



TUGAS AKHIR - SS141501

**KLASIFIKASI KELAS RISIKO PASIEN PNEUMONIA  
MENGUNAKAN METODE *SUPPORT VECTOR  
MACHINE-GENETIC ALGORITHM (SVM-GA) HYBRID***

**RAHAJENG RINDRASARI  
NRP 1314 100 117**

**Dosen Pembimbing  
Irhamah, M.Si, Ph.D**

**PROGRAM STUDI SARJANA  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2018**





**TUGAS AKHIR - SS141501**

**KLASIFIKASI KELAS RISIKO PASIEN PNEUMONIA  
MENGUNAKAN METODE *SUPPORT VECTOR  
MACHINE-GENETIC ALGORITHM (SVM-GA) HYBRID***

**RAHAJENG RINDRASARI  
NRP 1314 100 117**

**Dosen Pembimbing  
Irhamah, M.Si, Ph.D**

**PROGRAM STUDI SARJANA  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2018**

*(Halaman ini sengaja dikosongkan)*



**FINAL PROJECT - SS141501**

**CLASSIFICATION OF PNEUMONIA PATIENT USING  
SUPPORT VECTOR MACHINE-GENETIC ALGORITHM  
(SVM-GA) HYBRID METHOD**

**RAHAJENG RINDRASARI  
NRP 1314 100 117**

**Supervisor  
Irhamah, M.Si, Ph.D**

**UNDERGRADUATE PROGRAMME  
DEPARTMENT OF STATISTICS  
FACULTY OF MATHEMATICS, COMPUTING, AND DATA SCIENCES  
INSTITUTE TECHNOLOGY OF SEPULUH NOPEMBER  
SURABAYA 2018**

*(Halaman Ini Sengaja Dikosongkan)*

## LEMBAR PENGESAHAN

### KLASIFIKASI KELAS RISIKO PASIEN PNEUMONIA MENGGUNAKAN *SUPPORT VECTOR MACHINE-GENETIC ALGORITHM (SVM-GA) HYBRID*

#### TUGAS AKHIR

Diajukan Untuk Memenuhi Salah Satu Syarat  
Memperoleh Gelar Sarjana Sains  
pada  
Program Studi Sarjana Departemen Statistika  
Fakultas Matematika, Komputasi, dan Sains Data  
Institut Teknologi Sepuluh Nopember

Oleh :

**Rahajeng Rindrasari**  
NRP. 1314 100 117

Disetujui oleh Pembimbing:  
**Irhamah, M.Si., Ph.D**  
NIP : 19780406 200112 2 002

(  )

Mengetahui,  
Kepala Departemen



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

SURABAYA, JANUARI 2018



# **KLASIFIKASI KELAS RISIKO PASIEN PNEUMONIA MENGUNAKAN METODE *SUPPORT VECTOR MACHINE- GENETIC ALGORITHM (SVM-GA) HYBRID***

**Nama Mahasiswa** : Rahajeng Rindrasari  
**NRP** : 1314100117  
**Departemen** : Statistika  
**Dosen Pembimbing** : Irhamah, M.Si., Ph.D

## **Abstrak**

*Pneumonia, atau sering disebut radang paru-paru, dapat menimbulkan kerusakan pada jaringan paru-paru dan gangguan pertukaran gas setempat sehingga pneumonia dapat mengakibatkan sakit yang parah bahkan sampai berujung kematian. Dibutuhkan pengklasifikasian penyakit pneumonia untuk mempercepat dalam menentukan tingkat keparahan penyakit serta mengetahui tindakan yang paling tepat untuk penderita. Support Vector Machine dapat mengatasi permasalahan klasifikasi yang linier maupun tidak linier. Dalam penelitian ini, metode Genetic Algorithm dipakai dalam optimasi parameter. Penelitian ini menerapkan metode Grid Search Support Vector Machine dengan seleksi variabel dengan FCBF dan GA, serta Support Vector Machine-Genetic Algorithm (SVM-GA) Hybrid untuk mengklasifikasi kelompok kelas risiko berdasarkan rekam data medis pasien kemudian metode-metode tersebut dibandingkan nilai akurasi. Hasil analisis menunjukkan bahwa metode SVM-GA dengan seleksi variabel GA memberikan nilai ketepatan klasifikasi yang lebih tinggi dibandingkan metode SVM dengan tanpa seleksi variabel serta seleksi variabel GA. Selain itu, optimasi parameter menggunakan metode GA dapat meningkatkan nilai akurasi pada data.*

**Kata kunci** : *genetic algorithm, optimasi, pneumonia, seleksi variabel, support vector machine*

*(Halaman Ini Sengaja Dikosongkan)*

# CLASSIFICATION OF PNEUMONIA PATIENT USING SUPPORT VECTOR MACHINE-GENETIC ALGORITHM (SVM-GA) HYBRID

**Name** : Rahajeng Rindrasari  
**NRP** : 1314100117  
**Department** : Statistics  
**Supervisor** : Irhamah, M.Si., Ph.D

## **Abstract**

*Pneumonia, or it is often called pneumonia, can cause damage to lung tissue and local gas exchange disturbances so that pneumonia can cause severe pain even to death. The classification of pneumonia disease is needed to accelerate in determining the severity of the disease as well as knowing the most appropriate action for the patient. Support Vector Machine can solve linear or non-linear classification problem. In this research, Genetic Algorithm method is used in parameter optimization. In this research applies Grid Search Support Vector Machine method with variable selection with FCBF and GA, and Support Vector Machine-Genetic Algorithm Hybrid (SVM-GA) to classify risk class group based on patient medical data record then those methods are compared the value of accuracy. The results of the analysis show that the SVM-GA method with GA variable selection gives a higher classification precision value than the SVM method with no variable selection and GA variable selection. In addition, parameter optimization using the GA method can improve the accuracy value of the data.*

**Keywords:** *genetic algorithm, optimization, pneumonia, support vector machine, variable selection*

*(Halaman Ini Sengaja Dikosongkan)*

## KATA PENGANTAR

Puji syukur kehadirat Tuhan Yang Maha Esa, atas segala berkat dan karunia-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul **“Klasifikasi Kelas Risiko Pasien Pneumonia Menggunakan Metode *Support Vector Machine-Genetic Algorithm* (SVM-GA) Hybrid”** dengan lancar.

Keberhasilan penyusunan Tugas Akhir ini tidak lepas dari banyaknya bantuan yang diberikan berbagai pihak. Oleh karena itu, pada kesempatan ini penulis mengucapkan terima kasih kepada:

1. Irhamah, M.Si, Ph.D sebagai selaku dosen pembimbing dalam memberikan bimbingan selama penyusunan Tugas Akhir. Prof. Drs. NUR Iriawan, MIKom., Ph.D. dan Pratnya Paramitha Oktaviana, S.Si., M.Si. selaku dosen penguji yang telah memberikan banyak bantuan dan masukan.
2. Dr. Suhartono selaku Ketua Jurusan Statistika ITS dan Bapak Dr. Sutikno, S.Si, M.Si selaku ketua prodi S1 Statistika ITS. Dr. Dra. Agnes Tuti Rumiati, M.Sc selaku dosen wali atas dukungan yang diberikan selama proses perkuliahan.
3. Bapak Agus Nugroho dan Ibu Rohyati Santoen, Mas Mbak dan Keponakan penulis yang memberikan dukungan dan doa dalam menyelesaikan Tugas Akhir.
4. Mbak Neni, Mbak Atus, Mbak Ageng, Mas Adam, dan Raju yang selalu bersedia menjadi tempat konsultasi penulis dalam pembuatan Tugas Akhir.
5. Cristian Monang P. X. Lumbanbatu, Tanti, Nikita, Alifian, Erlin, Indah, Atiti, Intan, Putri, David, Harvey, Mbak Bella dan Mbak Retty, Irene Putri Rachel serta teman-teman RESPECT 2014 yang menjadi teman dan selalu mendukung penulis.

Penulis berharap hasil Tugas Akhir ini dapat bermanfaat bagi banyak pihak. Semoga kebaikan dan bantuan yang telah diberikan kepada penulis dibalas dengan kebaikan yang lebih besar lagi oleh Tuhan Yang Maha Esa. Amin.

Surabaya, Desember 2017

Penulis

*(Halaman Ini Sengaja Dikosongkan)*

## DAFTAR ISI

<b>HALAMAN JUDUL</b> .....	<b>i</b>
<b>PAGE TITLE</b> .....	<b>iii</b>
<b>LEMBAR PENGESAHAN</b> .....	<b>v</b>
<b>ABSTRAK</b> .....	<b>i</b>
<b>ABSTRACT</b> .....	<b>ix</b>
<b>KATA PENGANTAR</b> .....	<b>xi</b>
<b>DAFTAR ISI</b> .....	<b>xiii</b>
<b>DAFTAR TABEL</b> .....	<b>xv</b>
<b>DAFTAR GAMBAR</b> .....	<b>xvii</b>
<b>DAFTAR LAMPIRAN</b> .....	<b>xix</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	4
1.3 Tujuan Penelitian .....	4
1.4 Manfaat Penelitian .....	5
1.5 Batasan Masalah .....	5
<b>BAB II TINJAUAN PUSTAKA</b> .....	<b>7</b>
2.1 <i>Fast Correlation-Based Filter (FCBF)</i> .....	7
2.2 <i>Support Vector Machine (SVM)</i> .....	9
2.3 <i>Genetic Algorithm (GA)</i> .....	14
2.4 Evaluasi Kinerja Klasifikasi .....	21
2.5 <i>K-Fold Cross Validation</i> .....	22
2.6 Pneumonia .....	23
2.7 Diagnosis Pneumonia .....	25
<b>BAB III METODOLOGI PENELITIAN</b> .....	<b>29</b>
3.1 Sumber Data .....	29
3.2 Variabel Penelitian .....	29
3.3 Struktur Data .....	31
3.4 Langkah Analisis .....	31
<b>BAB IV ANALISIS DAN PEMBAHASAN</b> .....	<b>39</b>
4.1 Karakteristik Penderita Pneumonia .....	39
4.2 Klasifikasi dengan <i>Grid Search Support Vector Machine</i> .....	47
4.3 Optimasi Parameter <i>Support Vector Machine</i> dengan <i>Genetic Algorithm</i> .....	63
4.4 Perbandingan Nilai Ketepatan Klasifikasi .....	76

<b>BAB V KESIMPULAN DAN SARAN .....</b>	<b>79</b>
5.1 Kesimpulan .....	79
5.2 Saran .....	80
<b>DAFTAR PUSTAKA .....</b>	<b>81</b>
<b>LAMPIRAN .....</b>	<b>81</b>

## DAFTAR TABEL

<b>Tabel 2.1</b> Tabel Klasifikasi .....	21
<b>Tabel 2.2</b> Skor Prediksi CURB-65 .....	26
<b>Tabel 2.3</b> Kelas Risiko dengan Sistem CURB-65.....	26
<b>Tabel 2.4</b> Skor pada Pneumonia Komunitas Sistem PSI.....	26
<b>Tabel 2.5</b> Derajat Skor Penentuan Kelas Risiko Dengan Sistem PSI .....	27
<b>Tabel 3.1</b> Variabel Penelitian .....	29
<b>Tabel 3.2</b> Definisi Operasional.....	30
<b>Tabel 3.3</b> Struktur Data .....	31
<b>Tabel 4.1</b> Rata-rata tiap Variabel dan tiap Kelas Klasifikasi Pasien Pneumonia .....	40
<b>Tabel 4.2</b> Kombinasi <i>Range</i> Parameter untuk <i>Grid Search</i> SVM...	48
<b>Tabel 4.3</b> Hasil Percobaan <i>Grid Search</i> SVM Tanpa Seleksi Variabel .....	49
<b>Tabel 4.4</b> <i>Confusion Matrix</i> Klasifikasi pada Data <i>Testing</i> Pembagian Data Pertama.....	50
<b>Tabel 4.5</b> Hasil Kinerja Klasifikasi Data Pasien Pneumonia .....	51
<b>Tabel 4.6</b> Hasil Percobaan <i>Grid Search</i> SVM Seleksi Variabel FCBF .....	53
<b>Tabel 4.7</b> Hasil Kinerja Klasifikasi Dari Data Seleksi Variabel FCBF .....	54
<b>Tabel 4.8</b> Ilustrasi Kromosom Dalam Populasi Awal pada Seleksi Variabel .....	55
<b>Tabel 4.9</b> Ilustrasi Kromosom Dalam Populasi Awal pada Seleksi Variabel dengan Nilai <i>Fitness</i> .....	55
<b>Tabel 4.10</b> Ilustrasi Proses RWS pada Seleksi Variabel.....	56
<b>Tabel 4.11</b> Kromosom dan Nilai <i>Fitness Genetic Algorithm</i> pada Seleksi Variabel .....	58
<b>Tabel 4.12</b> <i>Cross Validation Genetic Algorithm</i> pada Seleksi Variabel.....	58
<b>Tabel 4.13</b> Hasil Percobaan <i>Grid Search</i> SVM Seleksi Variabel GA.....	59
<b>Tabel 4.14</b> Hasil Kinerja Klasifikasi dari Data Seleksi Variabel GA .....	59
<b>Tabel 4.15</b> <i>Cross Validation Genetic Algorithm</i> pada Seleksi Variabel GA .....	61

<b>Tabel 4.16</b> Hasil Percobaan <i>Grid Search</i> SVM Seleksi Variabel GA .....	61
<b>Tabel 4.17</b> Hasil Kinerja Klasifikasi dari Data Seleksi Variabel GA .....	62
<b>Tabel 4.18</b> Ilustrasi Kromosom Dalam Populasi Awal pada Optimasi Parameter .....	63
<b>Tabel 4.19</b> Ilustrasi Kromosom Dalam Populasi Awal pada Optimasi Parameter .....	64
<b>Tabel 4.20</b> Ilustrasi Proses RWS pada Optimasi Parameter .....	64
<b>Tabel 4.21</b> Hasil Percobaan SVM-GA Tanpa Seleksi Variabel dengan <i>Fitness AUC</i> .....	67
<b>Tabel 4.22</b> Hasil Kinerja Klasifikasi SVM-GA Data Tanpa Seleksi Variabel .....	67
<b>Tabel 4.23</b> Hasil Percobaan SVM-GA Tanpa Seleksi Variabel dengan <i>Fitness Akurasi</i> .....	68
<b>Tabel 4.24</b> Hasil Kinerja Klasifikasi SVM-GA Data Tanpa Seleksi Variabel .....	69
<b>Tabel 4.25</b> Hasil Percobaan SVM-GA Seleksi Variabel FCBF dengan <i>Fitness AUC</i> .....	70
<b>Tabel 4.26</b> Hasil Kinerja Klasifikasi SVM-GA Data Seleksi Variabel FCBF .....	70
<b>Tabel 4.27</b> Hasil Percobaan SVM-GA Seleksi Variabel FCBF dengan <i>Fitness Akurasi</i> .....	71
<b>Tabel 4.28</b> Hasil Kinerja Klasifikasi SVM-GA Data Seleksi Variabel FCBF .....	72
<b>Tabel 4.29</b> Hasil Percobaan SVM-GA Seleksi Variabel FCBF dengan <i>Fitness AUC</i> .....	73
<b>Tabel 4.30</b> Hasil Kinerja Klasifikasi SVM-GA Data Seleksi Variabel GA .....	74
<b>Tabel 4.31</b> Hasil Percobaan SVM-GA Seleksi Variabel GA dengan <i>Fitness Akurasi</i> .....	74
<b>Tabel 4.32</b> Hasil Kinerja Klasifikasi SVM-GA Data Seleksi Variabel GA .....	75
<b>Tabel 4.33</b> Perbandingan Nilai Ketepatan Klasifikasi <i>AUC</i> .....	76
<b>Tabel 4.34</b> Perbandingan Nilai Ketepatan Klasifikasi Akurasi .....	77

## DAFTAR GAMBAR

<b>Gambar 2.1</b> <i>Hyperplane</i> pada SVM <i>Linear Separable</i> (Kusumaningrum, 2017).....	9
<b>Gambar 2.2</b> Konsep <i>Hyperplane</i> pada SVM <i>linear nonseparable</i> (Kusumaningrum, 2017).....	12
<b>Gambar 2.3</b> <i>Hyperplane</i> pada <i>Nonlinear SVM</i> .....	13
<b>Gambar 2.4</b> Istilah dalam <i>Genetic Algorithm</i> .....	15
<b>Gambar 2.5</b> Ilustrasi Proses Mutasi.....	20
<b>Gambar 2.6</b> Ilustrasi Proses <i>5-Fold Cross-Validation</i> .....	23
<b>Gambar 3.1</b> Diagram Alir Penelitian.....	35
<b>Gambar 3.2</b> Diagram Alir Klasifikasi Menggunakan <i>Grid Search</i> SVM.....	36
<b>Gambar 3.3</b> Diagram Alir Menggunakan FCBF .....	37
<b>Gambar 3.4</b> Diagram Alir Menggunakan GA .....	38
<b>Gambar 4.1</b> Proporsi Tiap Kelas Pasien Pneumonia.....	39
<b>Gambar 4.2</b> Diagram Batang Jenis Kelamin terhadap Perbedaan Kelas Klasifikasi Pasien Pneumonia .....	40
<b>Gambar 4.3</b> Proporsi Tiap Kategori <i>Boxplot</i> dari Variabel Usia....	41
<b>Gambar 4.4</b> Proporsi Tiap Kategori <i>Boxplot</i> dari Variabel Tekanan Darah (a) Sistolik, (b) Diastolik.....	41
<b>Gambar 4.5</b> Proporsi Tiap Kategori <i>Boxplot</i> dari Variabel Denyut Nadi .....	42
<b>Gambar 4.6</b> Proporsi Tiap Kategori <i>Boxplot</i> dari Variabel Frekuensi Nafas tiap Menit.....	42
<b>Gambar 4.7</b> Proporsi Tiap Kategori <i>Boxplot</i> dari Variabel Suhu Tubuh Pasien .....	43
<b>Gambar 4.8</b> Proporsi Tiap Kategori <i>Boxplot</i> dari pH Darah.....	43
<b>Gambar 4.9</b> Proporsi Tiap Kategori <i>Boxplot</i> dari Variabel Kadar Natrium dalam Darah .....	44
<b>Gambar 4.10</b> Proporsi Tiap Kategori <i>Boxplot</i> dari Variabel Kadar Glukosa dalam Darah.....	45
<b>Gambar 4.11</b> Proporsi Tiap Kategori <i>Boxplot</i> dari Variabel Hematokrit .....	45
<b>Gambar 4.12</b> Rata-rata Hematokrit Pasien Pneumonia Berdasarkan Jenis Kelamin dan Kelas Klasifikasi.....	46
<b>Gambar 4.13</b> Proporsi Tiap Kategori <i>Boxplot</i> dari variabel Kadar PO <sub>2</sub> .....	46

<b>Gambar 4.14</b> Proporsi Tiap Kategori <i>Boxplot</i> dari Variabel Urea Nitrogen Darah .....	47
<b>Gambar 4.15</b> Representasi Satu Buah Kromosom Awal Data pada Seleksi Variabel .....	55
<b>Gambar 4.16</b> Ilustrasi <i>Single Point Crossover</i> pada Seleksi Variabel .....	56
<b>Gambar 4.17</b> Ilustrasi Proses Mutasi pada Seleksi Variabel .....	57
<b>Gambar 4.18</b> Ilustrasi Satu Buah Kromosom .....	63
<b>Gambar 4.19</b> Pindah Silang Dua Buah Kromosom .....	65
<b>Gambar 4.20</b> Ilustrasi Proses Mutasi pada Estimasi Parameter.....	66

## DAFTAR LAMPIRAN

<b>Lampiran 1</b> Data Penelitian.....	85
<b>Lampiran 2</b> Syntax Cross Validation .....	86
<b>Lampiran 3</b> Syntax Tuning Parameter GA.....	87
<b>Lampiran 4</b> Estimasi Parameter SVM.....	88
<b>Lampiran 5</b> Seleksi Variabel GA .....	89
<b>Lampiran 6</b> SVM-GA dengan <i>Fitness AUC</i> .....	91
<b>Lampiran 7</b> SVM-GA dengan <i>Fitness</i> akurasi.....	93

*(Halaman Ini Sengaja Dikosongkan)*

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Penyakit Infeksi Saluran Pernafasan Akut yang paling menjadi perhatian dalam kesehatan masyarakat adalah Pneumonia. Pneumonia merupakan penyakit batuk pilek disertai sesak nafas. Penyakit ini sering menyerang anak balita, namun juga dapat ditemukan pada orang dewasa dan pada orang usia lanjut. Pneumonia adalah infeksi yang menyebabkan paru-paru meradang (Misnadiarly, 2008). Pneumonia menyebabkan *alveoli* pada penderita dipenuhi nanah dan cairan sehingga menyebabkan bernafas menjadi menyakitkan dan oksigen yang masuk ke paru-paru menjadi berkurang. Kekurangan oksigen membuat sel-sel tubuh tidak bisa bekerja. Pneumonia, atau sering disebut radang paru-paru, dapat menimbulkan konsolidasi jaringan paru-paru dan gangguan pertukaran gas setempat serta menimbulkan angka kesakitan yang tinggi dengan gejala batuk, demam dan sesak nafas (Depkes R.I., 2002). Penyebab pneumonia bisa bermacam-macam dan diketahui ada 30 sumber infeksi, dengan sumber utama bakteri, virus, mikroplasma, jamur, berbagai senyawa kimia maupun partikel. Karena penyebab itulah pneumonia dapat mengakibatkan sakit yang parah bahkan sampai berujung kematian.

Di Indonesia, pneumonia merupakan penyebab kematian setelah kardiovaskuler dan tuberkulosis (Misnadiarly, 2008). Dari data *SEAMIC Health Statictics* 2001, influenza dan pneumonia merupakan penyebab kematian nomor 6 di Indonesia. Menurut hasil RIS-KESDAS (2013), prevalensi pneumonia di Indonesia tahun 2013 sebesar 4,5% dengan prevalensi pneumonia di Jawa Timur sebesar 4,2%. Beberapa rumah sakit besar di Indonesia, salah satunya RSUD Dr. Soetomo, mencatat bahwa kematian yang disebabkan pneumonia masih cukup tinggi. RSUD Dr. Soetomo merupakan rumah sakit terbesar dan terlengkap di Jawa Timur sehingga RSUD Dr. Soetomo sering menjadi rujukan beberapa kasus tertentu yang tidak bisa diatasi oleh rumah sakit lainnya, baik dari wilayah Jawa Timur atau provinsi lainnya. Data yang didapatkan dari RSUD Dr. Soetomo menunjukkan bahwa pneumonia komuniti menduduki peringkat keempat dari sepuluh penyakit terbanyak yang dirawat per tahun atau terdapat

sekitar 180 kasus pneumonia komunitas dengan angka kematian antara 20-30%. (PDPI, 2003).

Penyebab penyakit pneumonia sulit ditemukan dan memerlukan waktu beberapa hari untuk mendapatkan hasilnya. Pneumonia dapat menyebabkan kematian bila tidak segera diobati. Selain itu, penyakit pneumonia dapat menular melalui udara sehingga penyebaran penyakit ini cepat. Deteksi dini penderita pneumonia cukup sulit karena gejalanya hampir sama dengan batuk biasa. Selain menentukan deteksi dini, penilaian kelas risiko pneumonia merupakan hal yang penting dalam tatalaksana Pneumonia Komunitas. Perlunya ada penilaian kelas risiko penderita pneumonia adalah supaya penanganannya menjadi cepat dan tepat. Penilaian kelas risiko muncul karena pentingnya penanganan yang tepat dan cepat bagi pasien sehingga membuat munculnya berbagai sistem untuk menentukan tingkat keparahan penderita pneumonia seperti misalnya skor *Pneumonia Severity Index* (PSI) yang dikenal juga dengan nama *Patient Outcome Research Team* (PORT), *Confusion, Urea, Respiratory Rate, Blood pressure, age > 65 years* (CURB-65) yang direkomendasikan oleh *British Thoracic Society* (BTS), *modified ATS* (m-ATS) dan lain sebagainya. Sistem skor PSI dan CURB-65 mempunyai Kinerja yang sama dalam mengelompokkan kelas risiko pneumonia. Namun, kedua sistem skor tersebut belum teruji pada jumlah sampel yang besar. Penelitian lain pernah dilakukan oleh Firdausanti (2017) klasifikasi kelas risiko pneumonia dengan menggunakan metode *hybrid* Analisis Diskriminan dengan seleksi variabel metode *Particle Swarm Optimization* (PSO) mempunyai nilai akurasi yang tinggi dibandingkan metode analisis diskriminan. Oleh karena itu, diperlukan analisis mengenai klasifikasi kelas risiko pneumonia untuk mengetahui klasifikasi kelas risiko pneumonia dan variabel-variabel yang mempengaruhi tingkat keparahan penyakit pneumonia.

Metode *Support Vector Machine* (SVM) adalah salah satu metode yang dapat digunakan untuk melakukan klasifikasi. Dari penelitian yang dilakukan oleh Harafani dan Wahono (2015), SVM dapat mengatasi masalah klasifikasi dan regresi dengan kernel linier ataupun kernel nonlinier yang dapat menjadi satu kemampuan algoritma pembelajaran untuk klasifikasi serta regresi. Kelebihan

metode SVM lainnya adalah SVM dapat digunakan pada data dengan dimensi yang tinggi (Faihah, 2010). Namun, dibalik keunggulannya, SVM juga memiliki kelemahan yaitu sulitnya menentukan nilai parameter yang optimal (Harafani & Wahono, 2015). Dari kelemahan tersebut, diperlukan adanya langkah untuk mengoptimasi parameternya. Karena kesulitan metode SVM dalam menentukan nilai parameter yang optimal, maka dalam penelitian ini dilakukan optimasi menggunakan metode *Genetic Algorithm*. Seiring berkembangnya ilmu pengetahuan, banyak algoritma optimasi yang dapat dikombinasikan dengan metode statistika, misalnya pada penelitian oleh Rahimatin (2017) pada studi kasus penentuan klasifikasi kelas risiko pasien pneumonia menggunakan metode *hybrid* analisis diskriminan linier - *Genetic Algorithm* membuktikan bahwa nilai klasifikasi analisis diskriminan dengan optimasi *Genetic Algorithm* lebih baik daripada analisis diskriminan. Hasil eksperimen oleh Harafani dan Wahono (2015) pada kasus estimasi parameter kebakaran hutan membuktikan bahwa metode SVM (RBF)+GA memiliki nilai akurasi estimasi yang lebih baik dari pada metode regresi lainnya. Penelitian lain dilakukan oleh Kurnianto & Irhamah (2016) mengenai seleksi variabel dan estimasi parameter untuk klasifikasi menyimpulkan bahwa metode *Hybrid* Analisis Diskriminan dan *Genetic Algorithm* memiliki tingkat akurasi yang lebih tinggi dibanding metode analisis diskriminan. Maka dari itu, penelitian ini untuk mengklasifikasikan kelas risiko pneumonia dan variabel-variabel yang mempengaruhi tingkat keparahan penyakit pneumonia menggunakan metode *support vector machine* (SVM) dengan melakukan optimasi parameter menggunakan *Genetic Algorithm*.

