Nilai-nilai pada perubahan data yang hilang didapatkan dari rataan variabel pada periode yang sama pada tahun yang tidak mengalami kehilangan data.

Angka Bebas Jentik (ABJ) ditangani dengan cara melakukan pengisian rataan pada seluruh data sampel berdasarkan kecamatan.

**Tabel 6.4 Hasil Pengisian Data yang Hilang pada Variabel ABJ**

Kecamatan	Nilai Pengisian
Kepanjen	82.17
Donomulyo	81.89
Sumbermanjing Wetan	84.59
Wajak	91.52
Lawang	87.44
Karangploso	79.41
Tumpang	90.12

Tabel 6.4 menunjukkan pengisian data ABJ yang hilang pada masing-masing kecamatan yang tertera.

#### 6.4 Hasil Normalisasi Data

Implementasi metode normalisasi dilakukan dengan menggunakan *library* scikit-learn dengan metode normalisasi standar deviasi. Hasil dari normalisasi menggunakan metode ini adalah setiap variabel memiliki nilai rataan 0 dan nilai standar deviasi 1.

**Tabel 6.5 Hasil Normalisasi Data**

Variabel	Data Awal	Hasil Normalisasi
Suhu Udara	26.02	0.8620
	26.32	1.0930
	26.39	1.1469

	26.29	1.0699
	26.57	1.2855
<b>Kelembaban Udara</b>	84.74	0.8809
	84.64	0.8597
	85	0.9360
	86.17	1.1838
	85.61	1.0652
<b>Curah Hujan</b>	10.52	0.5748
	12.04	0.8362
	8.65	0.2531
	10.37	0.5490
	7.35	0.0295
<b>Kecepatan Angin</b>	0.84	-0.8237
	0.86	-0.7952
	0.55	-1.2359
	0.30	-1.5912
	0.58	-1.1932
<b>Kepadatan Penduduk</b>	1735	0.9828
	797	-0.6782
	1379	0.3531
	1736	0.9846
	1737	0.9864
<b>Angka Bebas Jentik</b>	86.96	0.0715
	86.72	0.0447
	86.50	0.0201
	86.28	-0.0044
	86.08	-0.0267
<b>Kejadian Demam Berdarah</b>	146	21.245
	31	4.0732
	5	0.1908
	38	5.1185
	24	3.0279

Tabel 6.5 merupakan 5 titik data yang diambil sebelum dilakukan normalisasi dan hasil setelah dilakukan normalisasi. Untuk memastikan bahwa setiap variabel telah memiliki nilai rata-rata 0 dan nilai standar deviasi 1, maka dilakukan

pengecekan. Tabel 6.6 merupakan ringkasan validasi hasil normalisasi yang telah dilakukan untuk memastikan setiap variabel data terpusat pada rata-rata 0 dan standar deviasi 1.

**Tabel 6.6 Validasi Hasil Normalisasi**

Variabel	Sebelum Normalisasi		Setelah Normalisasi	
	Mean	STD	Mean	STD
Suhu Udara	24.9	1.299	0	1
Kelembaban Udara	80.58	4.722	0	1
Curah Hujan	7.17	5.815	0	1
Kecepatan Angin	1.42	0.703	0	1
Kepadatan Penduduk	1180	564.84	0	1
Angka Bebas Jentik	86.31	8.958	0	1
Kejadian Demam Berdarah	3.72	6.698	0	1

## 6.5 Hasil Implementasi Oversampling

Implementasi metode *oversampling* dilakukan dengan menggunakan *library* imblearn. Metode ini memiliki konsep untuk membuat sampel dari kelas minoritas sehingga berjumlah sama dengan kelas mayoritas.

**Tabel 6.7 Hasil Implementasi Oversampling**

	Jumlah Awal	Jumlah Setelah Oversampling
Kelas KLB	295	1241
Kelas Bukan KLB	1241	1241
Total	1536	2482

Tabel 6.7 merupakan jumlah kelas target sebelum melakukan implementasi *oversampling* dan setelah melakukan implementasi *oversampling*.

## 6.6 Hasil Pembagian Data Latih dan Uji

Pembagian data dilakukan menggunakan metode *holdout* dengan komposisi 80% data latih dan 20% data validasi. Pembagian data juga dilakukan dengan menggunakan *5-fold cross validation* untuk melakukan validasi model prediksi tidak terjadi *overfitting* maupun *underfitting*.

Tabel 6.8 Ukuran Data Latih, Data Validasi dan Data Uji

Pembagian Data	Jumlah		Total
	KLB	Bukan KLB	
Data Latih	993	992	1985
Data Validasi	248	249	497
Data Uji (Dataset tahun 2018)	43	133	176

Tabel 6.8 merupakan persebaran jumlah data pada data latih dan data uji setelah dilakukan pembagian data.

Pada pembagian data menggunakan *5-fold cross validation*, jumlah data latih sebanyak 1985 dan jumlah data uji sebanyak 497. Data latih dan data uji berisikan data berbeda pada setiap lipatannya dengan skenario seperti Tabel 6.9.

Tabel 6.9 *5-fold cross validation*

Data										
CV1										
CV2										
CV3										
CV4										
CV5										
Kelas	NON KLB					KLB				

Note:  = Data uji,  = Data latih

## 6.7 Hasil Penyetelan Parameter

Pencarian parameter SVM (C dan gamma) menggunakan metode FPA dilakukan dengan beberapa iterasi dengan inisiasi parameter FPA yang berbeda. Berikut adalah semua inisiasi parameter FPA yang dilakukan.

Tabel 6.10 Skenario Inisiasi Parameter FPA

Jumlah Iterasi	10, 20, 50, 100
Jumlah Populasi (pop)	10, 20, 50, 100
Switch Probability (p)	0.2, 0.4, 0.6, 0.8, 1

### 6.7.1 Percobaan dengan 10 iterasi

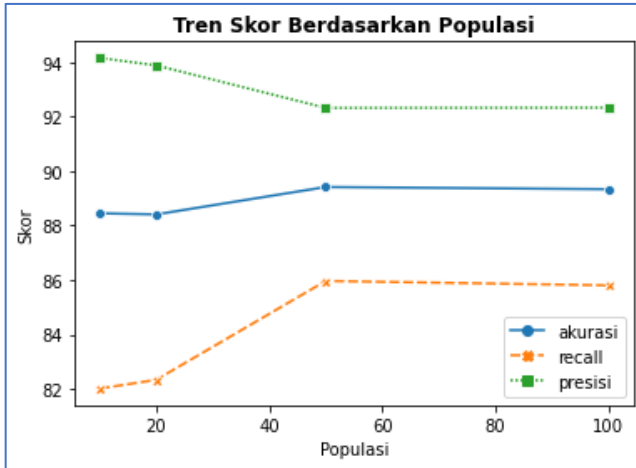
Pencarian parameter SVM menggunakan *Flower Pollination Algorithm* dengan 10 iterasi mendapatkan hasil yang tertera pada Tabel 6.11.