Proses seleksi variabel yang berpengaruh terhadap variabel respon dilakukan untuk memilih variabel prediktor mana saja yang digunakan dalam membangun model. Penelitian Kustiyo, Firqiani dan Giri (2008) tentang seleksi variabel menggunakan FCBF pada Algoritma Voting Feature Intervals 5 menyimpulkan bahwa nilai akurasi klasifikasi dengan seleksi variabel lebih baik daripada tanpa seleksi variabel. Penelitian Rusydina (2016) menyimpulkan bahwa seleksi variabel dengan metode FCBF pada *high dimensional* data mempunyai akurasi yang lebih besar dibanding metode seleksi

lainnya. Penelitian sebelumnya oleh Rachmatin dengan kasus klasifikasi kelas risiko pasien pneumonia menggunakan seleksi variabel metode *Genetic Algorithm*. Maka dari itu, dalam penelitian ini seleksi variabel yang digunakan adalah metode *Fast Correlation-Based Filter* (FCBF) dan *Genetic Algorithm* (GA).

Berdasarkan latar belakang yang telah dijelaskan di atas, metode SVM merupakan metode yang baik untuk klasifikasi dengan metode GA sebagai metode untuk mengoptimasi parameter SVM. Maka dari itu, penelitian ini akan menggunakan metode *Grid Search Support Vector Machine* (SVM), algoritma *Genetic Algorithm* (GA) dan *Support Vector Machine-Genetic Algorithm* (SVM-GA) *Hybrid* untuk mengklasifikasi kelas risiko pasien pneumonia di RSUD Dr. Soetomo serta menggunakan metode FCBF dan GA untuk seleksi *feature* yang berpengaruh terhadap kelas risiko pasien pneumonia.

## 1.2 Rumusan Masalah

Pneumonia merupakan penyakit infeksi saluran pernapasan akut (ISPA) yang paling berat karena dapat menyebabkan kematian. Kematian di Indonesia akibat pneumonia 15,5% per 2015, terdapat 554.650 kasus pneumonia. Pengambilan keputusan yang efektif dan efisien diperlukan dalam menangani kasus pneumonia. Penentuan tingkat keparahan penderita pneumonia diperlukan untuk mengetahui diagnosis awal yang tepat sehingga resiko kematian akibat penyakit ini dapat berkurang. Oleh karena itu, permasalahan utama yang akan dibahas dalam penelitian ini adalah bagaimana pengklasifikasian kelas risiko penderita pneumonia menggunakan metode *Support Vector Machine* (SVM) dan *Support Vector Machine-Genetic Algorithm* (SVM-GA) *Hybrid* serta perbandingan akurasi dari kedua metode tersebut.

## 1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang ada, adapun tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut.

1. Mengetahui variabel yang berpengaruh terhadap kelas risiko pasien pneumonia menggunakan metode *Fast Correlation-Based Filter* (FCBF) dan *Genetic Algorithm*.
2. Mengetahui klasifikasi kelas risiko pasien pneumonia dan nilai akurasi dengan menggunakan metode *Support Vector Machine*

(SVM) dan *Support Vector Machine-Genetic Algorithm (SVM-GA) Hybrid*.

3. Mengetahui perbandingan ketepatan klasifikasi antara metode *Support Vector Machine (SVM)* dan *Support Vector Machine-Genetic Algorithm (SVM-GA) Hybrid*.

#### **1.4 Manfaat Penelitian**

Berdasarkan tujuan yang ingin dicapai, adapun manfaat yang diharapkan dari hasil penelitian ini adalah sebagai berikut.

1. Bagi bidang kedokteran, diharapkan penelitian ini dapat memberikan kontribusi dalam klasifikasi kelompok kelas risiko pneumonia sehingga dapat digunakan sebagai alat pengambilan keputusan yang sesuai dalam menangani pasien pneumonia.
2. Bagi dunia pendidikan, penelitian ini dapat menjadi referensi dalam melakukan penelitian selanjutnya mengenai klasifikasi kelas risiko pasien pneumonia maupun metode *Hybrid Support Vector Machine-Genetic Algorithm (SVM-GA)*.

#### **1.5 Batasan Masalah**

Batasan masalah digunakan untuk memfokuskan permasalahan yang akan diselesaikan. Berikut adalah beberapa batasan masalah yang digunakan digunakan dalam penelitian ini

1. Data pasien yang digunakan adalah pasien pneumonia sebagai diagnosis utama
2. Kelas risiko yang digunakan adalah kelas risiko II, III, IV, dan V.

*(Halaman Ini Sengaja Dikosongkan)*

## BAB II TINJAUAN PUSTAKA

### 2.1 *Fast Correlation-Based Filter (FCBF)*

*Fast Correlation-Based Filter (FCBF)* adalah salah satu metode untuk seleksi variabel yang dikembangkan oleh Yu dan Liu (2003). FCBF merupakan metode *feature selection* yang bersifat multivariat dan mengukur kelas variabel dan korelasi antar variabel (Alonso, Noelia, & Veronica, 2015). Menurut algoritma FCBF, suatu variabel yang baik adalah variabel-variabel yang relevan terhadap kelas tapi tidak *redundant* atau berkorelasi tinggi dengan variabel prediktor lainnya terhadap variabel-variabel relevan yang lain.

Keuntungan menggunakan pendekatan ini adalah mudah untuk menghilangkan variabel yang tidak relevan dengan memilih variabel yang korelasinya 0 dan mampu mengurangi *redundant* pada variabel yang sudah dipilih. Ada dua pendekatan untuk mengukur korelasi antara dua variabel acak secara umum. Salah satunya didasarkan pada korelasi linier klasik dan yang lainnya didasarkan pada informasi dasar. Ukuran yang paling terkenal adalah korelasi korelasi linier. Untuk sepasang variabel ( $X, Y$ ), koefisien korelasi linier  $r$  adalah dengan rumus (2.1).

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x}_i)(y_i - \bar{y}_i)}{\sqrt{\sum_{i=1}^n (x_i - \bar{x}_i)^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y}_i)^2}} \quad (2.1)$$

dimana  $n$  adalah jumlah observasi,  $\bar{x}_i$  adalah *mean* dari  $X$ , dan  $\bar{y}_i$  adalah *mean* dari  $Y$ . Nilai  $r$  terletak antara -1 dan 1, inklusif. Jika  $X$  dan  $Y$  benar-benar berkorelasi,  $r$  mengambil nilai 1 atau -1; Jika  $X$  dan  $Y$  benar-benar independen,  $r$  adalah nol.

Ada beberapa manfaat untuk memilih korelasi linier sebagai ukuran kebaikan variabel untuk klasifikasi. Pertama, ini membantu menghilangkan variabel dengan korelasi nol dekat dengan kelas. Kedua, membantu mengurangi korelasi yang tinggi di antara variabel yang dipilih. Diketahui bahwa jika data dapat dipecah secara linier dalam representasi aslinya, masih dapat dipecah secara linier jika semua komponen yang terkait secara vertikal terlepas. Namun, tidak aman untuk selalu mengasumsikan korelasi linier antara variabel di dunia nyata. Pengukuran korelasi linier mungkin tidak mampu

menangkap korelasi yang tidak linier. Keterbatasan lainnya adalah perhitungan tersebut mengharuskan semua variabel mengandung nilai numerik. Untuk mengatasi kekurangan ini, digunakan pendekatan lain dan memilih ukuran korelasi berdasarkan konsep entropi informasi teoritis, ukuran ketidakpastian variabel acak. Dengan jumlah pengamatan sebanyak  $n$ , *Entropy* dari suatu variabel  $X$  didefinisikan seperti persamaan (2.2)

$$H(X) = - \sum_{i=1}^n P(x_i) \log_2(P(x_i)) \quad (2.2)$$

*Entropy* dari variabel  $X$  jika diketahui variabel  $Y$  didefinisikan pada persamaan (2.3)

$$H(X|Y) = - \sum_{i=1}^n P(y_i) \sum_{i=1}^n P(x_i|y_i) \log_2 P(x_i|y_i) \quad (2.3)$$

$P(y_i)$  adalah *prior probabilities* untuk semua nilai  $Y$  dan  $P(x_i|y_i)$  adalah *posterior probabilities* dari  $X$  jika diketahui  $Y$ . Jumlah dimana entropi  $X$  menurun mencerminkan informasi tambahan tentang  $X$  yang disediakan oleh  $Y$  dan disebut *Information Gain* (Quinlan, 1993). Dari *entropy* tersebut dapat diperoleh *Information Gain* seperti persamaan (2.4)

$$IG(X|Y) = H(X) - H(X|Y) \quad (2.4)$$

Pola data yang simetri adalah properti yang diinginkan untuk mengukur korelasi antar variabel. Namun, keuntungan informasi bias mendukung variabel dengan nilai lebih. Selanjutnya, nilai-nilai tersebut harus dinormalisasi untuk memastikannya sebanding dan memiliki nilai yang sama. Untuk mengukur korelasi antar variabel digunakan *symmetrical uncertainty*. Persamaan *symmetrical uncertainty* adalah pada persamaan (2.5)

$$SU(X, Y) = 2 \frac{IG(X|Y)}{H(X) + H(Y)} \quad (2.5)$$

Nilai *symmetrical uncertainty* adalah pada rentang 0 sampai 1. Nilai 1 yang menunjukkan bahwa variabel  $X$  dan  $Y$  dependen serta nilai 0 yang menunjukkan bahwa  $X$  dan  $Y$  adalah independen. Langkah berbasis entropi memerlukan variabel nominal, namun metode ini juga dapat mengukur korelasi antara variabel kontinu juga, jika nilainya didiskritisasi dengan benar sebelumnya. Oleh karena itu, kita menggunakan ketidakpastian simetris dalam pekerjaan ini.

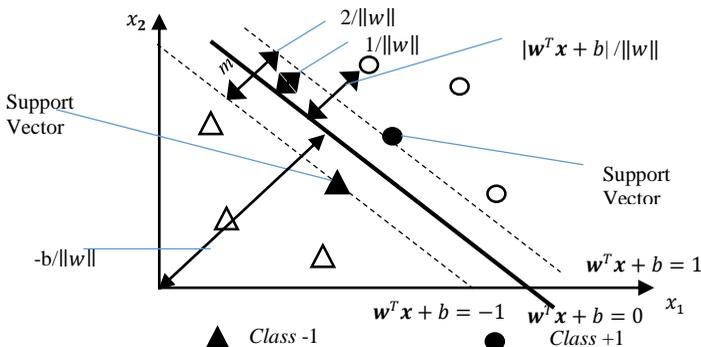
## 2.2 Support Vector Machine (SVM)

*Support Vector Machine* (SVM) pertama kali diperkenalkan oleh Vapnik tahun 1992. SVM merupakan suatu teknik untuk menemukan fungsi pemisah dalam pengklasifikasian yang dapat memisahkan dua atau lebih kelompok data yang berbeda. SVM merupakan salah satu kelas dalam *Neural Network*. Metode *neural network* bertujuan untuk menemukan solusi beberapa *local optimal* dan SVM untuk menemukan solusi *global optimal*. Metode SVM merupakan metode *machine learning* yang bekerja atas prinsip *Structural Risk Minimization* (SRM) dengan tujuan menemukan *hyperplane* terbaik yang memisahkan dua buah kelas pada *input space*. SVM dapat menemukan fungsi pemisah (*hyperplane*) terbaik diantara fungsi yang tidak terbatas jumlahnya untuk memisahkan obyek. *Hyperplane* terbaik terletak tepat di tengah antara dua set obyek dari dua kelas. Mencari *hyperplane* terbaik ekuivalen dengan memaksimalkan margin atau jarak antara dua set obyek dari dua kelas berbea. SVM bekerja untuk menemukan suatu fungsi pemisah dengan margin yang maksimal (Vapnik, 1999). SVM merupakan teknik klasifikasi dengan proses pelatihan (*supervised learning*).

### 2.1.1 Klasifikasi Linear Separable Case SVM

Klasifikasi Linier *separable* data adalah penerapan metode SVM pada data yang dapat dipisahkan secara linier.

Misal dengan data dengan dimensi  $m$ ,  $\mathbf{x}_i = \{\mathbf{x}_i, \mathbf{x}_{i+1}, \dots, \mathbf{x}_m\}$  adalah *dataset* dan  $y_i = \{+1, -1\}$  adalah label kategori untuk *dataset*. Ilustrasi linier *separable case* ditunjukkan oleh gambar 2.1.



**Gambar 2.1** Hyperplane pada SVM Linear Separable (Kusumaningrum, 2017)

Pada Gambar 2.1, garis putus-putus merupakan bidang membatasi kelas -1 dan kelas +1. Sedangkan garis tegas lurus diantara garis putus-putus merupakan bidang pemisah antara kelas +1 dan kelas -1. Garis pemisah tersebut merupakan *hyperplane* terbaik antara kedua kelas dapat ditemukan dengan mengukur *margin hyperplane* tersebut dan mencari titik maksimalnya. *Hyperplane* yang terbaik dilihat dari memaksimalkan nilai margin dan akan melewati pertengahan antara kedua kelas. *Hyperplane* sampel yang lokasinya paling dekat dengan *hyperplane* disebut *support vector*, dengan proses dalam SVM adalah mencari *support vector* untuk memperoleh *hyperplane* yang terbaik.

Pemisahan secara linier memisahkan data dari kelas yang berbeda. Diberikan data berdimensi  $m$ , data yang diamati sebanyak  $n$ , yaitu  $\mathbf{x}_i$  dengan  $i = 1, 2, \dots, n$  yang termasuk dalam kelas pertama atau kedua sehingga fungsi pemisahannya adalah (Abe, 2010)

$$D(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \quad (2.6)$$

dimana  $\mathbf{w}$  adalah vektor berdimensi  $m$  dan  $b$  adalah suatu konstanta yang disebut bias. Data  $\mathbf{x}_i$  termasuk dalam 1 ( $y_i = 1$ ) bila  $\mathbf{w}^T \mathbf{x}_i + b > 0$  dan data  $\mathbf{x}_i$  termasuk dalam -1 ( $y_i = -1$ ) bila  $\mathbf{w}^T \mathbf{x}_i + b < 0$ . Karena data tersebut dipisahkan secara linier maka persamaan *hyperplane* dapat ditulis seperti persamaan (2.7)

$$\mathbf{w}^T \mathbf{x}_i + b = 0 \quad (2.7)$$

oleh karena itu, pemisahan kelas dilakukan dengan mempertimbangkan persamaan (2.8)

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, i = 1, 2, \dots, n. \quad (2.8)$$

Nilai margin merupakan nilai jarak terdekat *hyperplane* dengan data yang paling dekat dengan *hyperplane* tiap kelas. Jarak antara data  $\mathbf{x}$  pada tiap kelas dengan *hyperplane* adalah pada persamaan (2.9)

$$d(\mathbf{w}, b; \mathbf{x}) = \frac{|\mathbf{w}^T \mathbf{x}_i + b|}{\|\mathbf{w}\|} \quad (2.9)$$

Dengan mengalikan  $b$  dan  $w$  dengan sebuah konstanta, akan dihasilkan nilai margin yang dikalikan dengan konstanta yang sama (Gunn, 1998). Jarak terdekat antara data  $\mathbf{x}$  dengan *hyperplane* pada kelas 1 dan 2 masing-masing adalah  $1/\|\mathbf{w}\|$  sehingga nilai margin antara bidang pembatas (berdasarkan rumus jarak garis ke titik pusat) yaitu pada persamaan (2.10)

$$\frac{2}{\|\mathbf{w}\|} \quad (2.10)$$

*Hyperplane* yang optimum diperoleh dengan memaksimalkan nilai margin. Nilai margin akan maksimal bila nilai  $\|\mathbf{w}\|$  minimal. Maka, pencarian bidang pemisah terbaik dengan nilai margin terbesar dapat dirumuskan menjadi masalah optimasi konstrain yaitu

$$\min \frac{1}{2} \|\mathbf{w}\|^2 \quad (2.11)$$

Solusi dari permasalahan persamaan kuadratik dengan fungsi batasan berupa pertidaksamaan tersebut dapat diperoleh dengan menggunakan *Lagrange Multiplier*, didapatkan persamaan (2.12).

$$L_d = \sum_{m=1}^n \alpha_m - \frac{1}{2} \sum_{m_1=1}^n \sum_{m_2=1}^n \alpha_{m_1} \alpha_{m_2} y_{m_1} y_{m_2} (\mathbf{x}_{m_1}^T \mathbf{x}_{m_2}) \quad (2.12)$$

Persamaan  $L_d$  didapatkan dengan mensubstitusikan nilai  $y_{m_1}$ ,  $y_{m_2}$ ,  $X_{m_1}$ , dan  $X_{m_2}$  ke persamaan (2.7). Persamaan  $L_d$  digunakan untuk mencari nilai-nilai  $\alpha_m$  (*support vector*) dengan membuat  $L_d$  optimum.  $L_d$  optimum didapat dengan cara mencari turunan parsial  $L_d$  terhadap  $\alpha$ . Setelah mendapatkan nilai  $\alpha$ , langkah selanjutnya adalah mencari nilai  $w$  dan  $b$  dengan persamaan sebagai berikut.

$$\mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i \quad \text{dan} \quad b = 1 - \mathbf{w}^T \mathbf{x} \quad (2.13)$$

### 2.1.2 Klasifikasi *Linear Nonseparable Case SVM*

Untuk mengatasi adanya misklasifikasi, formulasi yang telah dijelaskan sebelumnya akan diperluas sehingga dapat digunakan data *non-separable*. Masalah optimasi fungsi obyektif maupun kendala dimodifikasi dengan mengikuti *slack variable*  $\xi < 0$  yang merupakan sebuah ukuran kesalahan ini klasifikasi. Teknik *soft margin* dilakukan dengan memodifikasi persamaan (2.8) dengan memasukkan variabel *slack* pada persamaan tersebut (Ben-Hur & Weston, 2010). Berikut merupakan *constraint* yang sudah dimodifikasi untuk kasus *non-separable*.

$$y_i [(\mathbf{w}^T \mathbf{x}_i) + b] \geq 1 - \xi \quad (2.14)$$

untuk  $i = 1, 2, \dots, n$ . Variabel  $\xi$  merupakan variabel *slack* yang memungkinkan suatu data berada dalam margin ( $0 < \xi_i < 1$ ) yang disebut *margin error* atau misklasifikasi ( $\xi_i < 0$ ). *Hyperplane* atau

pemisah yang optimal ditentukan dengan vektor  $w$ , yaitu dengan meminimumkan fungsi

$$\Phi(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad (2.15)$$

dimana  $i = 1, 2, \dots, n$ ,  $\xi = (\xi_1, \dots, \xi_n)^T$ , dan  $C$  adalah parameter yang ditentukan. Nilai  $C$  merupakan parameter regulasi yang digunakan untuk mengontrol hubungan antara *slack variabel* dengan  $\|w\|^2$ . Bentuk dual dari masalah Lagrange adalah :

$$\max_{i=1} w(a) = \max \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i x_j) \quad (2.16)$$

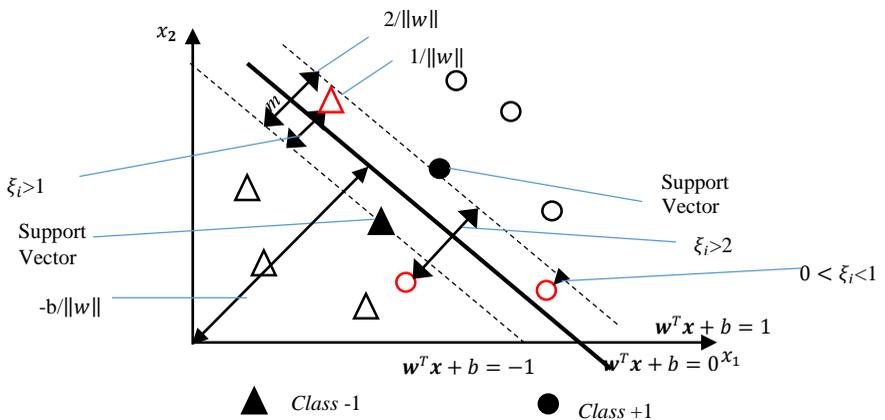
Solusi dari persamaan (2.16) adalah sebagai berikut

$$\arg \min \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i x_j) - \sum_{i=1}^n \alpha_i \quad (2.17)$$

dengan kendala pada *Lagrange Multiplier* persamaan (2.17) dihasilkan persamaan untuk mendapatkan  $w$  seperti persamaan berikut.

$$w = \sum_{i=1}^n \alpha_i y_i x_i \text{ dan } \hat{b} = -\frac{1}{2} w(x_r + x_s) \quad (2.18)$$

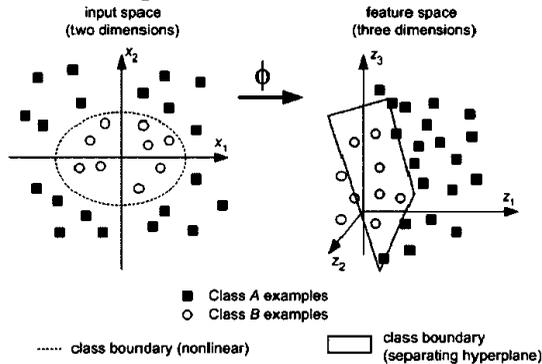
dimana  $x_r$  dan  $x_s$  adalah *support vector* dari tiap kelas, sehingga setelah mendapatkan nilai parameter tersebut, *hyperplane* dapat ditemukan (Cortes & Vapnik, 1995)



**Gambar 2.2** Konsep *Hyperplane* pada SVM linear nonseparable (Kusumaningrum, 2017)

### 2.1.3 Klasifikasi *Non-Linear* SVM

Untuk menyelesaikan problem *non-linear*, SVM dimodifikasi dengan menggunakan fungsi *Kernel*. Dalam *non-linear* SVM, data dipetakan oleh fungsi  $\Phi(\mathbf{x})$  ke ruang vektor yang berdimensi lebih tinggi. Pada ruang vektor yang baru ini, *hyperplane* yang memisahkan kedua kelas tersebut dapat dikonstruksikan.



(Sumber: Lessmann, 2004)

**Gambar 2.3** *Hyperplane* pada *Non-Linear* SVM

Pada ilustrasi Gambar 2.3 bentuk pemetaan *non-linear* tidak perlu diketahui secara eksplisit. Pemetaan pada vector input  $\mathbf{x}$  berdimensi  $m$  ke *feature space* berdimensi  $l$  dilakukan dengan fungsi pemetaan  $\mathbf{x}$  yaitu  $\phi(\mathbf{x}) = (\phi_1(\mathbf{x}), \dots, \phi_l(\mathbf{x}))^T$  berupa fungsi kernel yang ditentukan peneliti. Fungsi objektif SVM pada  $\Phi(\mathbf{x})$  dalam *feature space*  $F$  dengan mengganti *kernel linear*  $\mathbf{x}\mathbf{x}^T$  dengan *kernel non-linear*  $K(\mathbf{x}, \mathbf{x}^T) \in R^{m \times m}$ .

Sehingga berdasarkan persamaan (2.16) formulasi hasil dari *problem* menjadi

$$\max \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i \mathbf{x}_j) \quad (2.19)$$

dengan syarat  $\sum_{i=1}^l \alpha_i y_i = 0$  dengan mendapatkan nilai  $\alpha_i$  prediksi dapat dilakukan dengan rumus

$$f(\mathbf{x}) = \text{sign} \left( \sum_{i=1}^n \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + \hat{\mathbf{b}} \right) \quad (2.20)$$

dimana matriks  $K(\mathbf{x}_i, \mathbf{x})$  merupakan matriks *kernel*. Beberapa fungsi pembentuk matriks *kernel* untuk menyelesaikan permasalahan *nonlinear* adalah sebagai berikut.

- a. *Kernel Gaussian (RBF)*

$$K(\mathbf{x}_i, \mathbf{x}) = \exp(-\gamma(\|\mathbf{x} - \mathbf{x}_i\|^2)) \quad (2.21)$$

- b. *Kernel Linear*

$$K(\mathbf{x}_i, \mathbf{x}) \quad (2.22)$$

- c. *Kernel Polinomial*

$$K(\mathbf{x}_i, \mathbf{x}) = (\mathbf{x}, \mathbf{x}^T + d)^d \quad (2.23)$$

Ber macam-macam jenis *kernel* lain dapat digunakan untuk pemetaan pada *feature space* namun ketiga *kernel* tersebut yang paling umum digunakan (Santosa, 2007)

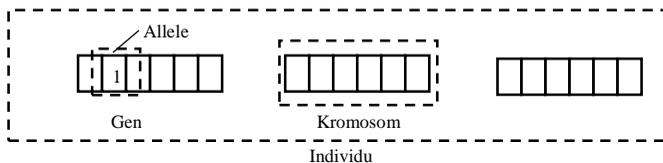
### 2.3 *Genetic Algorithm (GA)*

*Genetic Algorithm* ditemukan pada tahun 1975 di Universitas Michigan oleh John Holland dan murid-muridnya. Konsep *Genetic Algorithm* diadopsi dari teori evolusi oleh Darwin. *Genetic Algorithm* diilhami oleh ilmu genetika sehingga istilah-istilah yang digunakan pada metode ini diambil dari ilmu genetika (Setiawan, 2003). Pada proses evolusi akan didapatkan individu-individu harus mampu menyesuaikan diri dengan lingkungannya untuk bertahan hidup. Individu-individu ini telah secara berulang kali mengalami perubahan gen untuk dapat menyesuaikan diri dengan lingkungan hidupnya. Perubahan gen terjadi melalui proses perkembangbiakan, dalam *Genetic Algorithm*, proses perkembangbiakan ini merupakan dasar pemikiran dalam mendapatkan anak yang lebih sesuai. Konsep yang dibuat John Holland adalah ketahanan, dan juga keseimbangan antara efisiensi dan kemenangan sistem untuk bertahan hidup dalam setiap kondisi (Gunawan, 2012).

Salah satu kelebihan *Genetic Algorithm* yaitu algoritma ini mampu mengatasi berbagai jenis fungsi obyektif dari berbagai kromosom. *Genetic Algorithm* juga adaptif dan mudah dikombinasikan dengan metode lain (Gen & Cheng, 1997). Kelebihan lain dari *Genetic Algorithm* dibandingkan algoritma optimasi lainnya, yaitu kemampuan untuk menangani permasalahan kompleks dan parallel. *Genetic Algorithm* bisa digunakan untuk jumlah variabel

yang besar. *Genetic Algorithm* dapat menangani berbagai macam optimasi tergantung pada fungsi objektifnya (*fitness*), seimbang atau tidak seimbang, linier atau tidak linier, berkesinambungan atau tak berkesinambungan, atau dengan *random noise* (Zhang, 2011).

Konsep *Genetic Algorithm* yang didasarkan pada ilmu genetik menyebabkan istilah-istilah yang digunakan dalam *Genetic Algorithm* banyak diadaptasi dari ilmu tersebut. Gambar 2.4 merupakan ilustrasi istilah-istilah yang digunakan dalam *Genetic Algorithm*.



**Gambar 2.4** Istilah dalam *Genetic Algorithm*

Gen adalah nilai yang menyatakan satuan dasar yang membentuk suatu arti tertentu dalam satu kesatuan gen yang dinamakan kromosom. Nilai yang terkandung dalam gen disebut *allele*. Tipe data *allele* bisa berupa biner, *floating point*, atau integer tergantung representasi *genetic* yang digunakan. Sementara gabungan *allele* bisa memberi nilai pada kromosom yang disebut fenotip (Gunawan, 2012). Kromosom adalah gabungan gen-gen yang membentuk nilai tertentu. Individu menyatakan satu nilai atau keadaan yang menyatakan salah satu solusi yang mungkin dari permasalahan yang diangkat. Pada kasus ini, setiap individu terdiri dari satu kromosom.

Jumlah gen pada setiap kromosom pada GA untuk seleksi variabel adalah sebanyak jumlah variabel prediktor pada penelitian. *Allele* pada seleksi variabel adalah nilai biner yaitu 0 dan 1, dengan variabel yang tidak terpilih bernilai 0 dan variabel yang terpilih bernilai 1. Pada GA untuk optimasi parameter, jumlah gen pada satu kromosom adalah sebanyak parameter pada model yang akan dioptimasi. Pada klasifikasi menggunakan SVM dengan fungsi aktivasi Kernel *Gaussian*, parameter yang digunakan ada 2 yaitu  $C$  dan  $\gamma$ . *Allele* pada optimasi parameter adalah nilai *float* sehingga nilai optimum pada *allele* yang sudah dioptimasi menggunakan GA merupakan nilai parameter yang paling optimum untuk klasifikasi.

Prosedur yang harus terlebih dahulu ditetapkan (inisialisasi) sebelum melakukan analisis menggunakan *Genetic Algorithm* adalah sebagai berikut (Setiawan, 2003).

1. Menentukan Representasi Genetik. Bentuk kromosom harus ditentukan paling awal karena setiap *allele* yang bergabung harus disediakan tempat penampungannya. Pada kasus seleksi variabel, panjang kromosom sesuai dengan banyaknya variabel prediktor pada klasifikasi. Panjang kromosom dari program yang dibuat harus sesuai dengan banyaknya parameter yang ingin dioptimasi (pada masalah optimasi estimasi parameter model).
2. Menentukan Rumusan Fungsi *Fitness*. Fungsi *fitness* digunakan untuk mengetahui seberapa baik solusi yang diperoleh satu individu. Karena fungsi *fitness* dalam penelitian ini adalah ketepatan klasifikasi, jika semakin tinggi nilai *fitness* yang dihasilkan maka individu itu semakin baik. Melalui nilai *fitness* juga dapat dicari probabilitas kumulatif, yang berguna dalam pemilihan individu sebagai *parent*. Salah satu permasalahan yang sering terjadi adalah terdapat beberapa kromosom yang mendominasi populasi sehingga dapat mengakibatkan kondisi konvergen yang terlalu dini (*premature convergence*) dan jika kromosom tersebut mempunyai nilai *fitness* yang tinggi tetapi tidak optimal dapat mengakibatkan proses terjebak ke kondisi yang bersifat lokal (*local optimal*).
3. Menetapkan Nilai Parameter. Beberapa parameter yang perlu diberi nilai, antara lain sebagai berikut.
  - a. Ukuran populasi  
Ukuran populasi adalah banyaknya individu yang terdapat di dalam populasi. Populasi merupakan sekumpulan individu yang akan diproses bersama dalam satu siklus proses iterasi. Semakin banyak dan beragamnya individu yang terdapat di dalam populasi akan memberikan peluang yang lebih besar untuk menemukan individu yang mendekati sempurna.
  - b. Jumlah generasi  
Generasi pada GA merupakan satuan siklus dalam iterasi. Semakin besar jumlah generasi maka individu yang dihasilkan selalu baik dan sempurna. Tetapi, tidak berarti semakin besar jumlah generasi maka individu yang dihasilkan

selalu lebih baik karena ada suatu saat dimana nilai *fitness* semua individu akan menjadi sama. Jika hal itu terjadi maka generasi-generasi selanjutnya akan cenderung mempunyai nilai *fitness* yang sama dengan generasi sebelumnya, walaupun ada perubahan, perubahan tersebut tidak terlalu besar.

- c. Probabilitas pindah silang ( $P_c$ )  
Semakin besar probabilitas pindah silang maka semakin banyak kemungkinan terjadi perkawinan silang antara dua individu yang akan menghasilkan individu-individu baru.
- d. Probabilitas mutasi ( $P_m$ )  
Semakin besar probabilitas mutasi akan memperbesar kemungkinan operasi mutasi suatu kromosom yang akan menghasilkan individu-individu baru.
- e. Probabilitas reproduksi  
Probabilitas tersebut menentukan individu-individu yang pantas untuk bertahan dalam generasi mendatang dikarenakan *offspring* yang dihasilkan dengan *parent*.

Pada metode *Genetic Algorithm*, terdapat 7 komponen yang membangun metode ini (Suyanto, 2005) dengan rincian sebagai berikut.

1. Skema Pengkodean. Skema pengkodean yang paling umum digunakan dalam pengkodean kromosom antara lain:
  - i. *Binary Encoding*: tiap gen hanya bisa bernilai 0 atau 1. Pada penelitian ini, pengkodean *Binary Encoding* dipakai untuk seleksi variabel.
  - ii. *Real number encoding*: nilai gen berada dalam interval  $[0, R]$ , dengan  $R$  adalah bilangan real positif dan biasanya  $R=1$ . Pada penelitian ini, pengkodean *Real number encoding* dipakai untuk optimasi parameter.
  - iii. *Discrete encoding*: nilai gen berada dalam interval bilangan bulat  $[0, 9]$ .
  - iv. *Value encoding*: nilai gen yang berasal dari sembarang nilai yang sesuai dengan permasalahan (bilangan bulat maupun riil).
2. Nilai *Fitness*. Nilai *fitness* adalah ukuran performansi dari satu individu yang akan bertahan hidup. Di dalam evolusi alam,

individu yang memiliki nilai *fitness* tinggi akan tetap diikuti dalam iterasi berikutnya dan sebaliknya individu yang memiliki nilai *fitness* rendah tidak dapat diikuti dalam iterasi berikutnya. Fungsi *fitness* yang digunakan dalam penelitian ini adalah nilai ketepatan klasifikasi yaitu *AUC* dan akurasi, sehingga individu yang akan bertahan ke generasi selanjutnya adalah individu yang memiliki nilai *fitness* tertinggi.

3. Seleksi Orang Tua. Seleksi orang tua bertujuan untuk memberikan kesempatan reproduksi bagi anggota populasi yang memiliki *fitness* rendah. Nilai *fitness* setiap observasi akan dibandingkan dengan bilangan random yang telah dibangkitkan. Metode seleksi yang dipakai adalah *roulette wheel selection*. Pada metode ini, masing-masing kromosom menempati potongan lingkaran pada roda *roulette* secara proporsional sesuai dengan nilai *fitness*-nya. Keuntungan menggunakan metode ini adalah semua kromosom memiliki kesempatan untuk dipilih. Tahapan RWS adalah sebagai berikut (Jadaan, Rajamani, & Rao, 2005-2008).

- i. Menghitung nilai *fitness* masing-masing kromosom dengan ukuran populasi sebesar  $n$ .

- a. Kromosom untuk seleksi variabel.

Kromosom ke-	$X_1$	$X_2$	...	$X_{13}$	Nilai Fitness
1	1	0	...	1	...
...	...	...	...	...	...
$n$	...	...	...	...	...

- b. Kromosom untuk optimasi parameter.

Kromosom ke-	$c$	$\gamma$	Nilai Fitness
1	...	...	...
...	...	...	...
$n$	...	...	...

- ii. Menghitung total nilai *fitness* dalam populasi.

$$F = \sum_{h=1}^{n_{pop}} f(h) \quad (2.24)$$

- iii. Menghitung proporsi masing-masing kromosom.

$$P_h = \frac{f(h)}{F} \quad (2.25)$$

- iv. Menghitung nilai kumulatif proporsi untuk masing-masing kromosom

$$S_h = \sum_{q=1}^h P_q \quad (2.26)$$

- v. Membangkitkan sebuah  $r$  angka dengan range  $[0,1]$ .  
 vi. Jika  $r \leq S_1$ , maka pilih kromosom  $v_1$ , lainnya pilih  $v_h$ , sehingga  $S_{h-1} < r \leq S_h$ .  
 vii. Mengulangi tahapan v hingga semua kromosom yang berjumlah  $N$  terpilih semuanya.

Keterangan:

$F$  : total nilai *fitness* semua kromosom dalam populasi

$N$  : jumlah kromosom dalam satu populasi

$f(h)$  : nilai *fitness* kromosom ke- $h$

$P_h$  : nilai proporsi *fitness* kromosom ke- $h$

$S_h$  : nilai *fitness* kumulatif kromosom ke- $h$

$R$  : sebuah bilangan random

$v_i$  : kromosom ke- $i$

4. Pindah silang (*Crossover*). Proses pindah silang merupakan satu proses yang terjadi pada dua kromosom yang bertujuan untuk menambah keanekaragaman kromosom dalam satu populasi dengan penyilangan antar kromosom yang diperoleh dari proses reproduksi sebelumnya. Macam-macam proses pindah silang diantaranya yaitu pindah silang satu titik, dua titik, dan seragam. Salah satu contoh proses pindah silang satu titik potong sebagai berikut.

Orang tua 1 : [0 1 0 1 1 1 0 0]

Orang tua 2 : [1 0 1 0 0 0 1 1]

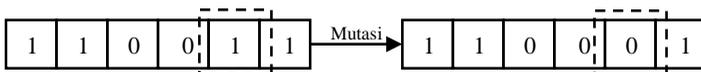
Apabila dilakukan pindah silang pada titik ke-4 maka anak yang akan dihasilkan adalah sebagai berikut.

Anak 1 : [0 1 0 1 0 0 1 1]

Anak 2 : [1 0 1 0 1 1 0 0]

Pindah silang dilakukan dengan suatu nilai probabilitas tertentu. Nilai probabilitas pindah silang merupakan seberapa sering proses pindah silang akan terjadi antara dua kromosom orang tua. Berdasarkan hasil penelitian *Genetic Algorithm* yang sudah pernah dilakukan sebaiknya nilai probabilitas pindah silang tinggi, yaitu antara 0,8-0,9 agar memberikan hasil yang baik.

5. Mutasi. Mutasi digunakan untuk mencegah algoritma yang terjebak pada solusi optimum dan melakukan tugasnya untuk mengembalikan atau membenahi material genetika yang hilang karena informasi acak genetika yang mengganggu. Proses mutasi cukup sederhana, jika bilangan random yang dibangkitkan kurang dari peluang mutasi yang ditentukan maka gen tersebut akan diubah menjadi kebalikannya (Sivanandam & Deepa, 2008). Nilai probabilitas mutasi menyatakan seberapa sering gen dalam kromosom akan mengalami mutasi. Proses mutasi ini bersifat acak sehingga tidak menjamin akan diperoleh kromosom dengan *fitness* yang lebih baik setelah terjadinya mutasi tersebut. Solusi yang optimum (konvergensi dini) dapat terjadi apabila proses pencarian solusi terperangkap dalam salah satu ruang pencarian kromosom dengan *fitness* yang rendah yang terus bertahan. Hal ini mengakibatkan tidak mempunyai mengeksplorasi bagian-bagian yang lain. Oleh karena itu diperlukan operator mutasi untuk menjaga perbedaan kromosom dalam populasi. Operasi mutasi juga ditentukan dengan probabilitas mutasi ( $P_m$ ) yang biasanya bernilai kecil, antara 0,001 sampai dengan 0,01 untuk memastikan agar solusi terbaik tidak menyimpang. Ada beberapa jenis mutasi yang telah berkembang, salah satunya adalah mutasi acak yang terjadi ketika beberapa posisi potongan terpilih secara acak dan nilai potongan tersebut akan menjadi negasinya. Ilustrasi proses mutasi yang terjadi pada posisi ke-35 dapat dilihat pada Gambar 2.5.



Gambar 2.5 Ilustrasi Proses Mutasi

6. Elitisme. Proses elitisme adalah suatu proses pengopian individu agar individu yang memiliki *fitness* tertinggi tidak hilang selama proses evolusi. Elitisme mengganti kromosom yang memiliki kualitas buruk pada populasi baru dengan kromosom terbaik pada populasi orang tua, jumlah kromosom yang diganti sebesar 10%-20% dari jumlah populasi. Tahapan ini dapat mempercepat iterasi *Genetic Algorithm* karena

konvergensi cepat tercapai. Hal ini dikarenakan individu yang memiliki *fitness* terendah tidak selalu terpilih karena proses seleksi dilakukan secara *random* (Jadaan, Rajamani, & Rao, 2005-2008).

7. Penggantian Populasi. Penggantian populasi yang berarti individu dalam satu populasi dari suatu generasi digantikan sekaligus oleh  $N$  individu baru hasil pindah silang dan mutasi. Presentasi populasi yang digantikan dalam tiap generasi dinyatakan dalam  $G$ . Nilai  $G = 1$  pada skema penggantian populasi dan untuk  $G = 1/N$  merupakan skema penggantian yang paling ekstrem dimana hanya mengganti satu individu pada tiap generasi. Dalam setiap generasi sejumlah  $NG$  individu harus dihapus agar ukuran populasi tetap  $N$ . Terdapat beberapa prosedur penghapusan individu ini seperti penghapusan individu yang paling tua atau individu yang memiliki nilai *fitness* yang paling rendah. Ada kemungkinan penghapusan individu dilakukan pada semua individu dalam populasi tersebut.

## 2.4 Evaluasi Kinerja Klasifikasi

Nilai akurasi hasil klasifikasi dapat dihitung menggunakan *confusion matrix*. *Confusion matrix* adalah alat yang digunakan untuk menganalisis seberapa baik *classifier* mengenali data dari kelas yang berbeda.

**Tabel 2.1** Tabel Klasifikasi

Aktual	Prediksi			
	Kelas 2	Kelas 3	Kelas 4	Kelas 5
Kelas 2	$n_{22}$	$n_{23}$	$n_{24}$	$n_{25}$
Kelas 3	$n_{32}$	$n_{33}$	$n_{34}$	$n_{35}$
Kelas 4	$n_{42}$	$n_{43}$	$n_{44}$	$n_{45}$
Kelas 5	$n_{52}$	$n_{53}$	$n_{54}$	$n_{55}$

$n_{ij}$  merupakan jumlah prediksi kelas ke- $j$  yang tepat diklasifikasikan ke kelas ke- $i$ . Dengan menggunakan *confusion matrix*, beberapa pengukuran Kinerja klasifikasi dapat dilakukan, yaitu akurasi, sensitifitas, spesifitas, dan presisi. Akurasi menunjukkan

efektifitas *classifier* secara menyeluruh. Semakin besar nilai akurasi, maka kinerja *classifier* semakin baik. Sensifitas digunakan untuk mengukur efektifitas sebuah *classifier* unuk mengidentifikasi kelas positif, sedangkan spesifikasi mengukur efektifitas *classifier* dalam mengidentifikasi kelas 22ender22a. Perhitungan akurasi klasikasi dengan *confusion matrix* dapat digunakan untuk komposisi data yang *balance*. Akurasi klasifikasi pada data *unbalance* dapat dihitung menggunakan *geometric mean* ( $G\_mean$ ) dan *Area Under ROC Curve* ( $AUC$ ).  $G\_mean$  untuk kasus mengelompokkan dengan jumlah kategori lebih dari dua dapat menggunakan *Expanding  $G\_mean$*  dengan rumus seperti persamaan (2.27) dan (2.28) dengan  $I$  adalah jumlah kelas klasifikasi yang terbentuk (Bekkar, Djemaa & Alitouch, 2013).

$$G\_mean = \left( \prod_{i=1}^I R_i \right)^{\frac{1}{I}} \quad (2.27)$$

$$AUC = \frac{1}{I} \sum_{i=1}^I R_i \quad (2.28)$$

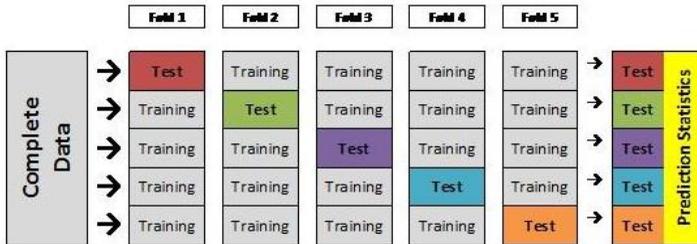
dimana

$$R_i = \frac{n_{ii}}{\sum_{l=1}^I n_{il}}, i = 1, 2, \dots, I \quad (2.29)$$

## 2.5 K-Fold Cross Validation

Metode validasi dengan *k-Fold Cross Validation* cocok digunakan dalam kasus yang mempunyai jumlah sampel terbatas. Dalam *k-Fold Cross Validation*, data ( $D$ ) dibagi ke dalam *subset* sejumlah  $k$  dengan ukuran masing-masing kelompok data  $D_1, D_2, \dots, D_k$  sama. Data yang digunakan untuk *training* sebanyak  $k-1$  *subset* yang dikombinasikan secara bersama-sama dan kemudian diaplikasikan untuk sisa *subset* sebagai data *testing*. Proses tersebut dilakukan sebanyak  $k$  kali sehingga masing-masing *subset* data menjadi data *testing*. Hasil akurasi adalah rata-rata dari hasil perhitungan akurasi setiap data *training* dan *testing*. Jumlah  $k$  yang

biasa digunakan adalah 3, 5, 10, dan 20 (Alonso, Noelia, & Veronica, 2015). Pada penelitian ini digunakan 5 *fold* karena data pada kelas pasien ke-5 hanya ada sedikit pengamatan yaitu hanya sebanyak 8. Gambar 2.6 merupakan ilustrasi untuk 5-fold *Cross Validation*.



Gambar 2.6 Ilustrasi Proses 5-Fold *Cross-Validation*

## 2.6 Pneumonia

Pneumonia merupakan peradangan yang mengenai *parenkim* paru, dari *bronkiolus terminalis* yang mencakup *bronkiolus respiratorius* dan *alveoli*, serta menimbulkan konsolidasi jaringan paru-paru dan gangguan pertukaran gas setempat (Sudoyo, Setiyohadi, Alwi, Simadibrata, & Setiati, 2009). Secara klinis, pneumonia didefinisikan sebagai suatu peradangan paru-paru yang disebabkan oleh mikroorganisme (bakteri, virus, jamur, parasit). Peradangan paru-paru yang disebabkan oleh non-mikroorganisme, misalnya bahan kimia, radiasi, aspirasi bahan *toxic*, disebut *pneumonitis* (PDPI, 2003).

Pneumonia atau sering disebut dengan paru-paru basah adalah infeksi yang memicu inflamasi pada kantong-kantong udara di salah satu atau kedua paru-paru. Kantong-kantong udara kecil di ujung saluran pernapasan dalam paru-paru akan membengkak dan dipenuhi cairan. Proses peradangan akan menyebabkan jaringan paru yang berupa *alveoli* (kantong udara) dapat dipenuhi cairan ataupun nanah. Akibat dari adanya cairan pada *alveoli* ini menyebabkan kemampuan paru-paru sebagai tempat pertukaran gas, terutama oksigen, akan terganggu. Kekurangan oksigen membuat sel-sel tubuh tidak bisa bekerja.

Pneumonia diklasifikasikan berdasarkan beberapa aspek yaitu berdasarkan klinis dan epidemiologi, bakteri penyebab, dan prediksi

infeksi (Rasyid & Kamso, 2004). Berdasarkan klinis dan epidemiologis, penyakit pneumonia diklasifikasikan menjadi 4 kelompok.

- a. Pneumonia komunitas (*community-acquired pneumonia*), yaitu infeksi pneumonia yang terjadi di masyarakat,
- b. Pneumonia nosokomial, (*hospital-acquired pneumonia / nosocomial pneumonia*), yaitu pneumonia yang infeksi didapat di rumah sakit,
- c. Pneumonia aspirasi, adalah pneumonia yang disebabkan oleh terhirupnya bahan-bahan ke dalam saluran pernafasan,
- d. Pneumonia pada penderita *immunocompromised*.

Berdasarkan bakteri penyebab, penyakit pneumonia diklasifikasikan menjadi 3 kelompok.

- a. Pneumonia bakteri/tipikal, dapat terjadi pada semua usia. Pneumonia bakterial sering diistilahkan dengan pneumonia akibat kuman. Para peminum alkohol, pasien yang terkebelakang mental, pasien pascaoperasi, orang yang menderita penyakit pernapasan lain atau infeksi virus adalah yang mempunyai sistem kekebalan tubuh rendah dan menjadi sangat rentan terhadap penyakit itu. Pada saat pertahanan tubuh menurun, misalnya karena penyakit, usia lanjut, dan malnutrisi, bakteri pneumonia akan dengan cepat berkembang biak dan merusak paru-paru. Bakteri *Pneumokokus* adalah kuman yang paling umum sebagai penyebab pneumonia bakterial tersebut,
- b. Pneumonia akibat virus, penyebab utama pneumonia virus adalah virus *influenza* (bedakan dengan bakteri hemofilus *influenza* yang bukan penyebab penyakit *influenza*, tetapi bisa menyebabkan pneumonia juga). Tipe pneumonia itu bisa ditumpangi dengan infeksi pneumonia karena bakteri. Hal itu yang disebut dengan superinfeksi bakterial. Salah satu tanda terjadi superinfeksi bakterial adalah keluarnya lendir yang kental dan berwarna hijau atau merah tua.
- c. Pneumonia jamur, sering merupakan infeksi sekunder. Predileksi terutama pada penderita dengan daya tahan lemah (*immunocompromised*).

Berdasarkan predileksi infeksi, penyakit pneumonia diklasifikasikan menjadi 3 kelompok.

- a. Pneumonia lobaris, pneumonia yang terjadi pada satu lobus (percabangan besar dari pohon bronkus) baik kanan maupun kiri.
- b. Pneumonia bronkopneumonia, pneumonia yang ditandai bercak-bercak infeksi pada berbagai tempat di paru. Bisa kanan maupun kiri yang disebabkan virus atau bakteri dan sering terjadi pada bayi atau orang tua. Pada penderita pneumonia, kantong udara paru-paru penuh dengan nanah dan cairan yang lain. Dengan demikian, fungsi paru-paru, yaitu menyerap udara bersih (oksigen) dan mengeluarkan udara kotor menjadi terganggu. Akibatnya, tubuh menderita kekurangan oksigen dengan segala konsekuensinya, misalnya menjadi lebih mudah terinfeksi oleh bakteri lain (super infeksi) dan sebagainya. Sukar penyembuhannya.

## 2.7 Diagnosis Pneumonia

Diagnosis penderita pneumonia dilakukan dengan mengamati gejala klinis dan kelainan fisik melalui pemeriksaan fisik yang dilakukan oleh dokter. Pemeriksaan penunjang seperti *rongent* dan pemeriksaan laboratorium masih diperlukan untuk memperkuat diagnosis apakah seseorang menderita pneumonia atau tidak (Misnadiarly, 2008). Sistem skor yang biasa digunakan untuk menentukan level risiko penyakit pneumonia. Lebel risiko tersebut menjadi acuan dalam pengobatan dan perawatan penderita pneumonia. Beberapa sistem untuk menentukan tingkat keparahan penderita pneumonia yang biasa digunakan misalnya skoring *Pneumonia Severity Index* (PSI) yang dikenal juga dengan nama *Patient Outcome Research Team* (PORT), *Confusion, Urea, Respiratory Rate, Blood pressure, age > 65 years* (CURB-65) yang direkomendasikan oleh *British Thoracic Society* (BTS).

Sistem CURB-65 sudah divalidasi dengan menggunakan sampel yang lebih sedikit dari sistem PSI. Kelebihan sistem CURB adalah penggunaannya yang mudah dan dirancang untuk lebih menilai keparahan penyakit dibanding menilai pasien pneumonia dengan risiko mortalitas. Variabel-variabel yang terdapat dalam skor prediksi menggunakan sistem CURB-65 dan jumlah poin tiap variabel ditunjukkan oleh Tabel 2.2. Skor CURB-65 tiap pasien pneumonia dijumlahkan dan dikategorikan berdasarkan tingkat keparahannya. Tingkat keparahan skor CURB-65 ditunjukkan pada Tabel 2.3.