Tabel 6.11 Pencarian Parameter dengan 10 Iterasi

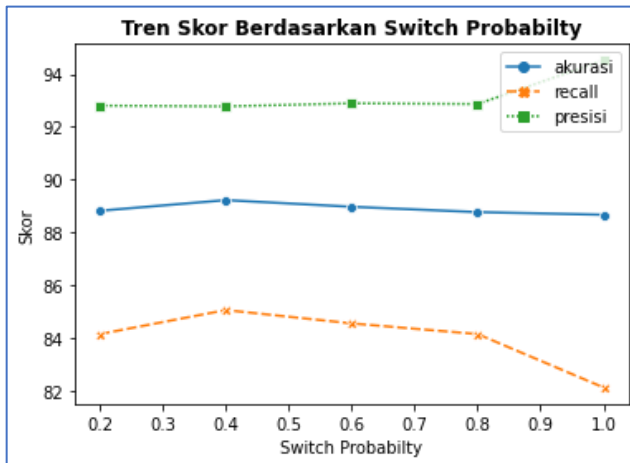
No	Pop	p	C	gamma	best
1.	10	0.2	1,852280	0,011119	88,058
2.	10	0.4	9,559361	0,006599	88,688
3.	10	0.6	6,177411	1,303472	86,891
4.	10	0.8	6,583266	0,028541	88,894
5.	10	1	5,378126	0,028644	88,904
6.	20	0.2	7,082249	0,026645	88,505
<b>7.</b>	<b>20</b>	<b>0.4</b>	<b>134,998469</b>	<b>0,014038</b>	<b>89,832</b>
8.	20	0.6	9,444606	0,017962	88,930
9.	20	0.8	7,946561	1,301724	86,891
10.	20	1	4,116458	1,289520	86,891
11.	50	0.2	7,654157	0,020653	89,122
12.	50	0.4	6,667471	0,026881	88,910
13.	50	0.6	140,663871	0,013757	89,618
14.	50	0.8	157,868123	0,010636	89,618
15.	50	1	6,129167	0,017625	88,904
16.	100	0.2	15,013524	0,016515	88,730
17.	100	0.4	5,235597	0,031316	88,692
18.	100	0.6	102,698249	0,016270	89,832
19.	100	0.8	111,965536	0,014354	89,424
20.	100	1	12,837860	0,005408	89,108

Pada percobaan dengan 10 iterasi diatas, dengan menggunakan 20 populasi dan 0.4 switch probability didapatkan skor tertinggi dengan nilai 89,832.

Berdasarkan Gambar 6.26, skor *recall* mengalami peningkatan dari populasi 10 hingga 50 sedangkan pada populasi 100 mengalami penurunan. Skor akurasi mengalami peningkatan seiring dengan penambahan jumlah populasi.



Gambar 6.26 Tren Skor Berdasarkan Populasi Pada Percobaan dengan 10 Iterasi



Gambar 6.27 Tren Skor Berdasarkan Switch Probability Pada Percobaan dengan 10 Iterasi

Berdasarkan Gambar 6.27, *switch probability* memberikan dampak terhadap skor *recall* namun dampak yang terjadi pada akurasi tidak besar.

### 6.7.2 Percobaan dengan 20 iterasi

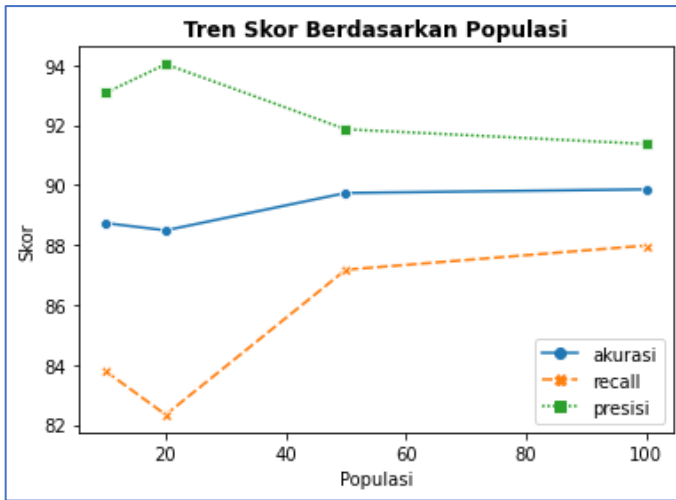
Pencarian parameter SVM menggunakan *Flower Pollination Algorithm* dengan 20 iterasi mendapatkan hasil yang tertera pada Tabel 6.12.

**Tabel 6.12 Pencarian Parameter dengan 20 Iterasi**

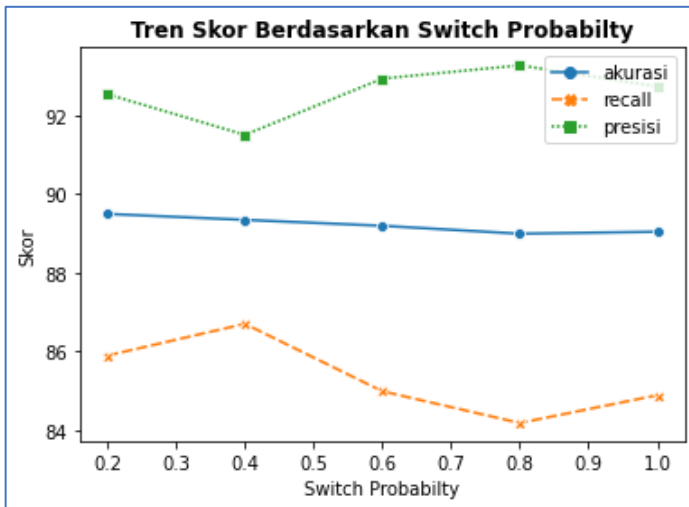
No	Pop	p	C	gamma	best
1.	10	0.2	7,721867	0,020816	89,122
2.	10	0.4	9,591819	0,020334	88,517
3.	10	0.6	7,013340	1,285271	86,891
4.	10	0.8	5,461338	0,030395	88,904
5.	10	1	109,875330	0,014144	89,231
6.	20	0.2	42,962735	0,007331	89,344
7.	20	0.4	13,679743	0,017287	88,944
8.	20	0.6	46,531852	0,024830	89,370
9.	20	0.8	5,484579	1,318499	86,891
10.	20	1	5,747164	1,271526	86,891
11.	50	0.2	5,492639	0,028612	88,904
12.	50	0.4	112,807224	0,014723	89,638
<b>13.</b>	<b>50</b>	<b>0.6</b>	<b>109,512116</b>	<b>0,016408</b>	<b>90,026</b>
<b>14.</b>	<b>50</b>	<b>0.8</b>	<b>110,862319</b>	<b>0,016187</b>	<b>90,026</b>
15.	50	1	41,982650	0,028688	89,370
16.	100	0.2	123,472325	0,014987	89,832
17.	100	0.4	98,292092	0,017833	89,599
18.	100	0.6	132,907817	0,014237	89,832
19.	100	0.8	371,298313	0,009279	89,404
<b>20.</b>	<b>100</b>	<b>1</b>	<b>115,721110</b>	<b>0,015601</b>	<b>90,026</b>

Pada percobaan dengan 20 iterasi diatas, skor terbaik yang didapatkan adalah 90,026 pada beberapa kombinasi skenario. Berdasarkan Gambar 6.28, banyaknya populasi memberikan dampak yang signifikan terhadap akurasi, *recall* dan presisi. 50

dan 100 populasi bunga mendapatkan rata-rata skor tertinggi diantara jumlah populasi lainnya.



Gambar 6.28 Tren Skor Berdasarkan Populasi Pada Percobaan dengan 20 Iterasi



Gambar 6.29 Tren Skor Berdasarkan Switch Probability Pada Percobaan dengan 10 Iterasi



Berdasarkan Gambar 6.29, skor akurasi memiliki tren yang sedikit menurun.

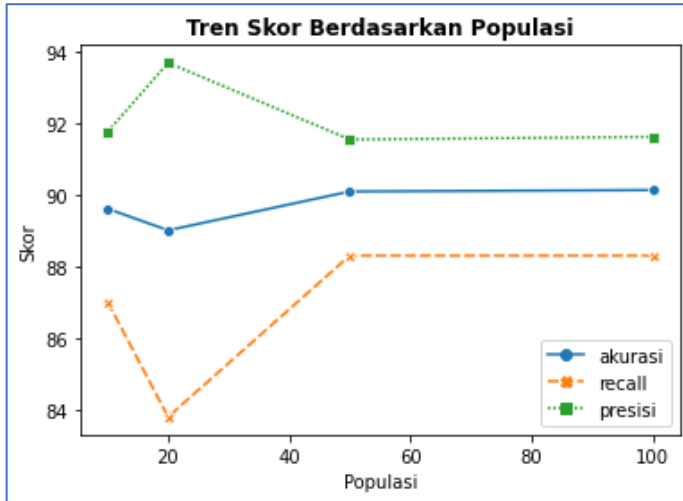
### 6.7.3 Percobaan dengan 50 iterasi

Pencarian parameter SVM menggunakan *Flower Pollination Algorithm* dengan 50 iterasi mendapatkan hasil yang tertera pada Tabel 6.13.

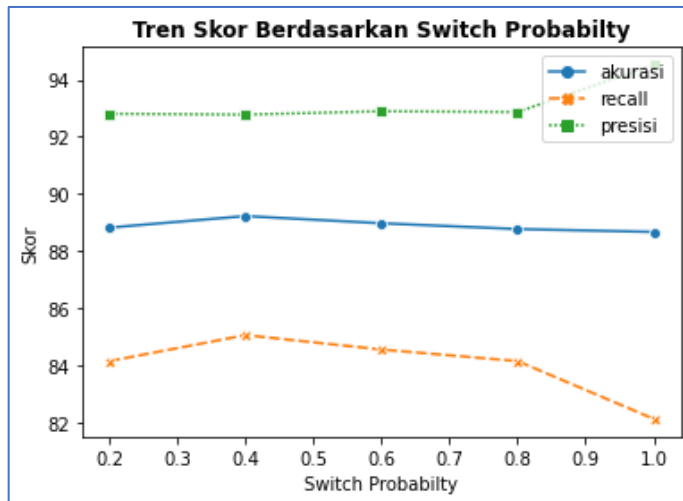
**Tabel 6.13 Pencarian Parameter dengan 50 Iterasi**

No	Pop	p	C	gamma	best
1.	10	0.2	159,607814	0,012498	89,813
2.	10	0.4	8,831193	0,019464	88,930
3.	10	0.6	113,446373	0,014559	89,638
4.	10	0.8	7,293768	0,021119	89,122
5.	10	1	135,523198	0,014262	89,832
6.	20	0.2	7,612701	0,021029	89,122
7.	20	0.4	7,385967	0,021267	89,122
8.	20	0.6	7,294261	0,021073	89,122
9.	20	0.8	130,780913	0,014328	89,832
10.	20	1	1,083191	1,275010	86,892
<b>11.</b>	<b>50</b>	<b>0.2</b>	<b>112,763537</b>	<b>0,015934</b>	<b>90,026</b>
<b>12.</b>	<b>50</b>	<b>0.4</b>	<b>107,067534</b>	<b>0,016452</b>	<b>90,026</b>
<b>13.</b>	<b>50</b>	<b>0.6</b>	<b>112,477926</b>	<b>0,015844</b>	<b>90,026</b>
14.	50	0.8	126,764192	0,014946	89,832
<b>15.</b>	<b>50</b>	<b>1</b>	<b>111,638194</b>	<b>0,015941</b>	<b>90,026</b>
<b>16.</b>	<b>100</b>	<b>0.2</b>	<b>111,188682</b>	<b>0,015872</b>	<b>90,026</b>
<b>17.</b>	<b>100</b>	<b>0.4</b>	<b>109,580968</b>	<b>0,016315</b>	<b>90,026</b>
<b>18.</b>	<b>100</b>	<b>0.6</b>	<b>114,362015</b>	<b>0,015727</b>	<b>90,026</b>
<b>19.</b>	<b>100</b>	<b>0.8</b>	<b>112,464797</b>	<b>0,015937</b>	<b>90,026</b>
<b>20.</b>	<b>100</b>	<b>1</b>	<b>106,010433</b>	<b>0,016507</b>	<b>90,026</b>

Pada percobaan dengan 50 iterasi diatas, skor terbaik yang didapatkan adalah 90.026 pada beberapa kombinasi skenario. Berdasarkan Gambar 6.30, skor akurasi memiliki tren yang meningkat.