**Tabel 2.2** Skor Prediksi CURB-65

Karakteristik	Poin
Penurunan Kesadaran	1
Urea nitrogen darah > 20 mg/Dl	1
Frekuensi Pernapasan $\geq$ 30 per menit	1
Tekanan darah (sistolik < 90 mmHg atau diastolik $\leq$ 60 mmHg)	1
Usia $\geq$ 65 tahun	1

**Tabel 2.3** Kelas Risiko dengan Sistem CURB-65

Jumlah Poin	Kelas Risiko	Rekomendasi Penanganan
0	Rendah	Rawat Jalan
1	Rendah	Rawat Jalan
2	Sedang	Rawat Inap/ Rawat Jalan
3	Sedang ke Berat	Rawat Inap/ Rawat Jalan
4 atau 5	Berat	Rawat Inap/ICU

Persatuan Dokter Paru Indonesia (PDPI) menetapkan sistem PSI menjadi acuan dalam mendiagnosis penyakit pneumonia. Menurut hasil *Pneumonia Patient Outcome Reserarch Team* (PORT), beberapa karakteristik pasien dapat dimasukkan ke dalam kelas risiko seperti ditunjukkan pada Tabel 2.4 (PDPI, 2003). Poin-poin dari hasil PORT dijumlahkan. Hasil penjumlahan poin tersebut dikategorikan menurut kelas risikonya sehingga dapat ditentukan penanganan dan pengobatan yang tepat (PDPI, 2003). Tabel 2.5 menunjukkan derajat skor penentuan kelas risiko dengan sistem PSI.

**Tabel 2.4** Skor pada Pneumonia Komunitas Sistem PSI

Karakteristik Penderita	Poin
Faktor Demografi	
Usia: Laki-laki	Umur (tahun)
Perempuan	Umur (tahun) – 10
Perawatan di rumah	+10
Penyakit Penyerta	
Keganasan	+30
Penyakit hati	+20
Gagal jantung kongenstif	+10
Penyakut serebrovaskuler	+10
Penyakit ginjal	+10

**Tabel 2.4** Skor pada Pneumonia Komunitas Sistem PSI (lanjutan)

Karakteristik Penderita	Poin
<b>Pemeriksaan Fisis</b>	
Perubahan status mental	+20
Pernafasan > 30kali/menit	+20
Tekanan darah sistolik < 90 mmHg	+20
Suhu tubuh < 35°C atau > 40 °C	+15
Nadi $\geq$ 125 kali/menit	+10
<b>Hasil laboratorium atau radiologi</b>	
Analisis gas darah arteri : Ph <7,35	+30
BUN ( <i>Blood Urea Nitrogen</i> ) > 30 mg/dl	+20
Natrium < 130 mEq/liter	+20
Glukosa > 250 mg/dl	+10
Hematokrit < 30%	+10
PO <sub>2</sub> $\leq$ 60 mmHg	+10
Efusi pleura	+10

**Tabel 2.5** Derajat Skor Penentuan Kelas Risiko Dengan Sistem PSI

Risiko	Kelas	Poin	Rekomendasi Penanganan
Rendah	I	0	Rawat Jalan
	II	<70	Rawat Jalan
	III	71 – 90	Rawat Inap/ Rawat Jalan
Sedang	IV	91 – 130	Rawat Inap
Berat	V	>130	Rawat Inap

*(Halaman ini sengaja dikosongkan)*

## BAB III METODOLOGI PENELITIAN

### 3.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari penelitian sebelumnya yaitu penelitian Firdausanti (2017) dengan judul “Klasifikasi Kelas Risiko Pasien Pneumonia menggunakan Metode *Hybrid Analisis Diskriminan Linier-Particle Swarm Optimization (ADL-PSO)* dan *Naïve Bayes Classification*” ditambah dengan variabel Jenis Kelamin. Data yang dianalisis adalah data rekam medis pasien pneumonia tahun 2015.

### 3.2 Variabel Penelitian

Variabel-variabel yang digunakan untuk mengklasifikasikan kelas risiko penderita pneumonia dapat dilihat pada Tabel 3.1. Data terdiri dari satu variabel respon dengan 12 variabel dengan skala data rasio dan 1 variabel dengan skala data nominal.

**Tabel 3.1** Variabel Penelitian

Variabel	Keterangan	Skala
$Y$	Kelas Risiko Pneumonia: II/III/IV/V	Ordinal
$X_1$	Jenis kelamin: Laki-Laki, Perempuan	Nominal
$X_2$	Usia (tahun)	Rasio
$X_3$	Tekanan Darah Sistolik (mmHg)	Rasio
$X_4$	Tekanan Darah Diastolik (mmHg)	Rasio
$X_5$	Denyut Nadi (denyut/menit)	Rasio
$X_6$	Frekuensi Nafas tiap menit	Rasio
$X_7$	Suhu Tubuh (°C)	Rasio
$X_8$	Analisis Gas Darah Arteri (pH)	Rasio
$X_9$	Kadar Natrium dalam Darah (mEq/liter)	Rasio
$X_{10}$	Kadar Glukosa dalam Darah (mg/dL)	Rasio
$X_{11}$	Persentase sel darah merah terhadap volume darah /Hematokrit (%)	Rasio
$X_{12}$	Tekanan oksigen dalam darah /Kadar PO <sub>2</sub> (mmHg)	Rasio
$X_{13}$	Urea Nitrogen Darah (mg/dL)	Rasio

Tabel 3.1 merupakan Tabel yang berisi keterangan variabel respon dan variabel prediktor yang digunakan dalam analisis dimana

variabel prediktor. Selanjutnya Tabel 3.2 memberikan keterangan mengenai definisi operasional variabel respon dan variabel prediktor.

**Tabel 3.2** Definisi Operasional

Variabel	Keterangan
Y	Kelas Risiko Pneumonia II : Risiko rendah dengan penanganan rawat jalan III : Risiko rendah dengan penanganan rawat inap IV : Risiko sedang dengan penanganan rawat inap V : Risiko berat dengan penanganan rawat inap
X <sub>3</sub> , X <sub>4</sub>	Tekanan Darah Sistolik (mmHg) adalah tekanan darah pada saat terjadi kontraksi otot jantung. Tekanan Darah Diastolik (mmHg) adalah jumlah tekanan darah atau angka bawah yang menunjukkan tekanan dalam arteri saat jantung beristirahat (di antara ketukan/detak). Tekanan darah biasanya digambarkan sebagai rasio tekanan sistolik terhadap diastolik. Rata-rata tekanan darah normal biasanya 120/80.
X <sub>5</sub>	Denyut Nadi (denyut/menit) adalah frekuensi denyut nadi pasien pneumonia tiap menit. Denyut nadi orang normal dewasa adalah 60-100 kali/menit.
X <sub>6</sub>	Frekuensi Nafas adalah frekuensi nafasi pasien pneumonia tiap menit. Frekuensi nafas orang normal dewasa adalah 12-20 kali/menit.
X <sub>7</sub>	Suhu tubuh adalah ukuran dari kemampuan tubuh dalam menghasilkan dan menyingkirkan hawa panas. Nilai normal suhu manusia pada adalah 36.7 °C.
X <sub>8</sub>	Ukuran keasaman atau pH darah pada darah pasien pneumonia dengan metode pengecekannya adalah menggunakan analisa gas darah arteri. Nilai pH yang normal bagi seseorang adalah 7,35.
X <sub>9</sub>	Natrium dalam tubuh manusia berfungsi untuk membantuk mengendalikan kadar air dalam tubuh. Kadar Natrium normal dalam darah adalah 135-145 mEq/liter
X <sub>10</sub>	Kadar Glukosa normal dalam darah adalah 70-150 mG/dL
X <sub>11</sub>	Hematocrit adalah jumlah sel darah merah dalam darah. Pemeriksaan hematokrit akan mendapatkan hasil perbandingan jumlah sel darah merah terhadap volume darah dalam satuan persen. Nilai normal hematokrit pada pria dewasa adalah 38,8%-50%, sedangkan pada wanita dewasa adalah 34,9%-44,5%.
X <sub>12</sub>	Kadar PO <sub>2</sub> adalah tekanan oksigen dalam darah. Kadar yang rendah menggambarkan hipoksemia dan pasien tidak bernafas dengan adekuat. PO <sub>2</sub> dibawah 60 mmHg mengindikasikan perlunya pemberian oksigen tambahan. Kadar normal PO <sub>2</sub> adalah 80-100 mmHg
X <sub>13</sub>	Urea nitrogen darah adalah konsentrasi serum atau plasma urea, yang ditentukan dengan kandungan nitrogen dengan satuan mg/dL. Kadar normal urea dalam darah adalah 7-25 mg/dL.

### 3.3 Struktur Data

Struktur data yang digunakan dalam penelitian ini dapat dilihat pada Tabel 3.3.

**Tabel 3.3** Struktur Data

Kelas	Sampel	Variabel Respon	Variabel Prediktor			
		$Y$	$X_1$	$X_2$	$\dots$	$X_{13}$
2	1	$y_{2,1}$	$x_{2,1,1}$	$x_{2,2,1}$	$\dots$	$x_{2,13,1}$
	2	$y_{2,2}$	$x_{2,1,2}$	$x_{2,2,2}$	$\dots$	$x_{2,13,2}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$
	$n_2$	$y_{2,n_2}$	$x_{2,1,n_2}$	$x_{2,2,n_2}$	$\dots$	$x_{2,13,n_2}$
3	1	$y_{3,1}$	$x_{3,1,1}$	$x_{3,2,1}$	$\dots$	$x_{3,13,1}$
	2	$y_{3,2}$	$x_{3,1,2}$	$x_{3,2,2}$	$\dots$	$x_{3,13,2}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$
	$n_3$	$y_{3,n_3}$	$x_{3,1,n_3}$	$x_{3,2,n_3}$	$\dots$	$x_{3,13,n_3}$
4	1	$y_{4,1}$	$x_{4,1,1}$	$x_{4,2,1}$	$\dots$	$x_{4,13,1}$
	2	$y_{4,2}$	$x_{4,1,2}$	$x_{4,2,2}$	$\dots$	$x_{4,13,2}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$
	$n_4$	$y_{4,n_4}$	$x_{4,1,n_4}$	$x_{4,2,n_4}$	$\dots$	$x_{4,13,n_4}$
5	1	$y_{5,1}$	$x_{5,1,1}$	$x_{5,2,1}$	$\dots$	$x_{5,13,1}$
	2	$y_{5,2}$	$x_{5,1,2}$	$x_{5,2,2}$	$\dots$	$x_{5,13,2}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$
	$n_5$	$y_{5,n_5}$	$x_{5,1,n_5}$	$x_{5,2,n_5}$	$\dots$	$x_{5,13,n_5}$

Keterangan

- $n_2$  : Jumlah observasi pada kategori kelas risiko II  
 $n_3$  : Jumlah observasi pada kategori kelas risiko III  
 $n_4$  : Jumlah observasi pada kategori kelas risiko IV  
 $n_5$  : Jumlah observasi pada kategori kelas risiko V

### 3.4 Langkah Analisis

Langkah analisis dibuat untuk mengetahui bagaimana langkah-langkah yang dilakukan dalam proses penelitian sehingga memberikan gambaran dalam proses analisis. Berikut adalah langkah analisis yang digunakan dalam mengklasifikasikan kelas risiko penderita pneumonia.

1. Melakukan analisis statistika deskriptif dan eksplorasi data.

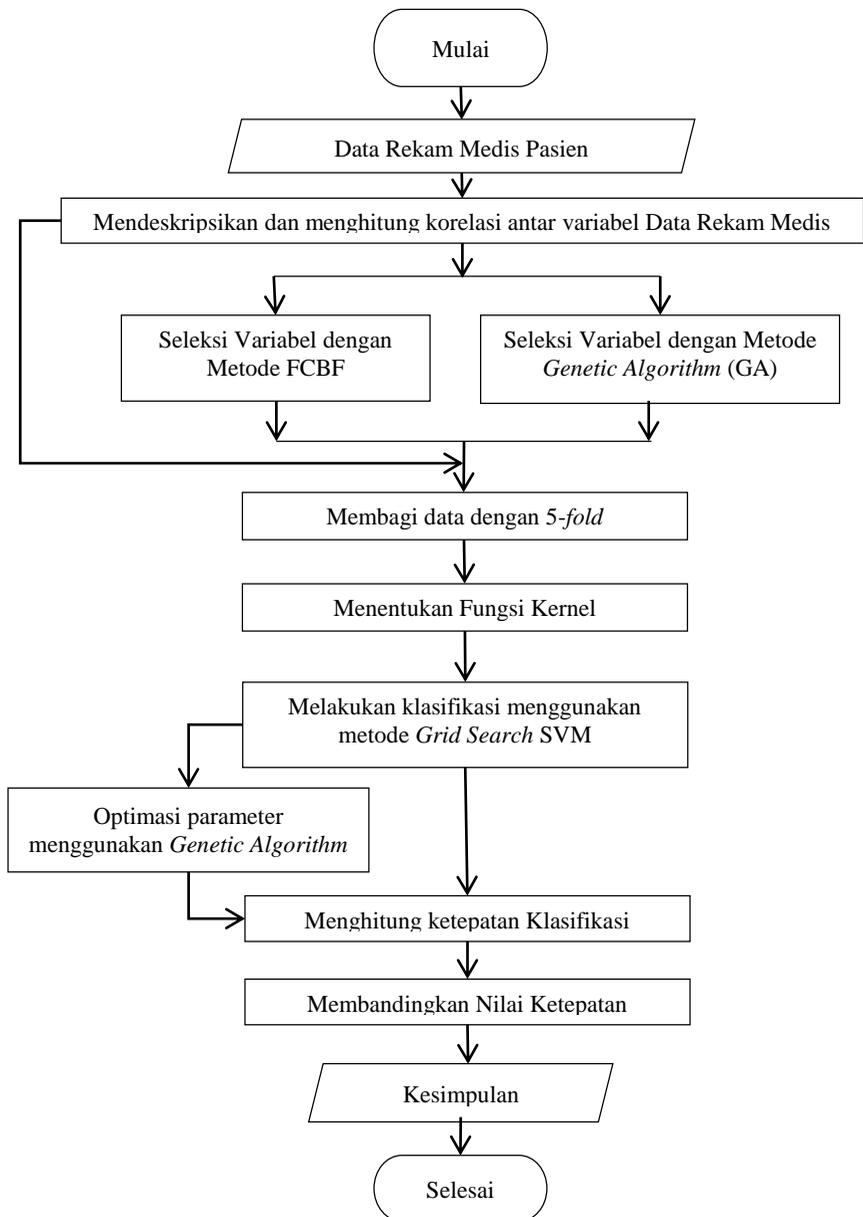
2. Melakukan seleksi variabel menggunakan metode FCBF dan *Genetic Algorithm*.
  - a. Seleksi variabel dengan metode FCBF
  - b. *Genetic Algorithm* (GA): berikut langkah *seleksi* variabel menggunakan GA
    - i. Merepresentasikan kromosom dan menentukan nilai inisialisasi. Representasi kromosom yang digunakan untuk seleksi variabel adalah *binary encoding*, yaitu 0 untuk variabel prediktor yang tidak masuk dalam model dan 1 untuk variabel prediktor yang masuk dalam model. Inisialisasi dilakukan dengan membangkitkan 100 individu secara *random*. Salah satu individu diisi dengan parameter yang telah diperoleh dari fungsi Kernel Gaussian (RBF)
    - ii. Mengevaluasi masing-masing kromosom menggunakan nilai *fitness*. Pada penelitian ini yang menjadi nilai *fitness* adalah nilai kesalahan klasifikasi
    - iii. Melakukan proses seleksi sebanyak 100 kromosom dari sejumlah 100 induk yang berasal dari populasi menggunakan seleksi *roulette wheel*
    - iv. Melakukan proses pindah silang jika nilai bilangan *random r* antara  $[0,1]$  yang dibangkitkan kurang dari probabilitas pindah silang ( $P_s$ ) = 0,8
    - v. Melakukan proses mutasi jika nilai bilangan *random r* antara  $[0,1]$  yang dibangkitkan kurang dari nilai probabilitas proses mutasi ( $P_m$ ) = 0,01
    - vi. Melakukan proses elitisme dimana dua kromosom dengan nilai *fitness* terbaik akan bertahan ke generasi selanjutnya
    - vii. Melakukan pergantian populasi lama dengan generasi baru dengan cara memilih kromosom terbaik berdasarkan nilai *fitness* yang telah melalui proses seleksi, pindah silang dan elitism
    - viii. Melakukan pengecekan terhadap solusi yang telah didapatkan. Solusi dikatakan telah mencapai kriteria apabila nilai *fitness* terbaik telah konvergen jika

kondisi ini belum terpenuhi maka proses *genetic algorithm* diulang dari langkah iv-vii.

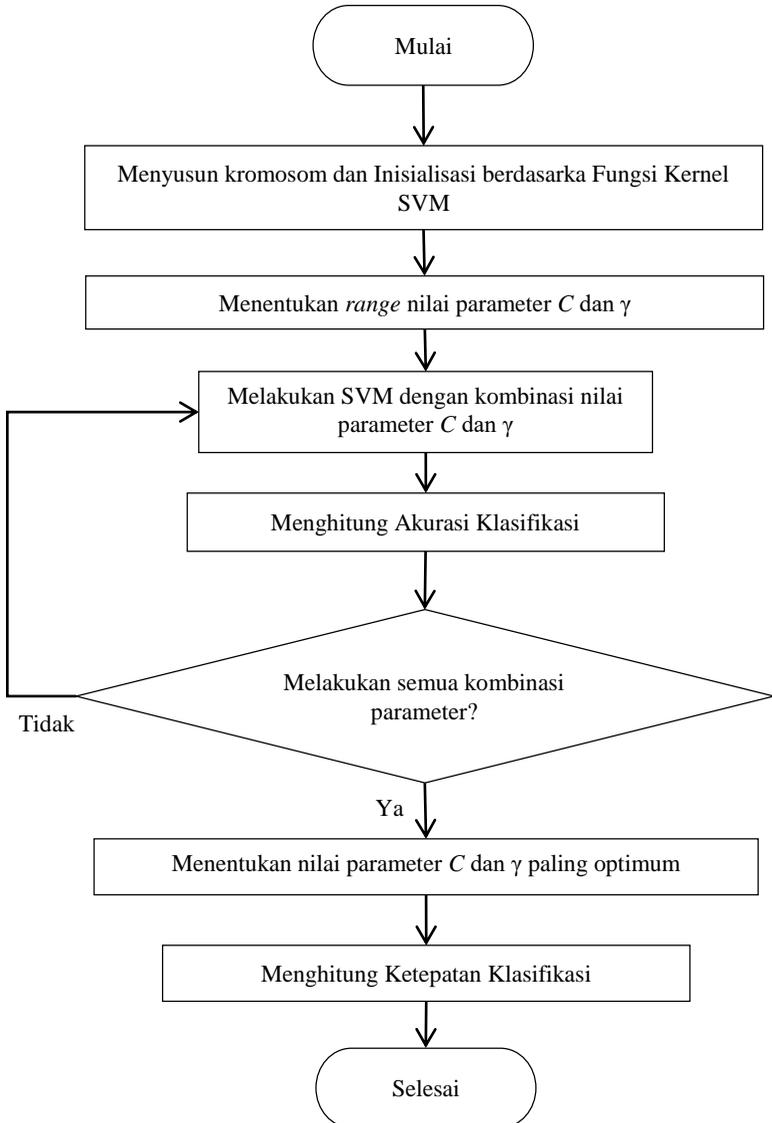
3. Membagi data dengan metode *k-Fold Crossvalidation* dengan  $k = 5$
4. Menganalisis klasifikasi menggunakan metode *grid search* SVM pada masing-masing data
  - a. Menentukan fungsi Kernel yang digunakan. Pada penelitian ini menggunakan fungsi Kernel Gaussian (RBF)
  - b. Menentukan *range* nilai parameter  $C$  dan  $\gamma$ .
  - c. Melakukan klasifikasi SVM dengan kombinasi nilai parameter  $C$  dan  $\gamma$ .
  - d. Menghitung akurasi klasifikasi.
  - e. Mengulangi langkah 4b sampai semua kombinasi sudah dilakukan.
  - f. Menentukan nilai parameter  $C$  dan  $\gamma$  yang paling optimal dari seluruh kombinasi yang dilakukan
  - g. Menghitung Kinerja klasifikasi
5. Melakukan analisis klasifikasi menggunakan *Hybrid Support Vector Machine-Genetic Algorithm* (SVM-GA) yang menghasilkan ketepatan klasifikasi tertinggi.
  - a. Menentukan nilai *fitness* untuk setiap kromosom yang didapatkan dari nilai kesalahan klasifikasi. Selain itu *stopping criteria* yang digunakan adalah nilai *fitness* konvergen, mencapai nilai 1, dan total generasi yang terbentuk adalah 1000.
  - b. Menyusun kromosom dengan membangkitkan 100 kromosom. Kromosom yang dibangkitkan terdiri dari 2 gen yang menunjukkan *hyperparameter* SVM, yaitu  $C$  dan  $\gamma$ . Nilai inisial kromosom diperoleh dari nilai parameter  $C$  dan  $\gamma$  yang paling optimal dari langkah 4f.
  - c. Melakukan optimasi parameter dengan GA seperti langkah 2.b.ii sampai 2.b.viii
  - d. Menghitung ketepatan klasifikasi.
  - e. Menentukan nilai  $C$  dan  $\gamma$  yang paling optimal.

6. Membandingkan ketepatan klasifikasi menggunakan metode *Grid Search SVM, Hybrid Support Vector Machine-Genetic Algorithm (SVM-GA)*.
7. Menarik kesimpulan dan memberi saran.

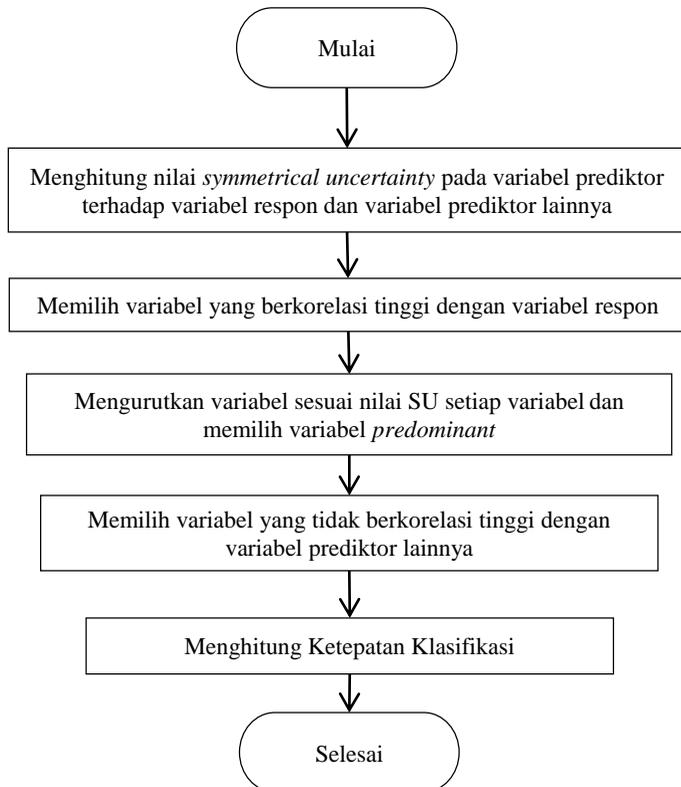
Diagram alir dalam penelitian ini menggunakan dua diagram alir. Diagram alir pertama adalah ditunjukkan pada Gambar 3.1 dimana digram alir ini adalah diagram alir berisi tahapan penelitian. Diagram alir kedua ditunjukkan pada Gambar 3.2 dimana diagram alir yang menunjukan tahapan klasifikasi menggunakan *Grid Search SVM* Gambar 3.3 dimana diagram alir yang menunjukan tahapan dari *Genetic Algorithm*.



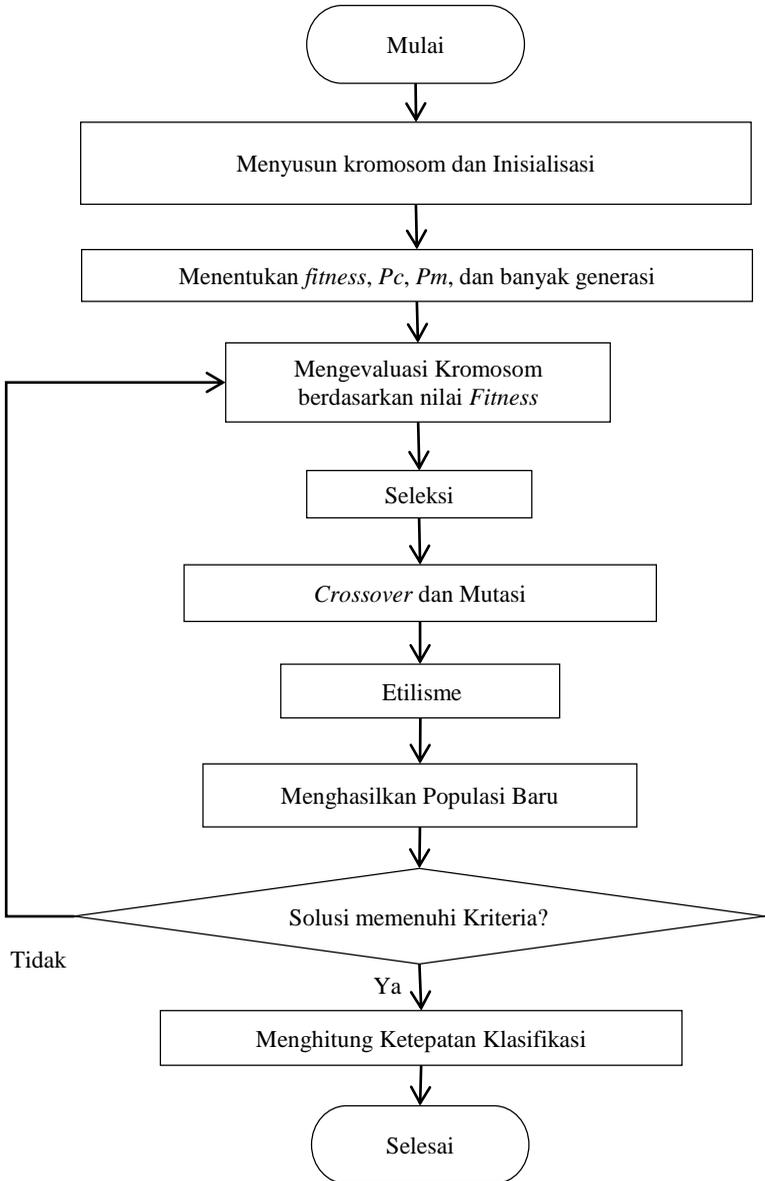
**Gambar 3.1** Diagram Alir Penelitian



**Gambar 3.2** Diagram Alir Klasifikasi Menggunakan *Grid Search* SVM



**Gambar 3.3** Diagram Alir Menggunakan FCBF

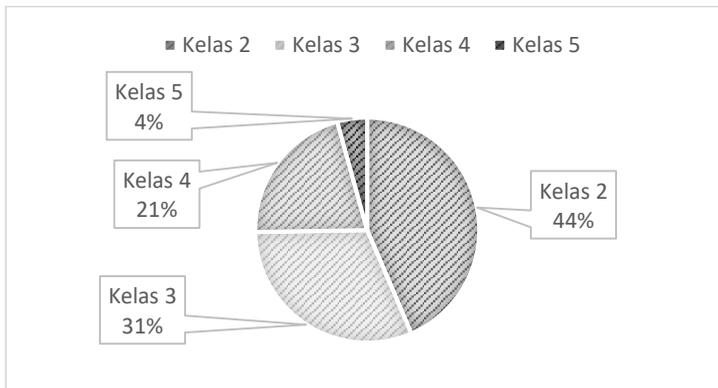


**Gambar 3.4** Diagram Alir Menggunakan GA

## BAB IV ANALISIS DAN PEMBAHASAN

### 4.1 Karakteristik Penderita Pneumonia

Analisis karakteristik data tiap variabel perlu dilakukan untuk mengetahui perbedaan karakteristik yang terdapat pada variabel independen. Gambar 4.1 merupakan persentase masing-masing kategori dari 4 kelas risiko pasien pneumonia. Persentase tiap kelas dari 186 pasien cukup berbeda, persentase pasien yang masuk ke kelas II sebesar 44% sedangkan persentase pasien yang masuk ke kelas V sebesar 4%. Artinya, data kelas risiko ini tidak seimbang (*unbalance*).

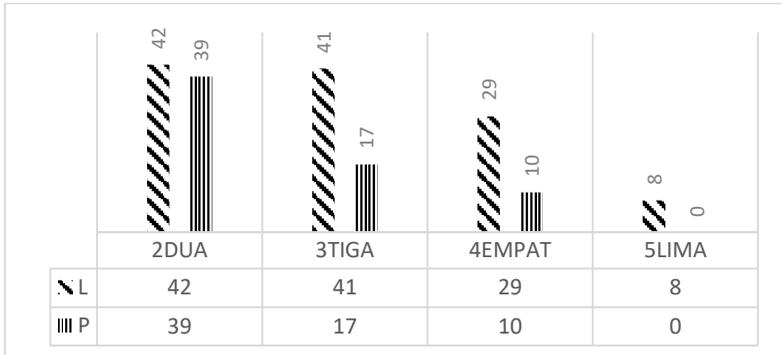


**Gambar 4.1** Proporsi Tiap Kelas Pasien Pneumonia

Pada penelitian ini, variabel prediktor yang digunakan untuk prediksi kelas pasien pneumonia ada 13 variabel. Variabel yang pertama adalah variabel jenis kelamin. Pada hasil klasifikasi PORT, poin untuk pasien perempuan berkurang 10 poin dari umurnya. Maka dari itu, pada penelitian ini akan dilakukan analisis apakah jenis kelamin pasien berpengaruh signifikan terhadap klasifikasi kelas pasien pneumonia.

Gambar 4.2 merupakan diagram batang dari jumlah pasien laki-laki dan perempuan setiap kelas. Pada kelas risiko pneumonia II, perbedaan jumlah pasien perempuan dan laki-laki tidak jauh berbeda. Namun pada kelas III, IV, dan V perbedaan jumlah pasien laki-laki dan perempuan semakin jauh. Dapat disimpulkan bahwa variabel

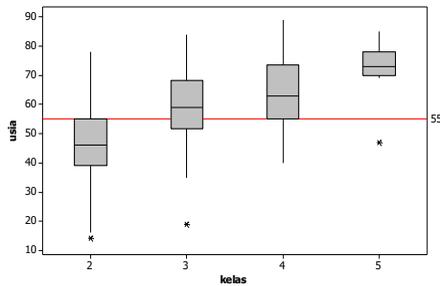
jenis kelamin diduga berpengaruh signifikan terhadap kelas risiko pneumonia. Pada Gambar 4.2 juga dapat diketahui bahwa jumlah pasien pneumonia laki-laki lebih banyak daripada perempuan pada setiap kelasnya. Dari lebih banyaknya jumlah pasien laki-laki ini ada dugaan bahwa orang dengan jenis kelamin laki-laki lebih mudah terkena pneumonia dibanding perempuan.



**Gambar 4.2** Diagram Batang Jenis Kelamin terhadap Perbedaan Kelas Klasifikasi Pasien Pneumonia

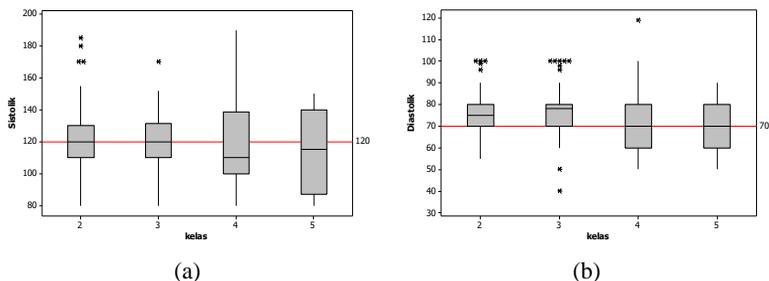
**Tabel 4.1** Rata-rata tiap Variabel dan tiap Kelas Klasifikasi Pasien Pneumonia

Variabel	Kelas				Mean Semua Data
	II	III	IV	V	
Usia	46,877	59,397	63,333	71,375	<b>55,285</b>
Sistolik	120,432	121,155	117,436	116,875	<b>119,876</b>
Diastolik	74,988	75,517	72,615	71,250	<b>74,495</b>
Denyut	103,691	103,466	107,846	119,500	<b>105,172</b>
Napas	26,420	28,552	29,436	33,500	<b>28,022</b>
Suhu	37,038	37,036	37,113	37,575	<b>37,076</b>
pH	7,451	7,433	7,357	7,340	<b>7,421</b>
Natrium	137,536	134,934	137,105	138,650	<b>136,682</b>
Glukosa	147,741	163,690	162,897	168,375	<b>156,780</b>
Hematokrit	38,211	37,690	38,064	38,500	<b>38,030</b>
PO <sub>2</sub>	109,075	95,236	125,969	98,375	<b>107,842</b>
BUN	14,878	18,969	25,797	41,500	<b>19,588</b>



**Gambar 4.3** Proporsi Tiap Kategori *Boxplot* dari Variabel Usia

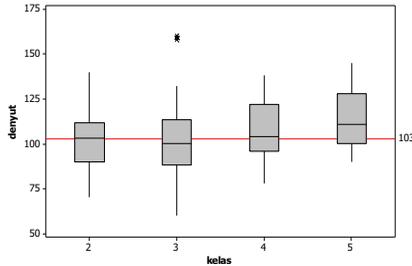
Tabel 4.1 menunjukkan rata-rata dari variabel-variabel prediktor dengan skala data rasio. Dapat diketahui bahwa rata-rata usia pasien pneumonia adalah 55 tahun dan rata-rata usia pasien meningkat setiap kelasnya. Gambar 4.3 menunjukkan bahwa *boxplot* dari variabel usia dari penderita pneumonia kelas II ke kelas V cenderung naik. Artinya, variabel usia diduga berpengaruh signifikan terhadap kelas risiko pneumonia karena adanya perbedaan dari setiap kelasnya.



**Gambar 4.4** Proporsi Tiap Kategori *Boxplot* dari Variabel Tekanan Darah (a) Sistolik, (b) Diastolik

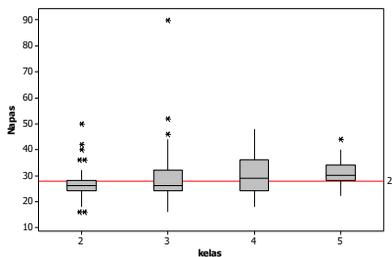
Rata-rata tekanan darah sistolik dan diastolik pasien pada penelitian ini adalah 119,976 dan 74,495. Rata-rata tekanan darah normal biasanya 120/80. Nilai rata-rata tekanan darah pasien pneumonia tidak terlalu jauh selisihnya dari ukuran normal. Gambar 4.4 (a) dan (b) menunjukkan bahwa *boxplot* dari variabel tekanan darah sistolik dan tekanan darah diastolik dari kelas II ke kelas V cenderung berada di sekitar median data keseluruhan kelas. Artinya,

variabel tekanan darah sistolik dan tekanan darah diastolik diduga tidak berpengaruh signifikan terhadap kelas risiko pneumonia karena tidak adanya perbedaan dari setiap kelasnya.



**Gambar 4.5** Proporsi Tiap Kategori *Boxplot* dari Variabel Denyut Nadi

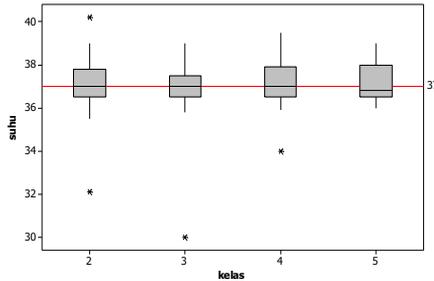
Batas normal dari jumlah denyut nadi setiap menit adalah 60-100 denyut/menit. Rata-rata jumlah denyut nadi pasien pneumonia pada penelitian ini adalah 105 denyut tiap menit. Jumlah denyut tersebut melebihi batas normal. Rata-rata jumlah denyut pada kelas ke 5 adalah 119 denyut/menit. Nilai tersebut cukup jauh dari batas normal dan rata-ratanya. Dapat dilihat pula pada Gambar 4.5 menunjukkan bahwa *boxplot* dari variabel denyut nadi dari kelas II ke kelas V cenderung naik pada kelas V, namun masih berada di sekitar median dari keseluruhan data. Artinya, variabel tekanan darah diastolik diduga tidak berpengaruh signifikan terhadap kelas risiko pneumonia karena tidak adanya perbedaan yang signifikan setiap kelasnya.



**Gambar 4.6** Proporsi Tiap Kategori *Boxplot* dari Variabel Frekuensi Napas tiap Menit

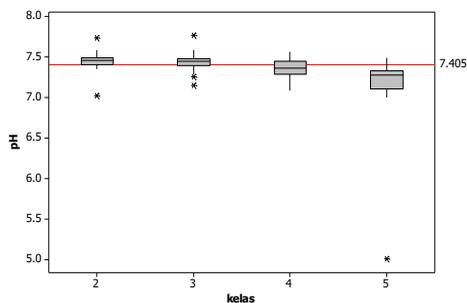
Gambar 4.6 menunjukkan bahwa *boxplot* dari variabel frekuensi nafas tiap menit dari kelas II ke kelas V cenderung berada di sekitar

median data keseluruhan kelas. Artinya, variabel frekuensi nafas tiap menit diduga tidak berpengaruh signifikan terhadap kelas risiko pneumonia karena tidak adanya perbedaan dari setiap kelasnya.



**Gambar 4.7** Proporsi Tiap Kategori *Boxplot* dari Variabel Suhu Tubuh Pasien

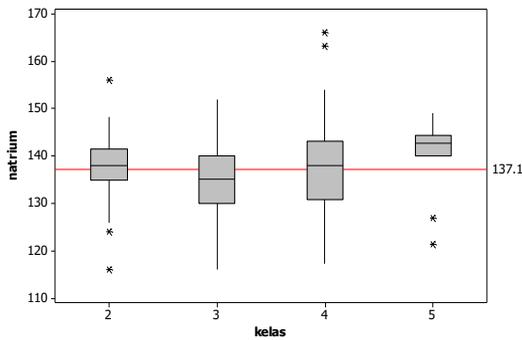
Suhu tubuh manusia normalnya adalah  $36,7^{\circ}\text{C}$ . Rata-rata suhu tubuh pasien pneumonia pada penelitian ini  $37,076^{\circ}\text{C}$ . Nilai tersebut melebihi suhu tubuh normal manusia. Namun bila dilihat dari rata-rata suhu tubuh antar kelas klasifikasi penyakit pneumonia tidak terlalu terlihat perbedaan yang signifikan karena masih berada di sekitar  $37^{\circ}\text{C}$ . Gambar 4.7 menunjukkan bahwa *boxplot* dari variabel suhu tubuh dari kelas II ke kelas V cenderung berada di sekitar median data keseluruhan kelas. Artinya, variabel suhu tubuh diduga tidak berpengaruh signifikan terhadap kelas risiko pneumonia karena tidak adanya perbedaan dari setiap kelasnya.



**Gambar 4.8** Proporsi Tiap Kategori *Boxplot* dari pH Darah

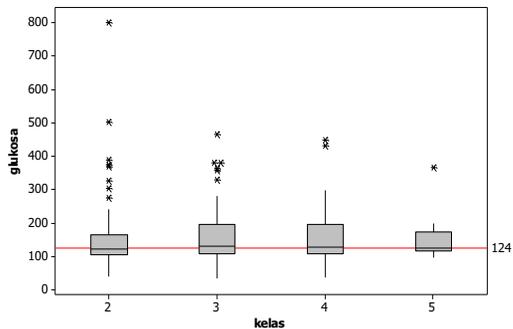
Ukuran keasaman atau pH darah pada darah pasien pneumonia dengan metode pengecekannya adalah menggunakan analisa gas

darah arteri. Nilai pH yang normal bagi seseorang adalah 7,35. Rata-rata pH darah pasien pneumonia pada penelitian ini adalah 7,42, yang tidak terlalu berbeda dari nilai normalnya. Namun bila dilihat kecenderungan pH darah dari tiap kelas seperti pada gambar 4.8 menunjukkan bahwa *boxplot* dari variabel pH atau derajat keasaman darah dari kelas II ke kelas V cenderung turun pada kelas V. terdapat nilai yang sangat rendah pada kelas V sampai mendekati nilai derajat keasaman 5. Artinya, variabel analisis gas darah arteri diduga berpengaruh signifikan terhadap kelas risiko pneumonia karena adanya perbedaan median setiap kelasnya.



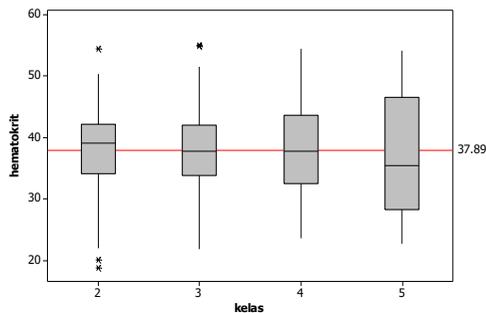
**Gambar 4.9** Proporsi Tiap Kategori *Boxplot* dari Variabel Kadar Natrium dalam Darah

Natrium dalam tubuh manusia berfungsi untuk membantu mengendalikan kadar air dalam tubuh. Kadar Natrium normal dalam darah adalah 135-145 mEq/liter. Rata-rata kadar natrium dalam darah pasien pneumonia pada penelitian ini adalah 136,68 mEq/liter. Artinya, rata-rata kadar natrium dalam darah pasien masih berada pada nilai yang normal. Gambar 4.9 menunjukkan bahwa *boxplot* dari variabel kadar natrium dalam darah dari kelas II ke kelas V cenderung naik pada kelas V, namun masih berada di sekitar median dari keseluruhan data. Artinya, variabel kadar natrium dalam darah diduga tidak berpengaruh signifikan terhadap kelas risiko pneumonia karena tidak adanya perbedaan yang signifikan setiap kelasnya.



**Gambar 4.10** Proporsi Tiap Kategori *Boxplot* dari Variabel Kadar Glukosa dalam Darah

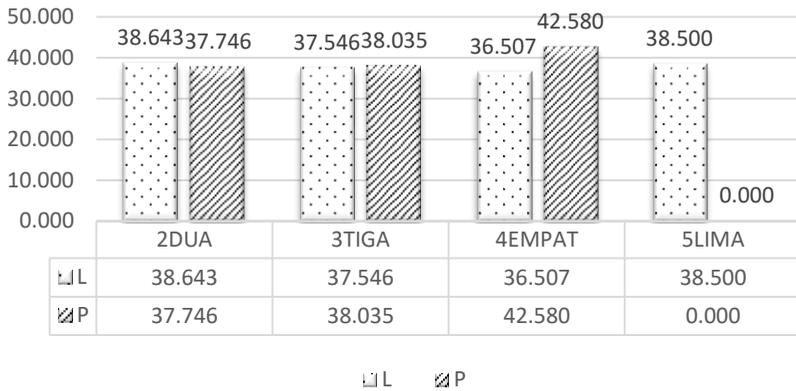
Kadar Glukosa normal dalam darah adalah 70-150 mG/dL. Rata-rata Gambar 4.10 merupakan *boxplot* dari variabel kadar glukosa dalam darah. Gambar 4.10 menunjukkan bahwa *boxplot* dari variabel kadar glukosa dalam darah dari kelas II ke kelas V cenderung berada di sekitar median data keseluruhan kelas. Artinya, variabel kadar glukosa dalam darah diduga tidak berpengaruh signifikan terhadap kelas risiko pneumonia karena tidak adanya perbedaan dari setiap kelasnya.



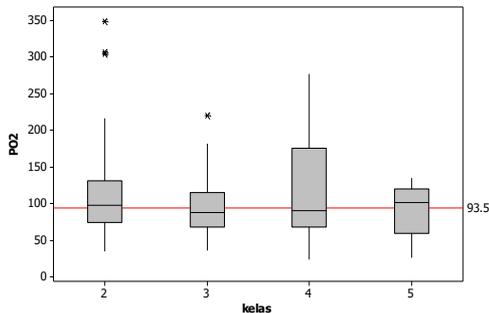
**Gambar 4.11** Proporsi Tiap Kategori *Boxplot* dari Variabel Hematokrit

Nilai normal hematokrit pada pria dewasa adalah 38,8%-50%, sedangkan pada wanita dewasa adalah 34,9%-44,5%. Rata-rata hematokrit pada pasien pneumonia adalah 38,03%, artinya, masih berada pada batas normal. Dari Gambar 4.12 dapat diketahui bahwa

rata-rata hematokrit pasien pneumonia laki-laki setiap kelas masih dalam batas normal, begitu pula pasien perempuan. Gambar 4.11 merupakan *boxplot* dari variabel persentase sel darah merah terhadap volume darah atau hematokrit Gambar 4.11 menunjukkan bahwa *boxplot* dari variabel hematokrit dari kelas II ke kelas V cenderung berada di sekitar median data keseluruhan kelas. Artinya, hematokrit diduga tidak berpengaruh signifikan terhadap kelas risiko pneumonia karena tidak adanya perbedaan dari setiap kelasnya.

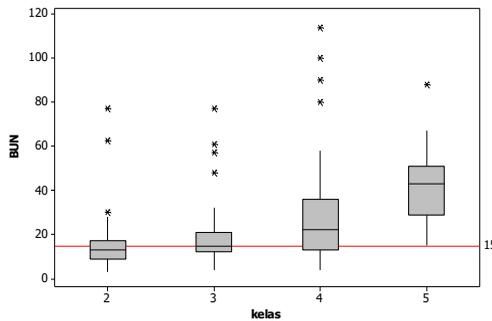


**Gambar 4.12** Rata-rata Hematokrit Pasien Pneumonia Berdasarkan Jenis Kelamin dan Kelas Klasifikasi



**Gambar 4.13** Proporsi Tiap Kategori *Boxplot* dari Variabel Kadar PO<sub>2</sub>

Kadar  $PO_2$  adalah tekanan oksigen dalam darah.  $PO_2$  dibawah 60 mmHg mengindikasikan perlunya pemberian oksigen tambahan. Kadar normal  $PO_2$  adalah 80-100 mmHg. Rata-rata kadar  $PO_2$  darah pada pasien pneumonia adalah 107,84 mmHg. Nilai ini melebihi batas normalnya. Bila ditinjau dari setiap kelasnya seperti yang diilustrasikan pada gambar 4.13 merupakan *boxplot* dari variabel tekanan oksigen dalam darah atau bisa juga disebut kadar  $PO_2$ . Gambar 4.13 kadar  $PO_2$  dari kelas II ke kelas V cenderung berada di sekitar median data keseluruhan kelas. Artinya, variabel kadar  $PO_2$  diduga tidak berpengaruh signifikan terhadap kelas risiko pneumonia karena tidak adanya perbedaan dari setiap kelasnya.



**Gambar 4.14** Proporsi Tiap Kategori *Boxplot* dari Variabel Urea Nitrogen Darah

Urea nitrogen darah adalah konsentrasi serum atau plasma urea, yang ditentukan dengan kandungan nitrogen dengan satuan mg/dL. Kadar normal urea dalam darah adalah 7-25 mg/dL. Rata-rata kadar urea nitrogen dalam darah pasien pneumonia pada penelitian ini adalah 19 mg/dL yang masih berada pada batas normalnya. Gambar 4.14 menunjukkan bahwa *boxplot* dari variabel urea nitrogen dalam darah dari kelas II ke kelas V cenderung naik. Artinya, variabel urea nitrogen dalam darah diduga berpengaruh signifikan terhadap kelas risiko pneumonia karena adanya perbedaan median setiap kelasnya.

## 4.2 Klasifikasi dengan *Grid Search Support Vector Machine*

Klasifikasi penderita pneumonia dengan menggunakan metode klasifikasi SVM. Prinsip *grid search* digunakan untuk memperoleh parameter yang membentuk model yang optimal.

Salah satu metode untuk mencari *hyperparameter* yang sesuai adalah metode *grid search*. Algoritma *grid search* membagi jangkauan parameter yang akan dioptimalkan ke dalam *grid* dan melintasi semua titik untuk mendapatkan parameter yang optimal (Yasi, Prahutama, & Utami, 2014). Algoritma *grid search* pada penelitian ini menggunakan fungsi *5-fold cross validation*. Kemudian dicoba beberapa variasi pasangan parameter agar mendapatkan *range* yang optimum dilihat dari akurasi data *testing cross validation* yang sudah dilakukan sebelumnya. *Range* yang optimum kemudian selanjutnya digunakan untuk mencari parameter *hyperplane* yang optimum. Parameter yang optimum yang sudah didapatkan kemudian digunakan untuk memprediksi klasifikasi data.

Klasifikasi penderita pneumonia menggunakan metode *grid search* SVM. Menurut penelitian sebelumnya oleh Kusumaningrum (2017), parameter optimum dari penelitiannya adalah  $C$  pada *range*  $2^3 - 2^7$  dan parameter  $\gamma$  pada *range*  $2^{-9} - 2^{-3}$ . Pada penelitian ini, akan dicoba beberapa kombinasi *range* dengan menggunakan kombinasi nilai parameter  $C$  pada *range*  $2^0 - 2^3$ ,  $2^3 - 2^7$ ,  $2^7 - 2^{12}$ , dan  $2^{12} - 2^{14}$  serta nilai parameter  $\gamma$  pada *range*  $2^{-15} - 2^{-9}$ ,  $2^{-9} - 2^{-3}$ , dan  $2^{-3} - 2^3$ . Karena setiap percobaan menghasilkan nilai parameter optimal yang berbeda maka percobaan dilakukan pada data sebanyak 5 kali untuk setiap kombinasi *range* parameter. *Range* parameter yang terpilih merupakan rata-rata Kinerja terbaik dari setiap kombinasi *range* parameter.

**Tabel 4.2** Kombinasi *Range* Parameter untuk *Grid Search* SVM

<b>Range Parameter</b>		<b>Rata-rata Akurasi</b>
<b><math>C</math></b>	<b><math>\gamma</math></b>	<b>(%)</b>
<b><math>2^0 - 2^3</math></b>	$2^{-15} - 2^{-9}$	63,101
	$2^{-9} - 2^{-3}$	64,197
	<b><math>2^{-3} - 2^3</math></b>	<b>64,831</b>
	$2^3 - 2^6$	43,592
<b><math>2^3 - 2^7</math></b>	$2^{-15} - 2^{-9}$	63,001
	$2^{-9} - 2^{-3}$	62,979
	$2^{-3} - 2^3$	62,148
	$2^3 - 2^6$	43,592

**Tabel 4.2** Kombinasi *Range* Parameter untuk *Grid Search* SVM (lanjutan)

Range Parameter		Rata-rata Akurasi
<i>C</i>	$\gamma$	(%)
$2^7 - 2^{12}$	$2^{-15} - 2^{-9}$	61,270
	$2^{-9} - 2^{-3}$	61,654
	$2^{-3} - 2^3$	62,148
	$2^3 - 2^6$	43,592
$2^{12} - 2^{14}$	$2^{-15} - 2^{-9}$	59,571
	$2^{-9} - 2^{-3}$	63,249
	$2^{-3} - 2^3$	62,148
	$2^3 - 2^6$	43,592

Dari Tabel 4.2, dapat disimpulkan bahwa nilai parameter *C* dan  $\gamma$  yang optimal berada pada *range*  $2^0 - 2^3$  dan  $2^{-3} - 2^3$ . Hal tersebut ditunjukkan dari nilai rata-rata akurasi yang diperoleh pada *range* tersebut merupakan rata-rata akurasi yang paling tinggi dibandingkan dengan rata-rata akurasi kombinasi *range* yang lain, yaitu sebesar 63,249%.