Gambar 6.30 Tren Skor Berdasarkan Populasi Pada Percobaan dengan 50 Iterasi



Gambar 6.31 Tren Skor Berdasarkan Switch Probability Pada Percobaan dengan 50 Iterasi

Berdasarkan Gambar 6.31, skor akurasi memiliki tren yang stabil sedangkan skor *recall* memiliki tren turun seiring dengan meningkatnya *switch probability*.

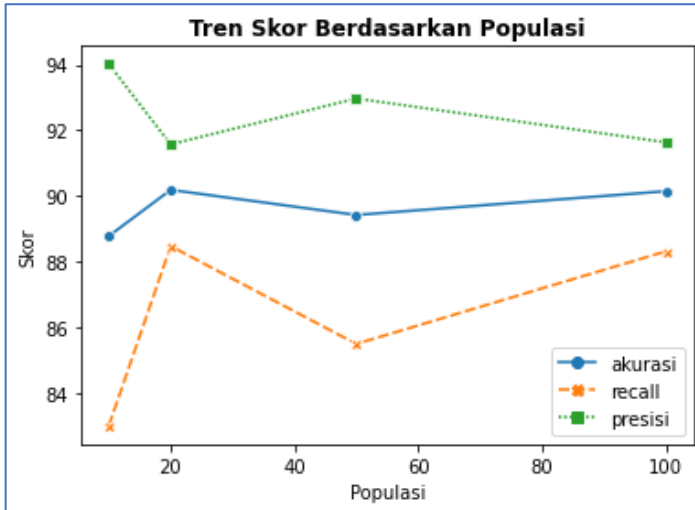
### 6.7.4 Percobaan dengan 100 iterasi

Pencarian parameter SVM menggunakan *Flower Pollination Algorithm* dengan 100 iterasi mendapatkan hasil yang tertera pada Tabel 6.14.

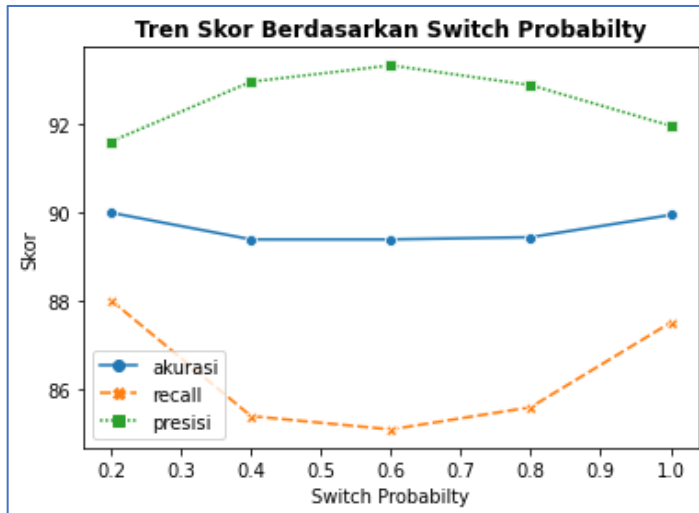
**Tabel 6.14 Pencarian Parameter dengan 100 Iterasi**

No	Pop	p	C	gamma	best
1.	10	0.2	115,080255	0,015718	90,026
2.	10	0.4	6,663923	1,262623	86,891
3.	10	0.6	111,346216	0,015915	90,026
4.	10	0.8	9,507593	1,300583	86,891
5.	10	1	7,794961	0,020330	89,122
6.	20	0.2	106,638509	0,016361	90,026
7.	20	0.4	110,287534	0,016061	90,026
8.	20	0.6	113,046718	0,015853	90,026
<b>9.</b>	<b>20</b>	<b>0.8</b>	<b>2829,0587</b>	<b>0,002501</b>	<b>90,259</b>
10.	20	1	112,930567	0,015790	90,026
11.	50	0.2	68,698390	0,021235	89,386
12.	50	0.4	116,464347	0,015596	90,026
13.	50	0.6	1,167524	1,297520	86,982
14.	50	0.8	112,737371	0,015886	90,026
15.	50	1	110,183547	0,016231	90,026
16.	100	0.2	112,918800	0,015738	90,026
17.	100	0.4	114,261243	0,015816	90,026
18.	100	0.6	109,485629	0,016176	90,026
19.	100	0.8	108,456297	0,016226	90,026
20.	100	1	109,261902	0,016137	90,026

Skor terbaik yang didapatkan adalah 90.259 yang terdapat pada beberapa percobaan diatas. Berdasarkan Gambar 6.32, skor akurasi memiliki tren yang stabil.