#### 4.2.1. Tanpa Seleksi Variabel

Setelah mendapatkan kombinasi *range* yang optimal dari nilai parameter *C* dan  $\gamma$ , langkah selanjutnya adalah mendapatkan nilai parameter yang optimal dari *range* yang berada pada *range* tersebut. Akurasi pada penelitian diperoleh dengan metode validasi *5-fold Cross-validation* pada data. Tabel 4.3 menunjukkan nilai parameter optimal dan akurasi yang diperoleh dari percobaan yang dilakukan.

**Tabel 4.3** Hasil Percobaan *Grid Search* SVM Tanpa Seleksi Variabel

Percobaan ke-	Range Parameter		Rata-rata Akurasi (%)
	<i>C</i>	$\gamma$	
1	$2^0$	$2^{-3}$	61,538
2	$2^1$	$2^{-3}$	68,421
3	$2^1$	$2^{-3}$	71,053
4	$2^1$	$2^{-3}$	66,667
5	$2^1$	$2^{-3}$	<b>71,429</b>

Dari Tabel 4.3, dapat diketahui bahwa dari 5 percobaan yang dilakukan, akurasi tertinggi yaitu pada nilai akurasi 71,429 % diperoleh pada percobaan ke-5. Dari percobaan tersebut, nilai parameter *C* dan  $\gamma$  yang optimum adalah  $2^1$  atau 2 dan  $2^{-3}$  atau 0,125.

Berdasarkan nilai parameter  $C$  dan  $\gamma$  yang telah didapatkan, fungsi *hyperplane* yang terbentuk untuk klasifikasi pada pasien Pneumonia dengan menggunakan metode *grid search SVM* adalah

$$f(\mathbf{x}) = \sum_{i=1}^n a_i y_i \exp(-0,125 \|\mathbf{x}_i - \mathbf{x}\|^2) + b$$

dengan  $\mathbf{x}$  adalah vektor observasi dari variabel prediktor dan  $n$  merupakan jumlah observasi. Setelah parameter  $C$  dan  $\gamma$  yang paling optimal sudah didapatkan, kemudian dibentuk model untuk mengetahui Kinerja klasifikasi pada data. Pembagian data dilakukan dengan menggunakan metode *5-fold cross-validation*.

Dari fungsi *hyperplane* yang sudah didapatkan kemudian dilakukan klasifikasi pada data dengan menggunakan aturan dari persamaan (2.27). Hasil klasifikasi pada data *testing* pada *fold* pertama ditunjukkan pada Tabel 4.4.

**Tabel 4.4** *Confusion Matrix* Klasifikasi pada Data *Testing* Pembagian Data Pertama

Aktual	Prediksi			
	Kelas 2	Kelas 3	Kelas 4	Kelas 5
Kelas 2	13	2	2	0
Kelas 3	3	6	3	0
Kelas 4	1	2	4	1
Kelas 5	0	0	2	0

Nilai *recall* ( $R$ ) untuk setiap kategori didapatkan dari persentase dari data yang tepat diklasifikasikan dibagi dengan total observasi tiap kategori. Perhitungan nilai  $R$  adalah untuk kelas 2, 3, 4, dan 5 adalah

$$R_2 = \frac{13}{13 + 2 + 2 + 0} \times 100\% = 76,471\%$$

$$R_3 = 50\%$$

$$R_4 = 50\%$$

$$R_5 = 0\%$$

Perbandingan performansi dilihat dari nilai  $AUC$  dan  $G\_mean$ . Perhitungan performansi klasifikasi dengan kriteria  $AUC$  seperti pada rumus (2.27) adalah

$$AUC = \frac{1}{4} (76,471\% + 50\% + 50\% + 0\%) = 44,118\%$$

Kemudian perhitungan performansi klasifikasi dengan kriteria  $G\_mean$  seperti pada rumus (2.8) adalah

$$G\_mean = (76,471\% \times 50\% \times 50\% \times 0\%)^{\frac{1}{4}} = 0\%$$

Karena ada beberapa pembagian data dengan kriteria  $G\_mean$  menghasilkan nilai 0, maka ditambah kriteria performansi klasifikasi dengan nilai akurasi. Maka nilai akurasi pada data testing tersebut adalah jumlah observasi yang tepat diklasifikasikan dibagi dengan jumlah seluruh observasi.

$$\begin{aligned} akurasi &= \frac{\sum_{i=2}^5 n_{ii}}{\sum_{i=2}^5 \sum_{j=2}^5 n_{ij}} = \frac{n_{22} + n_{33} + n_{44} + n_{55}}{n_{22} + n_{23} + \dots + n_{55}} \\ &= \frac{13 + 6 + 4 + 0}{13 + 2 + \dots + 0} = \frac{23}{39} \times 100\% = 58,974\% \end{aligned}$$

Tabel 4.5 merupakan tabel hasil kinerja klasifikasi dari data pasien pneumonia dengan kriteria performansi  $AUC$ ,  $G\_mean$ , dan akurasi.

**Tabel 4.5** Hasil Kinerja Klasifikasi Data Pasien Pneumonia

Pembagian data ke-	Data	$AUC$	$G\_mean$	Akurasi
1	<i>Training</i>	99,457	99,452	99,320
	<i>Testing</i>	44,118	0,000	58,974
2	<i>Training</i>	95,027	94,764	98,649
	<i>Testing</i>	53,125	0,000	68,421
3	<i>Training</i>	93,940	93,717	97,297
	<i>Testing</i>	53,125	0,000	71,053
4	<i>Training</i>	99,468	99,464	99,333
	<i>Testing</i>	46,449	0,000	66,667
5	<i>Training</i>	99,468	99,464	99,338
	<i>Testing</i>	53,369	0,000	71,429
Rata-rata	<i>Training</i>	97,472	97,372	98,787
	<i>Testing</i>	50,037	0,000	67,309

Hasil analisis pada Tabel 4.5 menunjukkan bahwa ketepatan klasifikasi pada data *training* adalah 98,787% sedangkan pada data *testing* 67,309%. Dari hasil tersebut, ketepatan klasifikasi dari model SVM yang terbentuk untuk klasifikasi adalah sebesar 67,309% ditinjau dari nilai akurasi, dan 50,037% ditinjau dari nilai  $AUC$ .

Ketepatan klasifikasi  $G\_mean$  sebesar 0% dikarenakan tidak ada hasil prediksi pada kelas ke lima tepat memprediksi observasi aktualnya.

Selisih yang cukup jauh antara akurasi data *training* dan *testing* diduga disebabkan oleh model *training* yang *overfitting*. Salah satu cara untuk mengatasi *overfitting* adalah dengan seleksi variabel.

#### 4.2.2. Seleksi Variabel *Fast Correlation Based Filter* (FCBF)

Seleksi variabel adalah metode untuk memilih variabel yang akan digunakan untuk klasifikasi menggunakan SVM. Variabel yang dipakai dalam pemodelan merupakan variabel yang berpengaruh signifikan terhadap variabel respon. Dengan semakin berkurangnya jumlah variabel, proses pemodelan menjadi lebih cepat. Pada penelitian ini, salah satu metode yang digunakan untuk seleksi variabel adalah *Fast Correlation Based Filter* (FCBF).

Langkah pertama dalam seleksi variabel menggunakan FCBF adalah menghitung entropi dari setiap variabel prediktor dan entropi setiap variabel prediktor bila diketahui variabel respon dan variabel prediktor lain. Dari nilai entropi yang sudah dihitung, diperoleh *information gain* seperti rumus (2.4). Setelah didapatkan nilai *information gain*, untuk mengukur korelasi antar variabel, maka digunakan *symmetrical uncertainty* dengan rumus (2.5).

Setelah mendapatkan nilai *symmetrical uncertainty* ( $SU$ ) untuk setiap variabel, langkah selanjutnya adalah milih variabel-variabel yang memiliki nilai  $SU$  yang tinggi terhadap variabel respon atau kelas klasifikasi pasien pneumonia. Pemilihan variabel tersebut adalah berdasarkan nilai *threshold* dan mengurutkannya sesuai  $SU$ . Nilai *threshold* yang digunakan pada penelitian ini adalah 0,5. Bila nilai  $SU$  variabel prediktor lebih dari nilai *threshold* maka variabel tersebut dianggap relevan dan berkorelasi tinggi dengan variabel respon. Variabel prediktor yang berkorelasi tinggi dimasukkan ke  $S'_{list}$ .

Tahapan berikutnya adalah menghilangkan variabel-variabel yang *redundant* atau saling berkorelasi tinggi. Variabel prediktor dengan nilai  $SU$  terhadap kelas tertinggi merupakan variabel yang dinamakan variabel *predominant* yang akan digunakan untuk menyaring variabel-variabel prediktor lainnya. Misalkan variabel yang menjadi variabel *predominant* adalah variabel  $x_1$ , dan variabel

yang ingin dibandingkan adalah variabel  $x_2$ , dan variabel prediktor disimbolkan dengan  $y$ . Bila nilai  $SU(x_1, x_2)$  lebih besar dari  $SU(x_2, y)$  maka variabel  $x_2$  berkorelasi tinggi dengan  $x_1$  sehingga variabel  $x_2$  tidak terpilih, begitu pula sebaliknya. Langkah tersebut dilakukan untuk semua variabel yang ada pada  $S'_{list}$ .

Klasifikasi pasien pneumonia pada penelitian ini terdapat sebanyak 13 variabel prediktor yang akan direduksi. Setelah dilakukan seleksi variabel dengan FCBF, variabel yang terpilih dalam pemodelan adalah variabel jenis kelamin, usia, pH darah dan kandungan urin dalam darah. Setelah terpilih variabel yang memiliki korelasi tinggi terhadap kelas klasifikasi pasien pneumonia, selanjutnya dilakukan analisis klasifikasi *grid search SVM* pada data hasil seleksi variabel.

Akurasi pada penelitian diperoleh dengan metode validasi *5-fold Cross-validation* pada data. Tabel 4.6 berikut menunjukkan nilai parameter optimal dan akurasi yang diperoleh dari percobaan yang dilakukan.

**Tabel 4.6** Hasil Percobaan *Grid Search SVM* Seleksi Variabel FCBF

Percobaan ke-	Parameter		Rata-rata Akurasi (%)
	$C$	$\gamma$	
1	$2^0$	$2^{-2}$	50,256
2	$2^2$	$2^{-2}$	58,421
3	$2^2$	$2^{-3}$	64,737
4	$2^2$	$2^{-1}$	62,778
5	$2^2$	$2^{-3}$	<b>72,571</b>

Dari Tabel 4.6, dapat diketahui bahwa dari 5 percobaan yang dilakukan, akurasi tertinggi yaitu pada nilai akurasi 72,571 % diperoleh pada percobaan ke-5. Dari percobaan tersebut, nilai parameter  $C$  dan  $\gamma$  yang optimum adalah  $2^2$  atau 4 dan  $2^{-3}$  atau 0,125.

Berdasarkan nilai parameter  $C$  dan  $\gamma$  yang telah didapatkan, fungsi *hyperplane* yang terbentuk untuk klasifikasi pada pasien Pneumonia dengan menggunakan metode *grid search SVM* adalah

$$f(\mathbf{x}) = \sum_{i=1}^n a_i y_i \exp(-0,125 \|\mathbf{x}_i - \mathbf{x}\|^2) + b$$

dengan  $x$  adalah vektor observasi dari variabel prediktor dan  $n$  merupakan jumlah observasi. Setelah parameter  $C$  dan  $\gamma$  yang paling optimal sudah didapatkan, kemudian dibentuk model untuk mengetahui ketepatan klasifikasi pada data. Pembagian data dilakukan dengan menggunakan metode *5-fold cross-validation*. Tabel 4.7 merupakan tabel hasil kinerja klasifikasi dari data pasien pneumonia.

**Tabel 4.7** Hasil Kinerja Klasifikasi Dari Data Seleksi Variabel FCBF

Pembagian data ke-	Data	AUC	G_mean	Akurasi
1	<i>Training</i>	62,769	59,176	74,150
	<i>Testing</i>	36,397	0,000	51,282
2	<i>Training</i>	52,297	43,834	68,243
	<i>Testing</i>	42,708	0,000	57,895
3	<i>Training</i>	50,044	0,000	68,243
	<i>Testing</i>	49,479	0,000	65,789
4	<i>Training</i>	55,672	51,612	67,333
	<i>Testing</i>	44,886	0,000	63,889
5	<i>Training</i>	50,794	42,124	65,563
	<i>Testing</i>	57,204	0,000	77,143
Rata-rata	<i>Training</i>	54,315	39,349	68,706
	<i>Testing</i>	46,135	0,000	63,200

Hasil analisis pada Tabel 4.7 menunjukkan bahwa ketepatan klasifikasi pada data *training* adalah 68,706% sedangkan pada data *testing* 63,2%. Dari hasil tersebut, ketepatan klasifikasi dari model SVM yang terbentuk untuk klasifikasi adalah sebesar 63,2% ditinjau dari nilai akurasi, dan 46,135% ditinjau dari nilai AUC. Ketepatan klasifikasi *G\_mean* sebesar 0% dikarenakan tidak ada hasil prediksi pada kelas ke lima tepat memprediksi observasi aktualnya.

#### 4.2.3. Seleksi Variabel *Genetic Algorithm* (GA)

Seleksi variabel menggunakan metode GA dengan jumlah kromosom berjumlah sama dengan jumlah variabel prediktor pada penelitian ini yaitu 13, dengan nilai *allele* merupakan *binary encoding* yaitu antara 1 dan 0. Nilai 1 pada kromosom menandakan bahwa variabel tersebut ikut dalam pemodelan sedangkan 0 menandakan bahwa variabel tersebut tidak ikut dalam pemodelan.

1	1	1	0	1	1	1	1	1	1	1	0	1
$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$X_{10}$	$X_{11}$	$X_{12}$	$X_{13}$

**Gambar 4.15** Representasi Satu Buah Kromosom Awal Data pada Seleksi Variabel

Variabel terpilih didapatkan dari kombinasi variabel yang menghasilkan nilai kesalahan klasifikasi terkecil. Gambar 4.15 merupakan nilai inisialisasi pada kromosom pertama. Inisialisasi bisa berbeda setiap penelitian. Inisialisasi yang digunakan yaitu dengan memasukkan 13 variabel independen yang memiliki kode 1 kedalam kromosom. Setelah inisialisasi, langkah selanjutnya adalah membentuk 100 kromosom awal yang dibangkitkan untuk populasi awal seperti pada Tabel 4.8.

**Tabel 4.8** Ilustrasi Kromosom Dalam Populasi Awal pada Seleksi Variabel

Populasi	Kromosom			
	$X_1$	$X_2$	...	$X_{12}$
1	1	1	...	1
2	1	0	...	1
3	1	0	...	0
⋮	⋮	⋮	⋮	⋮
99	1	1	...	1
100	1	0	...	1

Setelah memperoleh seluruh populasi, dilakukan evaluasi pada masing-masing populasi yang telah terbentuk adalah dengan metode evaluasi *AUC* sebagai fungsi *fitness* pada algoritma ini. Tabel 4.9 menunjukkan nilai *fitness* untuk masing-masing kromosom pada populasi awal.

**Tabel 4.9** Ilustrasi Kromosom Dalam Populasi Awal pada Seleksi Variabel dengan Nilai *Fitness*

Populasi	Kromosom				Nilai <i>Fitness</i>
	$X_1$	$X_2$	...	$X_{13}$	
1	1	1	...	1	0,513
2	1	0	...	1	0,739
3	1	0	...	0	0,743
⋮	⋮	⋮	⋮	⋮	⋮
99	1	1	...	1	0,726
100	1	0	...	1	0,573

Setelah mendapatkan nilai *fitness* dari masing-masing populasi, selanjutnya adalah melakukan pembentukan kromosom orang tua dengan metode *Roulette Wheel Selection* (RWS). Langkah awal pada proses RWS adalah menghitung nilai *fitness* yang dari 100 kromosom yang dibangkitkan. Setelah itu, semua nilai *fitness* dijumlahkan kemudian setiap kromosom dihitung nilai proporsi kumulatifnya dari nilai *fitness* dibagi dengan total *fitness*. Selanjutnya adalah menghitung nilai proporsi *fitness* kumulatif untuk semua kromosom.

**Tabel 4.10** Ilustrasi Proses RWS pada Seleksi Variabel

Populasi	<i>Fitness</i>	Proporsi Nilai <i>Fitness</i>	Nilai <i>Fitness</i> Komulatif	<i>Random</i> <i>Number</i>
1	0,625	0,008	0,008	0,028
2	0,683	0,012	0,020	0,628
3	0,669	0,012	0,031	0,827
⋮	⋮	⋮	⋮	⋮
99	0,781	0,011	0,991	0,688
100	0,625	0,009	1,000	0,236

Berdasarkan Tabel 4.10 diketahui bahwa bilangan *random* pertama yaitu sebesar 0,028, maka kromosom orangtua 1 didapatkan dari nilai *fitness* komulatif yang lebih besar dari 0,028 yang pertama yaitu kromosom 3, maka kromosom tersebut menjadi kromosom 1 pada populasi baru. Bilangan *random* diperoleh dengan membangkitkan 100 data berdistribusi *uniform* dengan *range* 0 sampai 1. Tahapan tersebut berhenti ketika telah diperoleh 100 kromosom orang tua berdasarkan bilangan *random* yang telah dibangkitkan.

Langkah selanjutnya adalah melakukan proses pindah silang atau *crossover*. Proses pindah silang menghasilkan kromosom baru dari hasil kombinasi 2 kromosom orang tua. Metode pindah silang yang digunakan pada penelitian adalah *single point crossover*.

Kromosom	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$X_{10}$	$X_{11}$	$X_{12}$	$X_{13}$	<i>random</i>
4	1	1	1	1	1	1	0	1	1	1	1	1	0	0.77
5	1	1	0	1	1	0	1	1	0	0	1	1	1	0.38
	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$X_{10}$	$X_{11}$	$X_{12}$	$X_{13}$	
4	1	1	1	1	1	1	1	1	0	0	1	1	1	
5	1	1	0	1	1	0	0	1	1	1	1	1	0	

**Gambar 4.16** Ilustrasi *Single Point Crossover* pada Seleksi Variabel

Proses pindah silang menggunakan *single point*, di mana dua kromosom dipotong sekali pada batas tertentu dan bagian setelah pemotongan ditukar antar kedua kromosom tersebut. Peluang pindah silang yang digunakan yaitu 0,8. Gambar 4.16 merupakan ilustrasi proses pindah silang terjadi pada kromosom orang tua 4 dan 5. Proses pindah silang dilanjutkan untuk semua pasang kromosom yang memiliki bilangan *random* kurang dari peluang pindah silang.

Setelah dilakukan proses pindah silang, proses selanjutnya dalam *genetic algorithm* adalah proses mutasi. Mutasi yang digunakan adalah mutasi *uniform* yaitu dengan memberikan kesempatan yang sama pada setiap gen untuk dilakukan proses mutasi. Tahapan awal mutasi adalah membangkitkan bilangan *random* pada setiap gen dan membandingkan dengan peluang mutasi sebesar 0,01. Bila nilai bangkitan pada suatu gen kurang dari peluang mutasi, maka proses mutasi dilakukan pada gen tersebut. Proses mutasi yaitu dengan mengubah *allele* 1 menjadi 0 atau 0 menjadi 1 pada gen yang mengalami proses mutasi. Berikut merupakan ilustrasi proses mutasi menggunakan metode mutasi *uniform*.

Kromosom	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$X_{10}$	$X_{11}$	$X_{12}$	$X_{13}$
	0,46	0,11	0,00	0,00	0,19	0,68	0,39	0,42	0,00	0,22	0,90	0,98	0,38
5	1	1	1	1	1	1	0	1	0	1	1	1	0
	MUTASI												
5	1	1	0	0	1	1	0	1	1	1	1	1	0

**Gambar 4.17** Ilustrasi Proses Mutasi pada Seleksi Variabel

Berdasarkan Gambar 4.17 proses mutasi pada kromosom 5, gen yang mengalami proses mutasi terdapat pada gen 3, gen 4 dan gen 9 karena memiliki peluang yang kurang dari 0,01. Gen 3 dan 4 yang awalnya memiliki kode 1 berubah menjadi 0 dan gen 9 yang pada kromosom awal memiliki kode *allele* 0 berubah menjadi 1.

Proses *genetic algorithm* berikutnya adalah proses elitisme untuk mempertahankan kromosom terbaik dalam populasi. Kromosom yang dipertahankan sebesar 20% dari total kromosom dalam populasi yaitu sebanyak 20 kromosom. Proses seleksi variabel *genetic algorithm* dilanjutkan hingga mendapatkan nilai *fitness* yang konvergen.

**Tabel 4.11** Kromosom dan Nilai *Fitness Genetic Algorithm* pada Seleksi Variabel

Populasi	Kromosom				Nilai <i>Fitness</i>
	$X_1$	$X_2$	...	$X_{13}$	
1	0	1	...	1	0,829
2	0	1	...	1	0,829
3	1	0	...	0	0,800
⋮	⋮	⋮	⋮	⋮	⋮
99	1	1	...	1	0,485
100	1	0	...	1	0,484

Hasil seleksi variabel dengan pembagian data *5-fold cross validation* ditunjukkan pada Tabel 4.11. Variabel yang terpilih dilihat dari nilai *fitness* yang paling tinggi. Nilai akurasi yang paling tinggi adalah pada percobaan ke-1 yaitu sebesar 82,857%. *Genetic algorithm* ini dilakukan sebanyak 5 kali. Rata-rata dari nilai ketepatan klasifikasi yang didapatkan menjadi penentu variabel terpilih untuk analisis selanjutnya.

Pada penelitian ini akan dilakukan seleksi dengan kriteria penerimaan yang menjadi nilai *fitness* adalah *AUC* dan Akurasi. Selanjutnya, seleksi variabel dilakukan menggunakan metode GA dengan nilai *fitness* yang digunakan adalah nilai *AUC*. Dari Tabel 4.12, rata-rata tertinggi adalah pada iterasi ke dua dengan ketepatan klasifikasinya sebesar 69,446%. Maka dari itu, variabel yang terpilih dengan metode seleksi variabel *Genetic Algorithm* adalah  $X_2$ ,  $X_3$ ,  $X_6$ ,  $X_8$ ,  $X_9$ ,  $X_{12}$ ,  $X_{13}$  yaitu variabel usia, tekanan darah sistolik, frekuensi nafas, derajat keasaman darah, kadar natrium dalam darah,  $PO_2$ , dan *Blood Urea Nitrogen*. Dapat disimpulkan bahwa variabel-variabel tersebut berpengaruh signifikan terhadap klasifikasi kelas pasien pneumonia.

**Tabel 4.12** *Cross Validation Genetic Algorithm* pada Seleksi Variabel

Percobaan ke-	Variabel Terpilih	Rata-rata <i>AUC</i> (%)
1	$X_2, X_3, X_5, X_6, X_{10}, X_{11}, X_{12}, X_{13}$	51,097
2	<b><math>X_2, X_3, X_6, X_8, X_9, X_{12}, X_{13}</math></b>	<b>51,539</b>
3	$X_2, X_3, X_5, X_8, X_9, X_{10}, X_{11}, X_{13}$	49,766
4	$X_1, X_2, X_5, X_9, X_{10}, X_{11}, X_{12}, X_{13}$	51,373
5	$X_1, X_2, X_3, X_6, X_8, X_9, X_{10}, X_{11}, X_{13}$	51,285

Akurasi pada penelitian diperoleh dengan metode validasi *5-fold Cross-validation* pada data. Tabel 4.13 menunjukkan nilai parameter optimal dan akurasi yang diperoleh dari percobaan yang dilakukan.

**Tabel 4.13** Hasil Percobaan *Grid Search SVM* Seleksi Variabel GA

Percobaan ke-	Parameter		Rata-rata <i>AUC</i> (%)
	<i>C</i>	$\gamma$	
1	$2^3$	$2^{-3}$	<b>57,831</b>
2	$2^3$	$2^{-3}$	48,646
3	$2^0$	$2^{-3}$	51,042
4	$2^0$	$2^{-3}$	47,869
5	$2^1$	$2^{-3}$	54,302

Dari Tabel 4.13, dapat diketahui bahwa dari 5 percobaan yang dilakukan, akurasi tertinggi yaitu pada nilai akurasi 57,831 % diperoleh pada percobaan ke-1. Dari percobaan tersebut, nilai parameter *C* dan  $\gamma$  yang optimum adalah  $2^3$  atau 8 dan  $2^{-3}$  atau 0,125.

Berdasarkan nilai parameter *C* dan  $\gamma$  yang telah didapatkan, fungsi *hyperplane* yang terbentuk untuk klasifikasi pada pasien Pneumonia dengan menggunakan metode *grid search SVM* adalah

$$f(\mathbf{x}) = \sum_{i=1}^n a_i y_i \exp(-0,125 \|\mathbf{x}_i - \mathbf{x}\|^2) + b$$

dengan  $\mathbf{x}$  adalah vektor observasi dari variabel prediktor dan *n* merupakan jumlah observasi. Setelah parameter *C* dan  $\gamma$  yang paling optimal sudah didapatkan, kemudian dibentuk model untuk mengetahui ketepatan klasifikasi pada data. Pembagian data dilakukan dengan menggunakan metode *5-fold cross-validation*. Tabel 4.14 merupakan tabel hasil kinerja klasifikasi dari data pasien pneumonia.