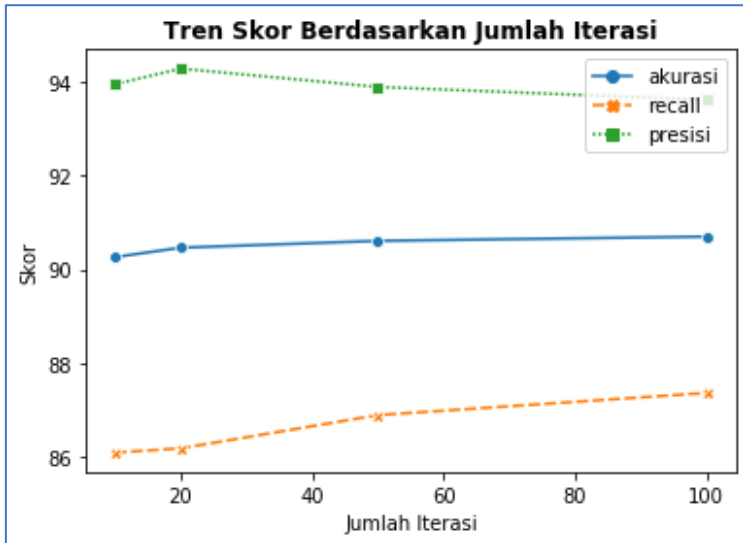


Gambar 6.32 Tren Skor Berdasarkan Populasi Pada Percobaan dengan 100 Iterasi



Gambar 6.33 Tren Skor Berdasarkan Switch Probability Pada Percobaan dengan 100 Iterasi

Berdasarkan Gambar 6.33, skor akurasi memiliki tren yang stabil.



Gambar 6.34 Tren Skor Berdasarkan Jumlah Iterasi

Berdasarkan Gambar 6.34, skor akurasi dan *recall* mengalami tren peningkatan seiring dengan bertambahnya jumlah iterasi.

Parameter yang akan digunakan untuk model SVM adalah  $C = 2829,0587$  dan  $\gamma = 0.002501$  pada percobaan dengan 100 iterasi, 20 populasi dan *switch probability* 0.8 karena mendapatkan nilai fungsi objektif tertinggi.

## 6.8 Prediksi SVM dan Evaluasi

Prediksi SVM dilakukan dengan menggunakan *library sklearn* dimana default parameter dari SVM terdapat pada sub bab 5.5.2 dan hasil pencarian parameter dengan *Flower Pollination Algorithm* terdapat pada bab 6.7.

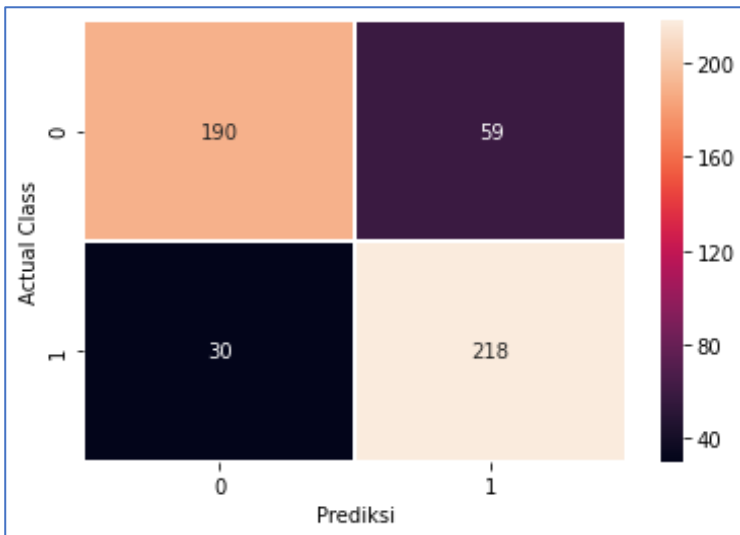
### 6.8.1 Prediksi SVM dengan Parameter Default

Parameter  $C$  dan  $\gamma$  yang digunakan sebagai parameter default berdasarkan *library sklearn* secara berturut-turut adalah 1 dan '*scale*'. Berikut adalah hasil evaluasi prediksi SVM dengan parameter default.

Tabel 6.15 Hasil Evaluasi Prediksi SVM dengan Parameter Default

Evaluasi	Skor Data Validasi
True Positive	218
True Negative	190
False Positive	59
False Negative	30
Akurasi	82,09
Recall	87,9
Precision	78,7

Berdasarkan Tabel 6.15, model SVM dengan parameter default tidak memiliki performa yang cukup baik karena memiliki akurasi  $> 80\%$ . Akurasi 80% merupakan standar akurasi yang harus di lampau karena data memiliki 80% bukan KLB, dimana jika model memprediksi semua data dengan bukan KLB maka model tersebut akan mendapatkan akurasi 80%.



Gambar 6.35 Confusion Matrix Model SVM dengan Default Parameter

Gambar 6.35 menunjukkan *confusion matrix* dari hasil prediksi model SVM dengan parameter *default*.

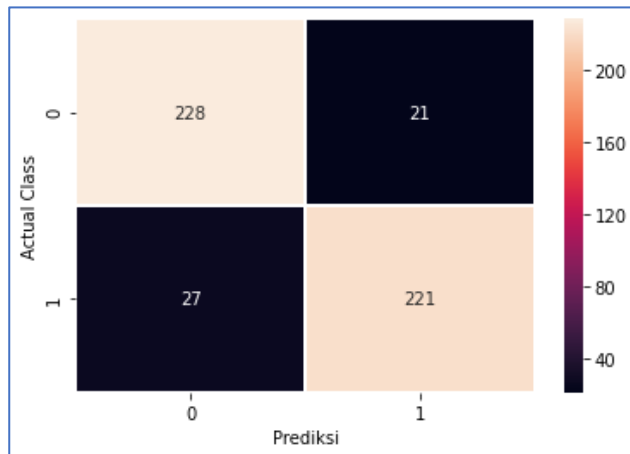
## 6.8.2 Prediksi SVM dengan Parameter Hasil Optimasi

Berikut adalah hasil evaluasi prediksi SVM dengan parameter terbaik pada hasil pencarian dengan *Flower Pollination Algorithm* ( $C = 2829,0587$  dan  $\text{gamma} = 0.002501$ ) di sub bab 6.7.

Tabel 6.16 Hasil Evaluasi Prediksi SVM dengan Parameter Hasil Optimasi

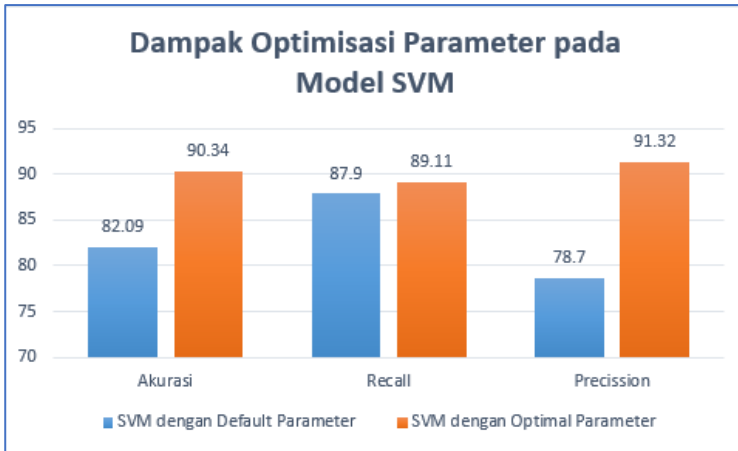
Evaluasi	Skor Data Validasi	Skor Data Tes
True Positive	221	38
True Negative	228	67
False Positive	21	66
False Negative	27	5
Akurasi	90,34	59,65
Recall	89,11	88,37
Precision	91,32	36,53

Berdasarkan Tabel 6.16, performa model SVM pada data validasi dengan menggunakan parameter hasil pencarian FPA sangat baik dimana model mendapatkan akurasi sebesar 90,34%, *recall* sebesar 89,11% dan *precision* sebesar 91,32%.



Gambar 6.36 Confusion Matrix Model SVM dengan Parameter Optimal Pada Data Validasi

Gambar 6.36 menunjukkan *confusion matrix* dari hasil prediksi model SVM pada data validasi dengan parameter optimal.



**Gambar 6.37 Dampak Optimisasi Parameter pada Model SVM**

Gambar 6.37 menunjukkan dampak dari optimisasi parameter pada model SVM dimana terjadi peningkatan yang signifikan terhadap akurasi dan presisi dari model.

Model SVM dengan parameter hasil pencarian FPA dilatih menggunakan data pembagian *5-fold cross validation*. Berikut adalah hasil evaluasi model tersebut.

**Tabel 6.17 Hasil Evaluasi Prediksi SVM dengan Parameter Hasil Optimisasi pada 5-fold cross validation**

Scoring	CV-1	CV-2	CV-3	CV-4	CV-5	Mean
<b>Akurasi</b>	64.4	71	62.2	62.5	61.4	64.3
<b>Recall</b>	84.3	93.8	98.5	98.5	97.4	94.5
<b>Precision</b>	60.3	64.5	57	57.3	56.6	59.1

Berdasarkan Tabel 6.17, model tidak mengalami *overfitting* maupun *underfitting* yang signifikan dengan skor yang memuaskan.



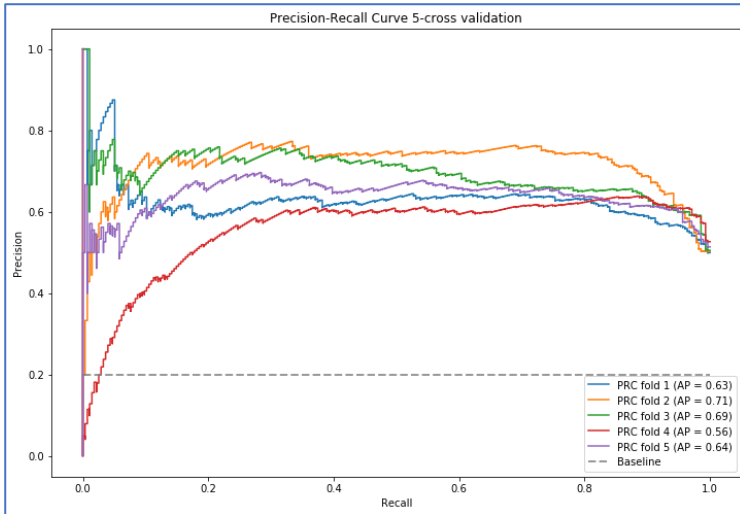
Tabel 6.18 Hasil Pencarian Feature Importance

Peringkat	Variabel	Information Gain
1	Kelembaban Udara	3,359
2	Curah Hujan	1,805
3	Kecepatan Angin	1,637
4	Kepadatan Penduduk	1,419
5	Suhu Udara	1,143
6	Angka Bebas Jentik	1,090

Tabel 6.18 merupakan hasil pencarian faktor-faktor yang mempengaruhi terjadinya KLB menggunakan nilai *information gain*. Kelembaban udara menjadi variabel yang paling berdampak terhadap KLB dan angka bebas jentik merupakan variabel yang paling tidak berdampak terhadap KLB. Periode yang paling berpotensi untuk terjadi KLB adalah bulan Februari dan Oktober.

Prediksi SVM pada data uji (dataset KLB tahun 2018) dilakukan dengan tujuan untuk menguji apakah model cukup baik untuk diterapkan sebagai model prediksi Kejadian Luar Biasa DBD di Kabupaten Malang.

Akurasi yang didapatkan dalam memprediksi KLB tahun 2018 adalah 59,65%, *recall* sebesar 88,37%, dan *precision* sebesar 36,53%. Berdasarkan hasil tersebut, dapat disimpulkan bahwa model dapat memprediksi 88,37% Kejadian Luar Biasa dengan benar pada data tahun 2018. Namun dari seluruh hasil prediksi yang menunjukkan KLB, terdapat hanya 36,53% prediksi yang tepat. Berdasarkan hal tersebut, model SVM mampu memprediksi KLB di Kabupaten Malang pada tahun 2018 tetapi model SVM melakukan banyak kesalahan prediksi pada data tidak KLB. Hasil akurasi dan *precision* yang didapatkan pada data uji masih dibawah 80%, hal ini kemungkinan dikarenakan nilai parameter C yang terlalu besar sehingga terjadi *overfitting* pada data latih. Oleh karena nilai akurasi, *recall* dan *precision* tidak dapat diinterpretasi langsung dari nilai masing-masing yang diakibatkan oleh pola data yang tidak berimbang sehingga dilakukan analisa PRC.



**Gambar 6.38 Precision-recall Curve dengan 5-fold Cross Validation**

Gambar 6.38 merupakan hasil visualisasi PRC dengan *5-fold cross validation*. Berdasarkan Gambar 6.38 tersebut, model prediksi memiliki performa yang sangat baik dari setiap lipatan yang ditunjukkan dengan tiap garis berada jauh diatas garis basis, dan rataan luas *precision* adalah 0,65.