**Tabel 4.14** Hasil Kinerja Klasifikasi dari Data Seleksi Variabel GA

Pembagian data ke-	Data	<i>AUC</i>	<i>G_mean</i>	Akurasi
1	<i>Training</i>	97,979	97,942	97,279
	<i>Testing</i>	65,993	61,535	61,538
2	<i>Training</i>	95,829	95,763	95,270
	<i>Testing</i>	52,083	0,000	65,789

**Tabel 4.14** Hasil Kinerja Klasifikasi dari Data Seleksi Variabel GA (lanjutan)

Pembagian data ke-	Data	<i>AUC</i>	<i>G_mean</i>	Akurasi
3	<i>Training</i>	96,109	96,037	95,946
	<i>Testing</i>	43,229	0,000	63,158
4	<i>Training</i>	96,407	96,364	96,000
	<i>Testing</i>	46,449	0,000	63,889
5	<i>Training</i>	94,080	93,929	93,377
	<i>Testing</i>	53,754	0,000	74,286
Rata-rata	<i>Training</i>	96,081	96,007	95,575
	<i>Testing</i>	<b>52,302</b>	12,307	65,732

Hasil analisis pada Tabel 4.14 menunjukkan bahwa ketepatan klasifikasi pada data *training* adalah 96,081% sedangkan pada data *testing* 52,302%. Dari hasil tersebut, ketepatan klasifikasi dari model SVM yang terbentuk untuk klasifikasi adalah sebesar 65,732% ditinjau dari nilai akurasi, dan 52,302% ditinjau dari nilai *AUC*. Ketepatan klasifikasi *G\_mean* sangat kecil hanya sebesar 12,307% dikarenakan ada hasil prediksi pada kelas ke lima tepat memprediksi observasi aktualnya.

Pada penelitian ini akan dilakukan seleksi dengan kriteria penerimaan yang menjadi nilai *fitness* selain nilai *AUC* nilai *fitness* juga dilihat dari nilai akurasinya. Selanjutnya, seleksi variabel dilakukan menggunakan metode GA dengan nilai *fitness* yang digunakan adalah nilai akurasi. Dari Tabel 4.15, rata-rata tertinggi adalah pada iterasi ke dua dengan ketepatan klasifikasinya sebesar 69,446%. Maka dari itu, variabel yang terpilih dengan metode seleksi variabel *Genetic Algorithm* adalah  $X_2$ ,  $X_3$ ,  $X_6$ ,  $X_7$ ,  $X_8$ ,  $X_9$ ,  $X_{10}$ ,  $X_{11}$ ,  $X_{12}$ ,  $X_{13}$  yaitu variabel usia, tekanan darah sistolik, frekuensi nafas, suhu tubuh, derajat keasaman darah, kadar natrium dalam darah, glukosa dalam darah, hematokrit,  $PO_2$ , dan *Blood Urea Nitrogen*. Dapat disimpulkan bahwa variabel-variabel tersebut berpengaruh signifikan terhadap klasifikasi kelas pasien pneumonia.

**Tabel 4.15** *Cross Validation Genetic Algorithm* pada Seleksi Variabel GA

Percobaan ke-	Variabel Terpilih	Rata-rata Akurasi (%)
1	$X_2, X_4, X_6, X_7, X_9, X_{10}, X_{11}$	60,583
2	$X_2, X_3, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13}$	<b>69,446</b>
3	$X_2, X_3, X_5, X_8, X_9, X_{10}, X_{11}, X_{13}$	65,799
4	$X_1, X_2, X_3, X_8, X_9, X_{10}, X_{13}$	66,432
5	$X_1, X_2, X_3, X_6, X_8, X_9, X_{10}, X_{11}, X_{13}$	67,518

Variabel terpilih dari seleksi variabel GA dengan nilai *fitness AUC* yaitu sebanyak 7 variabel berbeda dengan variabel terpilih dari seleksi variabel GA dengan nilai *fitness* akurasi yaitu sebanyak 10 variabel. Hal ini disebabkan karena perbedaan nilai *AUC* dihitung berdasarkan kelas klasifikasi pasien sedangkan nilai akurasi tidak memperhatikan ketepatan klasifikasi tiap kelas.

Akurasi pada penelitian diperoleh dengan metode validasi *5-fold Cross-validation* pada data. Tabel 4.16 menunjukkan nilai parameter optimal dan akurasi yang diperoleh dari percobaan yang dilakukan.

**Tabel 4.16** Hasil Percobaan *Grid Search SVM* Seleksi Variabel GA

Percobaan ke-	Parameter		Rata-rata Akurasi (%)
	$C$	$\gamma$	
1	$2^1$	$2^{-3}$	54,872
2	$2^1$	$2^{-3}$	57,895
3	$2^1$	$2^{-3}$	66,316
4	$2^1$	$2^{-3}$	67,778
5	<b><math>2^1</math></b>	<b><math>2^{-3}</math></b>	<b>79,429</b>

Dari Tabel 4.16, dapat diketahui bahwa dari 5 percobaan yang dilakukan, akurasi tertinggi yaitu pada nilai akurasi 79,429 % diperoleh pada percobaan ke-5. Dari percobaan tersebut, nilai parameter  $C$  dan  $\gamma$  yang optimum adalah  $2^1$  atau 2 dan  $2^{-3}$  atau 0,125.

Berdasarkan nilai parameter  $C$  dan  $\gamma$  yang telah didapatkan, fungsi *hyperplane* yang terbentuk untuk klasifikasi pada pasien Pneumonia dengan menggunakan metode *grid search SVM* adalah

$$f(\mathbf{x}) = \sum_{i=1}^n a_i y_i \exp(-0,125 \|\mathbf{x}_i - \mathbf{x}\|^2) + b$$

dengan  $\mathbf{x}$  adalah vektor observasi dari variabel prediktor dan  $n$  merupakan jumlah observasi. Setelah parameter  $C$  dan  $\gamma$  yang paling optimal sudah didapatkan, kemudian dibentuk model untuk mengetahui ketepatan klasifikasi pada data. Pembagian data dilakukan dengan menggunakan metode *5-fold cross-validation*. Tabel 4.17 merupakan tabel hasil kinerja klasifikasi dari data pasien pneumonia.

**Tabel 4.17** Hasil Kinerja Klasifikasi dari Data Seleksi Variabel GA

Pembagian data ke-	Data	<i>AUC</i>	<i>G_mean</i>	Akurasi
1	<i>Training</i>	84,257	81,317	94,558
	<i>Testing</i>	45,772	0,000	58,974
2	<i>Training</i>	84,954	84,157	91,216
	<i>Testing</i>	44,792	0,000	57,895
3	<i>Training</i>	89,927	89,798	92,568
	<i>Testing</i>	50,521	0,000	68,421
4	<i>Training</i>	90,470	90,378	92,000
	<i>Testing</i>	50,284	0,000	69,444
5	<i>Training</i>	90,685	90,399	89,404
	<i>Testing</i>	65,057	0,000	82,857
Rata-rata	<i>Training</i>	88,059	87,210	91,949
	<i>Testing</i>	51,285	0,000	<b>67,518</b>

Hasil analisis klasifikasi pada Tabel 4.17 menunjukkan bahwa akurasi dari klasifikasi pada data *training* adalah 91,949 % sedangkan pada data *testing* 67,518%. Dari hasil tersebut, ketepatan klasifikasi dari model SVM yang terbentuk untuk klasifikasi adalah sebesar 67,518% ditinjau dari nilai akurasi, dan 51,285% ditinjau dari nilai *AUC*. Ketepatan klasifikasi *G\_mean* sebesar 0% dikarenakan tidak ada hasil prediksi pada kelas ke lima tepat memprediksi observasi aktualnya.

### 4.3 Optimasi Parameter *Support Vector Machine* dengan *Genetic Algorithm*

Setelah melakukan seleksi variabel dan menentukan parameter SVM yang optimal, langkah selanjutnya adalah mengoptimasi nilai parameter menggunakan metode *Genetic Algorithm* (GA). Optimasi parameter ini bertujuan untuk mendapatkan nilai akurasi yang lebih tinggi. Nilai *fitness* yang digunakan adalah nilai *AUC*.

Gen pertama merupakan parameter diperoleh dari hasil *Grid Search* SVM sebelumnya. Langkah awal dari optimasi ini adalah melakukan inisialisasi kromosom sebanyak 100. Kromosom yang dibangkitkan memiliki dua gen yang terdiri dari parameter  $C$  dan  $\gamma$ . Ilustrasi kromosom pada optimasi ini ditunjukkan pada Gambar 4.18.

Parameter	$C$	$\gamma$
Kromosom	2	0,125

**Gambar 4.18** Ilustrasi Satu Buah Kromosom

Gambar 4.18 merupakan inisialisasi pada kromosom yang pertama. Kromosom yang terbentuk akan mengikuti proses GA untuk optimasi yang meliputi seleksi pindah silang, mutasi, dan elitisme. Nilai *fitness* yang digunakan adalah nilai *AUC*. Setelah dilakukan inisialisasi kemudian membentuk 100 kromosom awal yang dibangkitkan untuk populasi awal yang disajikan pada Tabel 4.18.

**Tabel 4.18** Ilustrasi Kromosom Dalam Populasi Awal pada Optimasi Parameter

Populasi	Kromosom	
	$C$	$\gamma$
1	2,666	0,9683
2	1,511	0,0462
3	1,373	0,7952
$\vdots$	$\vdots$	$\vdots$
99	2,099	1,1025
100	2,287	0,6724

Langkah selanjutnya adalah melakukan evaluasi terhadap masing-masing kromosom yang telah dibentuk menggunakan nilai *fitness* yang telah ditentukan. Nilai *fitness* yang digunakan adalah

*AUC* data *testing*. Tabel 4.19 menyajikan nilai *fitness* untuk masing-masing kromosom pada populasi awal.

**Tabel 4.19** Ilustrasi Kromosom Dalam Populasi Awal pada Optimasi Parameter

Populasi	Kromosom		Nilai
	$C$	$\gamma$	<i>Fitness</i>
1	2,666	0,9683	0,721
2	1,511	0,0462	0,633
3	1,373	0,7952	0,676
⋮	⋮	⋮	⋮
99	2,099	1,1025	0,635
100	2,287	0,6724	0,555

Kemudian langkah selanjutnya adalah pembentukan kromosom orang tua dengan menggunakan metode *Roulette Wheel Selection* (RWS). Ilustrasi untuk tahapan RWS dapat dilihat pada Tabel 4.20.

**Tabel 4.20** Ilustrasi Proses RWS pada Optimasi Parameter

Populasi	<i>Fitness</i>	Proporsi Nilai <i>Fitness</i>	Nilai <i>Fitness</i> Komulatif	<i>Random</i> <i>Number</i>
1	0,721	0,011	0,011	0,017
2	0,633	0,009	0,020	0,317
3	0,676	0,010	0,030	0,587
⋮	⋮	⋮	⋮	⋮
99	0,635	0,009	0,992	0,121
100	0,555	0,008	1,000	0,934

Berdasarkan Tabel 4.20 diketahui bahwa bilangan *random* pertama yaitu sebesar 0.017, maka kromosom orangtua 2 didapatkan dari nilai *fitness* komulatif yang lebih besar dari 0,017 yang pertama yaitu kromosom 2, maka kromosom tersebut menjadi kromosom 1 pada populasi baru. Bilangan *random* diperoleh dengan membangkitkan 100 data berdistribusi *uniform* dengan *range* 0 sampai 1. Tahapan tersebut berhenti bila telah diperoleh 100 kromosom orang tua berdasarkan bilangan *random* yang telah dibangkitkan. Langkah selanjutnya yang harus dilakukan dalam GA adalah pindah silang.

Orangtua	$C$	$\gamma$	Bil.acak
4	2,324	0,7744	0,32
5	1,059	0,2979	0,53
Anak	Pindah silang		
4	2,085	0,6843	
5	1,298	0,3880	

**Gambar 4.19** Pindah Silang Dua Buah Kromosom

Peluang pindah silang yang digunakan yaitu 0,8. Gambar 4.19 merupakan ilustrasi proses pindah silang terjadi pada kromosom orang tua 4 dan 5. Proses pindah silang dilanjutkan untuk semua pasang kromosom yang memiliki bilangan *random* kurang dari peluang pindah silang.

Proses pindah silang menggunakan metode *local arithmetic crossover*. Perhitungan *local arithmetic crossover* adalah (Dumitrescu, 2000)

$$C_4 = \alpha P_4 + (1 - \alpha)P_5 \text{ dan } C_5 = \alpha P_5 + (1 - \alpha)P_4$$

dimana  $P_4$  adalah kromosom orang tua (*parent*) ke-4 dan  $P_5$  adalah kromosom orang tua ke-5,  $C_4$  adalah kromosom anak hasil pindah silang pada kromosom ke-4 dan  $C_5$  adalah kromosom anak hasil pindah silang pada kromosom ke-5, serta  $\alpha$  adalah bilangan acak *uniform* yang bernilai pada *range* 0 sampai 1. Ilustrasi pada Gambar 4.18 diperoleh nilai  $\alpha = 0,811$ , sehingga kromosom anak hasil pindah silang dapat diketahui nilainya seperti ilustrasi Gambar 4.19. Perhitungan hasil *crossover* pada parameter  $C$  pada kromosom ke-5 adalah

$$C_4 = 0,811 \times 2,324 + (1 - 0,811) \times 1,059 = 2,085$$

Perhitungan pindah silang tersebut dilakukan untuk setiap parameter yang terpilih untuk mengikuti proses pindah silang. Setelah dilakukan proses pindah silang, proses selanjutnya dalam *genetic algorithm* adalah proses mutasi. Mutasi yang digunakan adalah mutasi *uniform* yaitu dengan memberikan kesempatan yang sama pada setiap gen untuk dilakukan proses mutasi. Tahapan awal mutasi adalah membangkitkan bilangan *random* pada setiap gen dan membandingkan dengan peluang mutasi sebesar 0,01. Bila nilai bangkitan pada suatu gen kurang dari peluang mutasi, maka proses

mutasi dilakukan pada gen tersebut. Proses mutasi yaitu dengan mengubah *allele* dengan bilangan real yang lain yang masih berada dalam range pada gen yang mengalami proses mutasi. Gambar 4.21 merupakan ilustrasi proses mutasi menggunakan metode mutasi *uniform*.

Kromosom	$C$	$\gamma$
	0,469	0,006
6	2,146	0,6499
	mutasi	
6	2,146	0,7359

**Gambar 4.20** Ilustrasi Proses Mutasi pada Estimasi Parameter

Proses *genetic algorithm* berikutnya adalah proses elitisme untuk mempertahankan kromosom terbaik dalam populasi. Kromosom yang dipertahankan sebesar 5% dari total kromosom dalam populasi yaitu sebanyak 5 kromosom. Proses optimasi parameter dilanjutkan hingga mendapatkan nilai *fitness* yang konvergen.

Optimasi parameter dilakukan untuk persamaan dengan mengikutsertakan semua variabel, variabel terpilih dari seleksi FCBF dan variabel terpilih dari seleksi menggunakan GA.

Analisis klasifikasi pasien pneumonia dengan menggunakan metode klasifikasi SVM. Optimasi parameter yang digunakan adalah dengan metode GA seperti yang sudah dijelaskan sebelumnya. Data dibagi dengan metode *5-fold cross validation*. Tabel 4.21 menunjukkan parameter yang optimum dari masing-masing pembagian data. Batas atas dan batas bawah dari parameter  $C$  adalah 1 dan 3 serta untuk parameter  $\gamma$  dengan *range* 0 sampai 1,125. Batas ini dipilih  $\pm 1$  dari parameter terpilih sebelumnya.

Nilai *fitness* yang digunakan dalam optimasi parameter pada penelitian ini adalah ketepatan klasifikasi dari nilai *AUC* dan nilai akurasi. Pada bagian ini optimasi parameter dengan nilai *fitness AUC* dengan variabel prediktor yang digunakan ada 13 variabel dengan tanpa seleksi variabel.

**Tabel 4.21** Hasil Percobaan SVM-GA Tanpa Seleksi Variabel dengan *Fitness AUC*

Percobaan ke-	Parameter		Rata-rata <i>AUC</i> (%)
	<i>C</i>	$\gamma$	
1	2,6031	0,0505	49,522
2	2,1759	0,1544	54,688
3	<b>1,9266</b>	<b>0,0805</b>	<b>57,813</b>
4	1,0352	0,0635	50,937
5	2,1230	0,0783	54,079

Dari Tabel 4.21, dapat diketahui bahwa dari 5 percobaan yang dilakukan, akurasi tertinggi yaitu pada nilai *AUC* adalah 57,813 % diperoleh pada percobaan ke-3. Dari percobaan tersebut, nilai parameter *C* dan  $\gamma$  yang optimum adalah 1,9266 dan 0,0805.

Berdasarkan nilai parameter *C* dan  $\gamma$  yang telah didapatkan, fungsi *hyperplane* yang terbentuk untuk klasifikasi pada pasien Pneumonia adalah

$$f(\mathbf{x}) = \sum_{i=1}^n a_i y_i \exp(-0,0805 \|\mathbf{x}_i - \mathbf{x}\|^2) + b$$

dengan  $\mathbf{x}$  adalah vektor observasi dari variabel prediktor dan  $n$  merupakan jumlah observasi. Setelah parameter *C* dan  $\gamma$  yang paling optimal sudah didapatkan, kemudian dibentuk model untuk mengetahui kinerja klasifikasi pada data. Pembagian data dilakukan dengan menggunakan metode *5-fold cross-validation*. Tabel 4.22 merupakan tabel hasil ketepatan klasifikasi dari data pasien pneumonia.

**Tabel 4.22** Hasil Kinerja Klasifikasi SVM-GA Data Tanpa Seleksi Variabel

Pembagian data ke-	Data	<i>AUC</i>	<i>G_mean</i>	Akurasi
1	<i>Training</i>	93,940	93,717	97,279
	<i>Testing</i>	47,243	0,000	61,538
2	<i>Training</i>	92,364	92,203	95,270
	<i>Testing</i>	48,438	0,000	63,158
3	<i>Training</i>	85,761	84,926	91,892
	<i>Testing</i>	57,813	0,000	76,316
4	<i>Training</i>	86,460	85,865	92,000
	<i>Testing</i>	48,722	0,000	69,444

**Tabel 4.22** Hasil Kinerja Klasifikasi Dari Data Pasien Pneumonia (lanjutan)

Pembagian data	Data	<i>AUC</i>	<i>G_mean</i>	Akurasi
5	<i>Training</i>	92,530	92,373	91,391
	<i>Testing</i>	50,507	0,000	68,571
Rata-rata	<i>Training</i>	90,211	89,817	93,566
	<i>Testing</i>	<b>50,544</b>	0,000	67,806

Hasil ketepatan dari klasifikasi *AUC* pada data *training* adalah 90,211% sedangkan pada data *testing* 50,544%. Dari hasil tersebut, ketepatan klasifikasi dari model SVM yang terbentuk untuk klasifikasi adalah sebesar 67,806% ditinjau dari nilai akurasi, dan 50,544% ditinjau dari nilai *AUC*. Ketepatan klasifikasi *G\_mean* sebesar 0% dikarenakan tidak ada hasil prediksi pada kelas ke lima tepat memprediksi observasi aktualnya.

Selanjutnya, optimasi parameter dilakukan dengan nilai akurasi sebagai nilai *fitness* pada optimasi parameter.

**Tabel 4.23** Hasil Percobaan SVM-GA Tanpa Seleksi Variabel dengan *Fitness* Akurasi

Percobaan ke-	Parameter		Rata-rata Akurasi (%)
	<i>C</i>	$\gamma$	
1	1,0778	0,04319	66,726
2	2,6164	0,16096	63,955
3	<b>1,7896</b>	<b>0,08900</b>	<b>67,806</b>
4	1,2010	0,02119	66,346
5	2,9628	0,00814	65,878

Dari Tabel 4.23, dapat diketahui bahwa dari 5 percobaan yang dilakukan, akurasi tertinggi yaitu pada nilai akurasi 67.806% diperoleh pada percobaan ke-3. Dari percobaan tersebut, nilai parameter *C* dan  $\gamma$  yang optimum adalah 1,7896 dan 0,08900.

Berdasarkan nilai parameter *C* dan  $\gamma$  yang telah didapatkan, fungsi *hyperplane* yang terbentuk untuk klasifikasi pada pasien Pneumonia dengan menggunakan metode *grid search* SVM adalah

$$f(\mathbf{x}) = \sum_{i=1}^n a_i y_i \exp(-0,089 \|\mathbf{x}_i - \mathbf{x}\|^2) + b$$

dengan  $\mathbf{x}$  adalah vektor observasi dari variabel prediktor dan  $n$  merupakan jumlah observasi. Setelah parameter *C* dan  $\gamma$  yang paling

optimal sudah didapatkan, kemudian dibentuk model untuk mengetahui kinerja klasifikasi pada data. Pembagian data dilakukan dengan menggunakan metode *5-fold cross-validation*. Tabel 4.24 merupakan tabel hasil ketepatan klasifikasi dari data pasien pneumonia.

**Tabel 4.24** Hasil Kinerja Klasifikasi SVM-GA Data Tanpa Seleksi Variabel

Pembagian data ke-	Data	<i>AUC</i>	<i>G_mean</i>	Akurasi
1	<i>Training</i>	93,940	93,717	97,279
	<i>Testing</i>	47,243	0,000	61,538
2	<i>Training</i>	92,364	92,203	95,270
	<i>Testing</i>	48,438	0,000	63,158
3	<i>Training</i>	86,689	85,776	93,243
	<i>Testing</i>	57,813	0,000	76,316
4	<i>Training</i>	91,644	91,550	94,000
	<i>Testing</i>	48,722	0,000	69,444
5	<i>Training</i>	93,311	93,187	92,053
	<i>Testing</i>	51,806	0,000	68,571
Rata-rata	<i>Training</i>	91,590	91,286	94,369
	<i>Testing</i>	50,804	0,000	<b>67,806</b>

Hasil analisis pada Tabel 4.24 menunjukkan bahwa akurasi dari klasifikasi pada data *training* adalah 94,369% sedangkan pada data *testing* 67,806%. Dari hasil tersebut, ketepatan klasifikasi dari model SVM yang terbentuk untuk klasifikasi adalah sebesar 67,806% ditinjau dari nilai akurasi, dan 50,804% ditinjau dari nilai *AUC*. Ketepatan klasifikasi *G\_mean* sebesar 0% dikarenakan tidak ada hasil prediksi pada kelas ke lima tepat memprediksi observasi aktualnya.

Analisis klasifikasi pasien pneumonia dengan menggunakan metode klasifikasi SVM. Variabel yang dipakai pada analisis ini adalah variabel terpilih dari seleksi variabel FCBF. Optimasi parameter yang digunakan adalah dengan metode GA seperti yang sudah dijelaskan sebelumnya. Data dibagi dengan metode *5 fold cross validation*. Tabel 4.25 menunjukkan parameter yang optimum dari masing-masing pembagian data. Batas atas dan batas bawah dari parameter *C* adalah 3 dan 5 serta untuk parameter  $\gamma$  dengan *range* 0 sampai 1,125. Batas ini dipilih  $\pm 1$  dari parameter terpilih sebelumnya.

Nilai *fitness* yang digunakan dalam optimasi parameter pada penelitian ini adalah ketepatan klasifikasi dari nilai *AUC* dan nilai

akurasi. Pada bagian ini optimasi parameter dengan nilai *fitness AUC* dengan variabel prediktor yang digunakan ada 4 variabel dengan seleksi variabel FCBF.

**Tabel 4.25** Hasil Percobaan SVM-GA Seleksi Variabel FCBF dengan *Fitness AUC*

Percobaan ke-	Parameter		Rata-rata <i>AUC</i> (%)
	<i>C</i>	$\gamma$	
1	4,0760	0,1464	48,152
2	4,0232	0,5778	51,613
3	<b>4,2038</b>	<b>0,6579</b>	<b>52,238</b>
4	4,5574	0,3711	50,108
5	3,9409	0,2188	50,586

Dari Tabel 4.25, dapat diketahui bahwa dari 5 percobaan yang dilakukan, akurasi tertinggi yaitu pada nilai akurasi 52,238% diperoleh pada percobaan ke-1. Dari percobaan tersebut, nilai parameter *C* dan  $\gamma$  yang optimum adalah 4,2038 dan 0,6579.

Berdasarkan nilai parameter *C* dan  $\gamma$  yang telah didapatkan, fungsi *hyperplane* yang terbentuk untuk klasifikasi pada pasien Pneumonia dengan menggunakan metode *grid search* SVM adalah

$$f(\mathbf{x}) = \sum_{i=1}^n a_i y_i \exp(-0,6579 \|\mathbf{x}_i - \mathbf{x}\|^2) + b$$

dengan  $\mathbf{x}$  adalah vektor observasi dari variabel prediktor dan  $n$  merupakan jumlah observasi. Setelah parameter *C* dan  $\gamma$  yang paling optimal sudah diapatkan, kemudian dibentuk model untuk mengetahui Kinerja klasifikasi pada data. Pembagian data dilakukan dengan menggunakan metode *5-fold cross-validation*. Tabel 4.26 merupakan tabel hasil kinerja klasifikasi dari data pasien pneumonia.

**Tabel 4.26** Hasil Kinerja Klasifikasi SVM-GA Data Seleksi Variabel FCBF

Pembagian data ke-	Data	<i>AUC</i>	<i>G_mean</i>	Akurasi
1	<i>Training</i>	85,336	84,590	82,993
	<i>Testing</i>	40,135	34,074	43,590
2	<i>Training</i>	82,892	81,513	81,081
	<i>Testing</i>	53,646	0,000	68,421

**Tabel 4.26** Hasil Kinerja Klasifikasi Dari Data Pasien Pneumonia  
(lanjutan)

Pembagian data	Data	<i>AUC</i>	<i>G_mean</i>	Akurasi
3	<i>Training</i>	84,401	83,425	82,432
	<i>Testing</i>	51,563	0,000	68,421
4	<i>Training</i>	80,860	79,401	78,000
	<i>Testing</i>	45,597	0,000	63,889
5	<i>Training</i>	82,138	80,768	80,132
	<i>Testing</i>	70,252	65,299	65,714
Rata-rata	<i>Training</i>	83,126	81,939	80,928
	<i>Testing</i>	<b>52,238</b>	19,875	62,007

Hasil analisis klasifikasi ada Tabel 4.28 menunjukkan bahwa akurasi dari klasifikasi pada data *training* adalah 80,928% sedangkan pada data *testing* 62,007%. Dari hasil tersebut, ketepatan klasifikasi dari model SVM yang terbentuk untuk klasifikasi adalah sebesar 62,007% ditinjau dari nilai akurasi, dan 52,238% ditinjau dari nilai *AUC*. Ketepatan klasifikasi *G\_mean* sebesar 19,857% dikarenakan ada hasil prediksi pada kelas ke lima yang tidak dapat tepat memprediksi observasi aktualnyayaitu pada *fold* ke 2,3, dan 4.