## **BAB VII**

### **KESIMPULAN DAN SARAN**

Dalam bab ini dijelaskan tentang kesimpulan dan saran yang didapatkan dari proses pelaksanaan pengerjaan tugas akhir ini.

#### **7.1 Kesimpulan**

Kesimpulan yang didapatkan dari pengerjaan tugas akhir yang telah dilakukan adalah sebagai berikut:

1. Model prediksi berbasis Support Vector Machine yang melibatkan Flower Pollination Algorithm (SVM-FPA) untuk memprediksi kejadian luar biasa (KLB) pada kasus demam berdarah dengue (DBD) mampu memberikan hasil prediksi dengan cukup baik. Hal ini ditunjukkan oleh hasil eksperimen bahwa model tersebut mampu memprediksi lebih dari 80% periode KLB baik pada data validasi maupun data tes.
2. Dengan menggunakan pengaturan nilai parameter yang optimal dan evaluasi kinerja model menggunakan 5-fold cross validation, kinerja model SVM-FPA yang diimplementasikan dalam Tugas Akhir ini mampu memberikan hasil akurasi, recall, dan presisi berturut-turut sebesar 64,3%, 94,5%, dan 59,1%. Tetapi, oleh karena variabel yang dijadikan target mempunyai label kelas yang tidak berimbang, maka diperlukan analisis precision-recall curve (PRC) untuk mengevaluasi hasil prediksi dari model. Hasil plotting PRC menunjukkan bahwa model SVM-FPA mempunyai kinerja yang cukup baik dan berada jauh di atas garis basis dengan nilai average precision (AP) sebesar 0,65.
3. Pengaturan nilai parameter cost dan gamma pada model SVM-FPA sangat mempengaruhi hasil prediksi. Hal ini ditunjukkan oleh hasil eksperimen bahwa pengaturan nilai parameter default yang tidak optimal menghasilkan akurasi prediksi yang kurang baik. Di lain pihak, dengan menentukan nilai parameter yang optimal (nilai cost dan gamma berturut-turut sebesar 2829,0587 dan 0,002801),

model SVM-FPA mampu memprediksi KLB dengan lebih akurat. Selain itu faktor yang mempengaruhi terjadinya KLB adalah kelembaban udara, curah hujan, kecepatan angin, kepadatan penduduk, suhu udara, dan angka bebas jentik (berurutan dari yang paling signifikan).

4. Secara umum, pengaturan nilai parameter optimal dari model SVM-FPA mampu memberikan hasil prediksi yang cukup baik. Namun demikian, beberapa kasus diprediksi sebagai KLB walaupun seharusnya tidak terjadi KLB. Kesalahan prediksi ini kemungkinan dikarenakan jumlah data latih yang digunakan relatif sedikit yang hanya mampu memprediksi KLB untuk satu periode ke depan.

## 7.2 Saran

Seperti dijelaskan dalam kesimpulan bahwa model prediksi SVM-FPA yang diimplemmentasikan dalam Tugas Akhir ini memberikan hasil prediksi data tes yang kurang memuaskan. Selain itu, terdapat kasus yang seharusnya tidak terjadi KLB tetapi diprediksi KLB. Berikut ini diberikan beberapa kemungkinan perbaikan yang dapat dilakukan untuk penyempurnaan implementasi model SVM-FPA.

1. Memperbanyak kombinasi skenario pengaturan parameter SVM-FPA untuk menemukan kombinasi nilai parameter yang lebih baik yang pada gilirannya akan meningkatkan kinerja prediksi dari model.
2. Menambah beberapa variabel yang berkorelasi dengan kejadian luar biasa DBD, seperti frekuensi pengurasan air dan frekuensi penanganan dini (*fogging*) karena beberapa variabel tersebut mempengaruhi populasi dari nyamuk penyebab DBD.
3. Menambah data uji coba dengan memperluas cakupan daerah yang digunakan dalam pembangunan model prediksi. Keterbatasan data di daerah Kabupaten Malang semata-mata dikarenakan adanya kesulitan untuk memperoleh data pada masa pandemi Covid-19, sehingga terjadi keadaan *overfitting*, di mana hasil prediksi data tes lebih jelek dibandingkan hasil prediksi data validasi.

*Halaman ini sengaja dikosongkan*

## DAFTAR PUSTAKA

- [1] WHO, “Dengue and Severe Dengue,” 15 April 2019. [Online]. Available: <https://www.who.int/en/news-room/fact-sheets/detail/dengue-and-severe-dengue>. [Diakses 30 September 2019].
- [2] D. Widiastuti, A. P. Kesuma dan N. Pramestuti, “Indeks Entomologi Dan Transmisi Transovari Yang Mendukung Peningkatan Kasus Demam Berdarah Dengue Di Kabupaten Banjarnegara,” *Loka Litbang P2B2 Baturaja*, vol. 8, pp. 30-37, 2016.
- [3] Y. A. Indrayani dan T. Wahyudi, “Situasi Penyakit Demam Berdarah Di Indonesia Tahun 2017,” Pusat Data dan Informasi Kementerian Kesehatan RI, 2018.
- [4] S. Hardiyanto, “JawaPos.com,” PT Jawa Pos Grup Multimedia, 31 1 2019. [Online]. Available: <https://www.jawapos.com/jpg-today/31/01/2019/januari-72-warga-di-kabupaten-malang-terjangkit-demam-berdarah/>. [Diakses 4 10 2019].
- [5] K. Mulligan, “Dengue and development: a critical political”.
- [6] L. Yongqian, S. Ying , D. Infield, Z. Yu , H. Shuang dan Y. Jie , “A Hybrid Forecasting Method for Wind Power Ramp Based on Orthogonal Test and Support Vector Machine (OT-SVM),” *IEEE TRANSACTIONS ON SUSTAINABLE ENERGY*, vol. 8, no. 2, pp. 451-457, 2017.
- [7] Y. Yusof dan Z. Mustafa, “Dengue Outbreak Prediction: A Least Squares Support Vector Machines Approach,”

*International Journal of Computer Theory and Engineering*, vol. 3, no. 4, pp. 489-493, 2011.

- [8] K. Kesorn, P. Ongruk, J. Chompoonsri, A. Phumee, U. Thavara, A. Tawatsin dan P. Siriyasatien, "Morbidity Rate Prediction of Dengue Hemorrhagic Fever (DHF) Using the Support Vector Machine and the *Aedes aegypti* Infection Rate in Similar Climates and Geographical Areas," *Plus One*, 2015.
- [9] J.-C. Coetsier dan R. Jiamthapthaksin, "Parallelized FPA-SVM: Parallelized Parameter Selection and Classification using Flower Pollination Algorithm and Support Vector Machine," dalam *2017 14th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, Nakhon Si Thammarat, 2017.
- [10] Z. Mustaffa, M. H. Sulaiman, F. Ernawan dan Y. Yusuf, "Dengue Outbreak Prediction: Hybrid Meta-heuristic," dalam *2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, Busan, 2018.
- [11] C. C. Ho, T. Choo-Yee dan D. B. Raja, "Using Public Open Data to Predict Dengue Epidemic: Assessment of Weather Variability, Population Density, and Land use as Predictor Variables for Dengue Outbreak Prediction using Support Vector Machine," *Indian Journal of Science and Technology*, vol. 11, no. 4, 2018.
- [12] S. Haryanto, B. Yohan, M. S. Santoso, R. F. Hayati, D. Denis, G. Udjung, H. R. I. Kendarsari, H. Trimarsanto dan R. T. Sasmono, "Clinical features and virological confirmation of perinatal dengue infection in Jambi,

Indonesia: A case report,” *International Journal of Infectious Diseases*, vol. 86, pp. 197-200, 2019.

- [13] F. S. Keman dan C. U. Wahyun, “PERAN FAKTOR LINGKUNGAN DAN PERILAKU TERHADAP PENULARAN DEMAM BERDARAH DENGUE DI KOTA MATARAM,” *JURNAL KESEHATAN LINGKUNGAN*, vol. 2, no. 1, pp. 1-10, 2005.
- [14] “Peraturan Menteri Kesehatan Republik Indonesia No. 949 tahun 2004 tentang Pedoman Penyelenggaraan Sistem Kewaspadaan Dini Kejadian Luar Biasa”.
- [15] W. ANGGRAENI, I. P. A. A. PRAMANA, F. SAMOPA, E. RIKSAKOMARA, R. P. WIBOWO, L. C. T dan P. , “FORECASTING THE NUMBER OF DENGUE FEVER CASES IN MALANG REGENCY INDONESIA USING FUZZY INFERENCE SYSTEM MODELS,” *Journal of Theoretical and Applied Information Technology*, vol. 95, no. 1, 2017.
- [16] A. Purwar dan S. . K. Singh, “Hybrid prediction model with missing value imputation for medical data,” *Expert Systems with Applications*, vol. 42, no. 13, pp. 5621-5631, 2015.
- [17] A. Fernandez, S. Garc’ia, F. Herrera dan N. Chawla, “SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary,” *Journal of Artificial Intelligence Research*, no. 61, pp. 863-905, 2018.
- [18] S. Defiyanti, “Analisis dan Prediksi Kinerja Mahasiswa Menggunakan Teknik Data Mining,” *Jurnal Informatika*, vol. 2, no. 1, 2013.



- [19] A. Abdelaziz, E. Ali dan S. A. Elazim, "Implementation of flower pollination algorithm for solving economic load dispatch and combined economic emission dispatch problems in power systems," *Energy*, vol. 101, pp. 506-518, 2016.
- [20] V. Puspita, E. dan D. M. Tarigan, "Time Series untuk Peramalan Konsumsi Listrik Menggunakan Metode Support Vector Machine (SVM)," dalam *Annual Research Seminar (ARS)*, Palembang, 2017.
- [21] A. Darmawan, N. Kustian dan W. Rahayu, "IMPLEMENTASI DATA MINING MENGGUNAKAN MODEL SVMUNTUK PREDIKSI KEPUASAN PENGUNJUNG TAMAN TABEBUYA," *Jurnal String*, vol. 2, no. 3, pp. 299-3017, 2018.
- [22] T. Saito dan M. Rehmsmeier, "The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets," *PLOS ONE*, vol. 10, no. 3, 2015.

*Halaman ini sengaja dikosongkan*

## BIODATA PENULIS



Penulis bernama Yusnardo Tendio, lahir di Palu, Sulawesi Tengah pada tanggal 7 Februari 1999, yang merupakan anak pertama dari tiga bersaudara. Penulis telah menempuh jenjang Pendidikan formal di beberapa sekolah yaitu: SMPN 1 Palu (2012-2013), dan SMAN 1 Palu (2013-2016). Penulis melanjutkan Pendidikan sarjana di Departemen Sistem Informasi Fakultas Teknologi Elektro dan Informatika Cerdas (FTEIC) Institut Teknologi

Sepuluh Nopember (ITS) pada tahun 2016 yang terdaftar sebagai mahasiswa dengan nrp 05211640000149.

Selama menjadi mahasiswa, penulis aktif mengikuti beberapa kepanitiaan seperti Education Fair Sulawesi Tengah 2017 dan *Indonesian Youth for SDGs* 2017. Di bidang akademik, penulis aktif menjadi asisten dosen pada mata kuliah Statistika dan asisten praktikum pada mata kuliah Pengantar Sistem Operasi. Selain itu pada tahun 2020, penulis mengikuti perlombaan dan menjadi finalis pada kompetisi Data Science Arkavidia 6.0 yang diadakan oleh Himpunan Mahasiswa Informatika STEI ITB dan finalis Data Mining Competition JOINTS UGM yang diadakan oleh Fakultas Matematika dan Ilmu Pengetahuan Alam UGM. Penulis juga sempat magang sebagai *Software Engineering* di PT. Sribu Digital Kreatif.

Untuk mendapatkan gelar S.Kom (Sarjana Komputer), penulis mengambil topik penelitian tugas akhir klasifikasi pada laboratorium Rekayasa Data dan Intelegensi Bisnis (RDIB). Untuk kepentingan penelitian, Penulis dapat dihubungi melalui email [yusnardotendio@yahoo.com](mailto:yusnardotendio@yahoo.com)

*Halaman ini sengaja dikosongkan*

## LAMPIRAN A HASIL NORMALISASI

Adapun seluruh data hasil normalisasi dan *oversampling* akan dilampirkan pada *link* [https://bit.ly/hasil\\_normalisasi](https://bit.ly/hasil_normalisasi).