Selanjutnya, optimasi parameter dilakukan dengan nilai akurasi sebagai nilai *fitness* pada optimasi parameter.

**Tabel 4.27** Hasil Percobaan SVM-GA Seleksi Variabel FCBF dengan *Fitness*  
Akurasi

Percobaan ke-	Parameter		Rata-rata Akurasi (%)
	<i>C</i>	$\gamma$	
1	3,4736	0,00559	70,857
2	3,8397	0,66159	68,421
3	4,5584	0,59570	68,421
4	4,3092	0,47595	66,667
5	<b>4,4297</b>	<b>0,12904</b>	<b>79,429</b>

Dari Tabel 4.27, dapat diketahui bahwa dari 5 percobaan yang dilakukan, akurasi tertinggi yaitu pada nilai akurasi 79,429% diperoleh pada percobaan ke-5. Dari percobaan tersebut, nilai parameter *C* dan  $\gamma$  yang optimum adalah 4,4297 dan 0,12904.

Berdasarkan nilai parameter *C* dan  $\gamma$  yang telah didapatkan, fungsi *hyperplane* yang terbentuk untuk klasifikasi pada pasien Pneumonia dengan menggunakan metode *grid search* SVM adalah

$$f(\mathbf{x}) = \sum_{i=1}^n a_i y_i \exp(-0,12904 \|\mathbf{x}_i - \mathbf{x}\|^2) + b$$

Setelah parameter  $C$  dan  $\gamma$  yang paling optimal sudah didapatkan, kemudian dibentuk model untuk mengetahui Kinerja klasifikasi pada data. Pembagian data dilakukan dengan menggunakan metode *5-fold cross-validation*. Tabel 4.28 merupakan tabel hasil kinerja klasifikasi dari data pasien pneumonia.

**Tabel 4.28** Hasil Kinerja Klasifikasi SVM-GA Data Seleksi Variabel FCBF

Pembagian data ke-	Data	AUC	G_mean	Akurasi
1	<i>Training</i>	62,226	58,692	73,469
	<i>Testing</i>	36,397	0,000	51,282
2	<i>Training</i>	52,841	44,207	68,919
	<i>Testing</i>	42,708	0,000	57,895
3	<i>Training</i>	54,210	46,291	68,919
	<i>Testing</i>	46,354	0,000	63,158
4	<i>Training</i>	55,672	51,612	67,333
	<i>Testing</i>	44,886	0,000	63,889
5	<i>Training</i>	50,941	42,321	65,563
	<i>Testing</i>	59,476	0,000	80,000
Rata-rata	<i>Training</i>	55,178	48,625	68,841
	<i>Testing</i>	45,964	0,000	<b>63,245</b>

Hasil analisis pada Tabel 4.28 menunjukkan bahwa akurasi dari klasifikasi pada data *training* adalah 68,841% sedangkan pada data *testing* 63,245%. Dari hasil tersebut, ketepatan klasifikasi dari model SVM yang terbentuk untuk klasifikasi adalah sebesar 63,245% ditinjau dari nilai akurasi, dan 45,964% ditinjau dari nilai *AUC*. Ketepatan klasifikasi *G\_mean* sebesar 0% dikarenakan tidak ada hasil prediksi pada kelas ke lima tepat memprediksi observasi aktualnya.

Analisis klasifikasi pasien pneumonia dengan menggunakan metode klasifikasi SVM. Variabel yang dipakai pada analisis ini adalah variabel terpilih dari seleksi variabel dengan metode GA dengan nilai *fitness AUC* maupun akurasi. Optimasi parameter yang digunakan adalah dengan metode GA seperti yang sudah dijelaskan sebelumnya. Data dibagi dengan metode *5-fold cross validation*.

Nilai *fitness* yang digunakan dalam optimasi parameter pada penelitian ini adalah ketepatan klasifikasi dari nilai *AUC* dan nilai akurasi. Pada bagian ini optimasi parameter dengan nilai *fitness AUC* dengan variabel prediktor yang digunakan ada  $X_2, X_3, X_6, X_8, X_9, X_{12}, X_{13}$  yaitu variabel usia, tekanan darah sistolik, frekuensi nafas, derajat keasaman darah, kadar natrium dalam darah,  $PO_2$ , dan *Blood Urea Nitrogen*. Tabel 4.29 menunjukkan parameter yang optimum dari masing-masing pembagian data. Batas atas dan batas bawah dari parameter  $C$  adalah 7 dan 9 serta untuk parameter  $\gamma$  dengan *range* 0 sampai 1.125. Batas ini dipilih  $\pm 1$  dari parameter terpilih sebelumnya.

**Tabel 4.29** Hasil Percobaan SVM-GA Seleksi Variabel FCBF dengan *Fitness AUC*

Percobaan ke-	Parameter		Rata-rata <i>AUC</i> (%)
	$C$	$\gamma$	
1	8,61407	0,18044	50,853
2	7,38025	0,16997	51,430
3	7,22389	0,04236	50,966
4	7,42294	0,05512	52,789
5	<b>8,04008</b>	<b>0,04901</b>	<b>53,712</b>

Dari Tabel 4.29, dapat diketahui bahwa dari 5 percobaan yang dilakukan, akurasi tertinggi yaitu pada nilai akurasi 53,712% diperoleh pada percobaan ke-5. Dari percobaan tersebut, nilai parameter  $C$  dan  $\gamma$  yang optimum adalah 8,04008 dan 0,04901.

Berdasarkan nilai parameter  $C$  dan  $\gamma$  yang telah didapatkan, fungsi *hyperplane* yang terbentuk untuk klasifikasi pada pasien Pneumonia dengan menggunakan metode *grid search SVM* adalah

$$f(x) = \sum_{i=1}^n a_i y_i \exp(-0,04901 \|x_i - x\|^2) + b$$

dengan  $x$  adalah vektor observasi dari variabel prediktor dan  $n$  merupakan jumlah observasi. Setelah parameter  $C$  dan  $\gamma$  yang paling optimal sudah didapatkan, kemudian dibentuk model untuk mengetahui kinerja klasifikasi pada data. Pembagian data dilakukan dengan menggunakan metode *5-fold cross-validation*. Tabel 4.30

merupakan tabel hasil ketepatan klasifikasi dari data pasien pneumonia.

**Tabel 4.30** Hasil Kinerja Klasifikasi SVM-GA Data Seleksi Variabel GA

Pembagian data ke-	Data	<i>AUC</i>	<i>G_mean</i>	Akurasi
1	<i>Training</i>	89,269	89,102	90,476
	<i>Testing</i>	45,588	0,000	61,538
2	<i>Training</i>	87,545	87,442	89,865
	<i>Testing</i>	52,083	0,000	65,789
3	<i>Training</i>	79,403	78,998	83,784
	<i>Testing</i>	60,417	57,508	68,421
4	<i>Training</i>	79,073	78,803	82,667
	<i>Testing</i>	50,284	0,000	69,444
5	<i>Training</i>	74,922	73,938	81,457
	<i>Testing</i>	60,187	0,000	80,000
Rata-rata	<i>Training</i>	82,043	81,657	85,650
	<i>Testing</i>	<b>53,712</b>	11,502	69,039

Hasil *AUC* dari klasifikasi pada data *training* adalah 81,043% sedangkan pada data *testing* 53,712%. Dari hasil tersebut, ketepatan klasifikasi dari model SVM yang terbentuk untuk klasifikasi adalah sebesar 69,039% ditinjau dari nilai akurasi, dan 53,712% ditinjau dari nilai *AUC*. Ketepatan klasifikasi *G\_mean* sebesar 11,502% dimana nilai tersebut sangat kecil dikarenakan ada hasil prediksi pada kelas ke lima yang tidak tepat memprediksi observasi aktualnya.

Selanjutnya, optimasi parameter dilakukan dengan nilai akurasi sebagai nilai *fitness* pada optimasi parameter.

**Tabel 4.31** Hasil Percobaan SVM-GA Seleksi Variabel GA dengan *Fitness* Akurasi

Percobaan ke-	Parameter		Rata-rata Akurasi (%)
	<i>C</i>	$\gamma$	
1	2,4593	0,08168	70,033
2	1,6009	0,09689	69,520
3	<b>2,5863</b>	<b>0,09340</b>	<b>70,046</b>
4	1,0133	0,14371	67,925
5	1,3563	0,10984	69,520

Dari Tabel 4.31, dapat diketahui bahwa dari 5 percobaan yang dilakukan, akurasi tertinggi yaitu pada nilai akurasi 70,046% diperoleh pada percobaan ke-3. Dari percobaan tersebut, nilai parameter  $C$  dan  $\gamma$  yang optimum adalah 2,5863 dan 0,09340.

Berdasarkan nilai parameter  $C$  dan  $\gamma$  yang telah didapatkan, fungsi *hyperplane* yang terbentuk untuk klasifikasi pada pasien Pneumonia dengan menggunakan metode *grid search SVM* adalah

$$f(\mathbf{x}) = \sum_{i=1}^n a_i y_i \exp(-0,09340 \|\mathbf{x}_i - \mathbf{x}\|^2) + b$$

dengan  $\mathbf{x}$  adalah vektor observasi dari variabel prediktor dan  $n$  merupakan jumlah observasi. Setelah parameter  $C$  dan  $\gamma$  yang paling optimal sudah didapatkan, kemudian dibentuk model untuk mengetahui kinerja klasifikasi pada data. Pembagian data dilakukan dengan menggunakan metode *5-fold cross-validation*. Tabel 4.32 merupakan tabel hasil ketepatan klasifikasi dari data pasien pneumonia.

**Tabel 4.32** Hasil Kinerja Klasifikasi SVM-GA Data Seleksi Variabel GA

Pembagian data ke-	Data	AUC	G_mean	Akurasi
1	<i>Training</i>	92,488	92,282	94,558
	<i>Testing</i>	47,855	0,000	61,538
2	<i>Training</i>	89,664	89,523	92,568
	<i>Testing</i>	55,208	0,000	71,053
3	<i>Training</i>	90,190	90,032	92,568
	<i>Testing</i>	52,083	0,000	71,053
4	<i>Training</i>	89,389	89,314	91,333
	<i>Testing</i>	50,284	0,000	69,444
5	<i>Training</i>	93,458	93,369	92,053
	<i>Testing</i>	59,923	0,000	77,143
Rata-rata	<i>Training</i>	91,038	90,904	92,616
	<i>Testing</i>	53,071	0,000	<b>70,046</b>

Hasil akurasi dari klasifikasi pada data *training* adalah 92,616% sedangkan pada data *testing* 70,046%. Artinya, ketepatan klasifikasi dari model SVM yang terbentuk untuk klasifikasi adalah sebesar 70,046% ditinjau dari nilai akurasi, dan 53,071% ditinjau dari nilai

*AUC*. Ketepatan klasifikasi *G\_mean* sebesar 0% dikarenakan tidak ada hasil prediksi pada kelas ke lima tepat memprediksi observasi aktualnya.

#### 4.4 Perbandingan Nilai Ketepatan Klasifikasi

Tabel 4.33 menunjukkan perbandingan antara klasifikasi dengan metode SVM yang parameternya belum serta yang sudah dioptimasi menggunakan perbandingan *AUC* dan akurasi.

**Tabel 4.33** Perbandingan Nilai Ketepatan Klasifikasi *AUC*

Metode Seleksi Variabel	Variabel	Metode Evaluasi	Nilai Klasifikasi (%)	
			<i>AUC</i> *	Akurasi
Tanpa Seleksi Variabel	$X_1, X_2, X_3, X_4,$	SVM	50,037	67,309
	$X_5, X_6, X_7,$	SVM-GA	50,544	67,806
	$X_8, X_9, X_{10}$ $X_{11}, X_{12}, X_{13}$			
FCBF	$X_1, X_2, X_8,$	SVM	46,135	63,200
	$X_{13},$	SVM-GA	52,238	62,007
<i>Genetic Algorithm</i>	$X_2, X_3, X_6,$	SVM	52,302	65,732
	$X_8, X_9, X_{12},$ $X_{13}$	SVM-GA	<b>53,052</b>	<b>69,039</b>

(\*) sebagai nilai *fitness* pada seleksi variabel dan optimasi GA

Berdasarkan hasil analisis yang telah dilakukan dengan akurasi acuan untuk optimasi adalah nilai *AUC*, dilihat dari nilai ketepatan klasifikasi dari *AUC* yang paling tinggi sebesar 53,052%, dan akurasi sebesar 69,039% yaitu metode yang digunakan adalah SVM-GA *hybrid* dengan variabel prediktor yang digunakan adalah dengan seleksi variabel menggunakan metode GA. Dari analisis ini dapat disimpulkan bahwa optimasi parameter menggunakan GA dapat meningkatkan akurasi data *testing* seperti pada analisis tanpa seleksi variabel akurasi data *testing* meningkat dari 67,309% menjadi 67,806%.

**Tabel 4.34** Perbandingan Nilai Ketepatan Klasifikasi Akurasi

Metode Seleksi Variabel	Variabel	Metode Evaluasi	Nilai Klasifikasi (%)	
			<i>AUC</i>	<b>Akurasi*</b>
Tanpa Seleksi Variabel	$X_1, X_2, X_3, X_4$	SVM	50,037	67,309
	$X_5, X_6, X_7, X_8, X_9, X_{10}$ $X_{11}, X_{12}, X_{13}$	SVM-GA	50,804	67,806
FCBF	$X_1, X_2, X_8,$	SVM	46,135	63,200
	$X_{13},$	SVM-GA	45,964	63,245
<i>Genetic Algorithm</i>	$X_2, X_3, X_6, X_7$	SVM	51,285	67,518
	$X_8, X_9, X_{10},$ $X_{11}, X_{12}, X_{13}$	SVM-GA	<b>53,071</b>	<b>70,046</b>

(\*) sebagai nilai *fitness* pada seleksi variabel dan optimasi GA

Berdasarkan hasil analisis pada Tabel 4.34 dengan akurasi acuan untuk optimasi adalah nilai akurasi, dilihat dari nilai ketepatan klasifikasi dari *AUC* yang paling tinggi sebesar 53,071%, dan akurasi sebesar 70,046% yaitu metode yang digunakan adalah SVM-GA *hybrid* dengan variabel prediktor yang digunakan adalah dengan seleksi variabel menggunakan metode GA. Dari analisis ini dapat disimpulkan bahwa optimasi parameter menggunakan GA dapat meningkatkan akurasi data *testing* seperti pada analisis tanpa seleksi variabel akurasi data *testing* meningkat dari 67,309% menjadi 70,046%.

*(Halaman ini sengaja dikosongkan)*

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan analisis yang telah dilakukan, kesimpulan yang dapat diambil adalah

1. Variabel yang berpengaruh signifikan terhadap klasifikasi kelas pasien pneumonia dengan metode seleksi FCBF adalah jenis kelamin pasien, usia pasien, ukuran keasaman atau pH darah pada darah pasien pneumonia, dan kadar urea nitrogen darah yang menghasilkan evaluasi kinerja klasifikasi *AUC* dan akurasi sebesar 46,135% dan 63,200% sebelum parameternya dioptimasi dan 52,238% dan 63,245%. Variabel yang berpengaruh signifikan terhadap klasifikasi kelas pasien pneumonia menggunakan metode seleksi GA dengan nilai fitness ketepatan klasifikasi *AUC* adalah variabel usia, tekanan darah sistolik, frekuensi nafas, derajat keasaman darah, kadar natrium dalam darah,  $PO_2$ , dan kadar urea nitrogen darah yang menghasilkan ketepatan klasifikasi *AUC* dan akurasi sebesar 52,302%, dan 65,732% sebelum parameternya dioptimasi dan 53,052%, dan 69,309%. Sedangkan variabel yang berpengaruh signifikan terhadap klasifikasi kelas pasien pneumonia menggunakan metode seleksi GA dengan nilai fitness ketepatan klasifikasi akurasi adalah variabel usia, tekanan darah sistolik, frekuensi nafas, suhu tubuh, derajat keasaman darah, kadar natrium dalam darah, glukosa dalam darah, hematokrit,  $PO_2$ , dan kadar urea nitrogen darah yang menghasilkan ketepatan klasifikasi *AUC* dan akurasi sebesar 51,285%, dan 67,518% sebelum parameternya dioptimasi dan 53,071%, dan 70,046%.
2. Nilai evaluasi kinerja klasifikasi *AUC* dan akurasi tanpa seleksi variabel menggunakan metode klasifikasi SVM berturut-turut sebesar 51,285% dan 67,518%. Sedangkan nilai evaluasi kinerja klasifikasi *AUC* dan akurasi dengan parameter yang dioptimasi adalah sebesar 53,052% dan 70,046%.
3. Metode klasifikasi yang menghasilkan nilai akurasi terbesar yaitu metode klasifikasi *support vector machine – genetic algorithm (SVM-GA) hybrid*.

## 5.2 Saran

Berdasarkan penelitian yang telah dilakukan, saran yang diberikan bagi dunia kedokteran adalah metode SVM-GA dapat digunakan untuk klasifikasi kelas risiko pasien pneumonia karena pada penelitian ini metode tersebut menghasilkan nilai akurasi dan *AUC* paling besar dibandingkan dengan metode lainnya.

Saran untuk penelitian selanjutnya adalah menggunakan metode klasifikasi untuk klasifikasi kelas yang mengatasi *imbalance*, atau dapat menambah observasi untuk pasien kelas ke-5 agar prediksi menjadi lebih baik. Saran yang lainnya adalah penelitian selanjutnya dihibau untuk mengumpulkan lebih banyak data agar ketepatan klasifikasinya semakin meningkat.

## DAFTAR PUSTAKA

- Abe, S. (2010). *Support Vector Machines for Pattern Classification 2nd Edition*. London: Springer-Verlag.
- Alonso, A., Noelia, S., & Veronica, B. (2015). *Feature Selection for High-Dimensional Data*. Artificial Intelligence: Foundation, Theory, and Algorithms. Springer International Publishing Switzerland.
- Cortes, & Vapnik. (1995). Support Vector Networks. *Machine Learning I September pp*, 273-297.
- Depkes R.I. (2002). *Pedoman Pemberantasan Penyakit Infeksi Saluran Pernafasan Akut untuk Penanggulangan Pneumonia pada Balita dalam Pelita VI*. Jakarta: Dirjen PPM & PLP.
- Dumitrescu, dkk. (2000). *Evolutionary Computation*. Florida: CRC Press.
- Faihah, R. T. (2010). *Makalah Data Mining Support Vector Machine (SVM)*. Madura: Universitas Trunojoyo.
- Firdausanti, N. A. (2017). *Klasifikasi Kelas Risiko Pasien Pneumonia Menggunakan Metode Hybrid Analisis Diskriminan-Particle Swarm Optimization (ADL-PSO) dan Naive Bayes Classification*. Tugas Akhir. Surabaya: ITS.
- Gen, M., & Cheng, R. (1997). *Genetic Algorithms and Engineering Design*. New York: John Wiley & Sons, Inc.
- Gunawan, dkk. (2012). Evolutionary Neural Network for Othello Game. *Procedia-Social and Behavioral Sciences*, 419-425.
- Gunn, S. (1998). *Support Vector Machine for Classification and Regression*. Southampton: University of Southampton.
- Harafani, H., & Wahono, R. S. (2015). Optimasi Parameter pada Support Vector Machine Berbasis Algoritma Genetika untuk Estimasi Kebakaran Hutan . *Journal of Intelligent Systems*, Vol. 1, No. 2, December 82-90.
- Hidayat, M. R. (2016). *Seleksi Variabel dan Estimasi Parameter Menggunakan Metode Hybrid Regresi Logistik Biner dan Algoritma Genetika*. Tugas Akhir. Surabaya: ITS.

- Jadaan, O. A., Rajamani, L., & Rao, C. R. (2005-2008). Improved Selection Operator for GA. *Journal of Theoretical and Applied Information Technology (JATIT)*, 269-277.
- Kurnianto, I., & Irhamah. (2016). *Seleksi Variabel dan Estimasi Parameter Menggunakan metode Hybrid Analisis Diskriminan Multi-Class dan Algoritma Genetika untuk Klasifikasi*. Surabaya: Institut Teknologi Sepuluh Nopember.
- Kustiyo, A., Firqiani, H. N., & Giri, E. P. (2008). Seleksi Fitur menggunakan Fast Correlation Based Filter pada Algoritma Voting Feature Intervals 5. *Jurnal Ilmiah Ilmu Komputer Edisi 11 Vol. 6 No. 2*, 41-47.
- Kusumaningrum, A. P. (2017). *Optimasi Parameter Support Vector Machine menggunakan Genetic Algorithm untuk Klasifikasi Microarray Data*. Surabaya: ITS.
- Misnadiarly. (2008). *Penyakit Infeksi Saluran Napas Pneumonia pada Balita, Orang Dewasa, Usia Lanjut*. Jakarta: Pustaka Obor Populer.
- PDPI. (2003). *Pneumonia Komuniti Pedoman Diagnosis dan Penatalaksanaan di Indoneia*. Perhimpunan Dokter Paru Indonesia.
- Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. USA: Morgan Kaufman Publishers, Inc.
- Racmatin, S. M. (2017). *Klasifikasi Penderita Pneumonia Menggunakan Metode Analisis Diskriminan, Hybrid Analisis Diskriminan-Algoritma Genetika dan Naive Bayes*. Tugas Akhir. Surabaya: ITS.
- Rusydina, A. W. (2016). *Perbandingan Metode Feature Selection pada High Dimensional Data dan Klasifikasi menggunakan Support Vector Machine (SVM)*. Tugas Akhir. Surabaya: ITS.
- Santosa, B. (2007). *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Saputra, G. K. (2016). *Metode Regresi Logistik Multinomial dan Hybrid Regresi Logistik Multinomial Genetic Algorithm untuk Klasifikasi Jenis Kaca sebagai Barang Bukti Kejahatan*. Tugas Akhir. Surabaya: ITS.
- Setiawan, K. (2003). *Paradigma Sistem Cerdas*. Malang: Bayumedia.

- Sivanandam, S. N., & Deepa, S. N. (2008). *Introduction to Genetic Algorithms*. New York: Springer-Verlag Berlin Heidelberg.
- Sudoyo, A., Setiyohadi, B., Alwi, I., Simadibrata, M., & Setiati, S. (2009). *Buju Ajar Ilmu Penyakit Dalam Jilid II Edisi V*. Jakarta: Interna Publishing.
- Suyanto. (2005). *Algoritma Genetika dalam Matlab*. Yogyakarta: Andi Offset.
- Vapnik, V. N. (1999). *The Nature of Statistical Learning Theory 2nd Edition*. Springer-Verlag: New York Berlin Heidelberg.
- Yasi, H., Prahutama, A., & Utami, T. W. (2014). PREDIKSI HARGA SAHAM MENGGUNAKAN SUPPORT VECTOR REGRESSION DENGAN ALGORITMA GRID SEARCH. *Media Statistika, Vol. 7, No. 1, , 29-35*.
- Zhang, D. L. (2011). Parameter Optimization for Support Vector Regression Based on Genetic Algorithm with Simplex Crossover Operator. *Journal of Information & Computational Science*, 911–920.

*(Halaman ini sengaja dikosongkan)*

**LAMPIRAN****Lampiran 1** Data Penelitian

<b>Nomer</b>	<b>Usia</b>	<b>Sistol</b>	<b>Diastol</b>	<b>...</b>	<b>RR</b>	<b>BUN</b>	<b>Y</b>
1.	24	100	70		24	18	2
2.	47	100	60		30	11	2
3.	50	100	70		24	15	2
4.	42	110	80		28	12	2
5.	44	110	80		30	21	2
⋮	⋮	⋮	⋮		⋮	⋮	⋮
181.	54	120	80		30	12	4
182.	70	127	70		30	43	5
183.	78	110	60		30	33	5
184.	73	150	80		32	15	5
185.	73	115	70		32	88	5
186.	73	140	80		34	44	5

**Lampiran 2** Syntax Cross Validation

```

data = read.csv('E:/dataTAsand.csv', header=TRUE)

crossval=function(data)
{
  library(cvTools)
  nn=ncol(data)
  datax=data[order(data[,nn]),]
  datay=data[,nn]
  n=nrow(datax)
  k=5
  subset=seq(1,n,1)
  nl=matrix(0,4)
  ns=matrix(0,4)
  sn=0
  for(1 in 2:5)
  {
    koding=data[subset[datay==1],]
    nl[1]=nrow(koding)
    sn=sn+nrow(koding)
    ns[1]=sn
  }
  a=1
  fold=0
  for (i in 2:5)
  {
    folds=cvFolds(nl[i],K=k)
    fold[a:ns[i]]=folds$which
    a=a+nl[i]
  }
  datax=cbind(data,fold)
  datatraining=datax[subset[datax[, (nn+1)]!=5],]
  datetesting=datax[subset[datax[, (nn+1)]=5],]
  list(training=datatraining[, (1:nn)],testing=datetesting[, (1:nn)])
}

kol=ncol(data)
dataa=crossval(data)$training[,1:kol-1]
datatest=crossval(data)$testing[,1:kol-1]

#write.csv(dataa,file = 'E:/5data.csv')
#write.csv(datatest,file = 'E:/5datatest.csv')

```

**Lampiran 3** Syntax Tunning Parameter SVM

```

library(e1071)
set.seed(202)

yuhu=ncol(dataa)
dataa.x=dataa[,-yuhu]
dataa.y=dataa[,yuhu]
datatest.x=datatest[,-yuhu]
datatest.y=datatest[,yuhu]

ptm<-proc.time()
cv=5
#menentukan range parameter
range_cost = 2^seq(0,3, by=1)
range_gamma = 2^seq(-3,3, by=1)
hasil = matrix(0,cv,4)
auc= function(model)
{
  pred = predict(model,datatest[,1:(yuhu-1)],type="class")
  Tabel=table(pred=pred,datatest.y)
  r=matrix(0,4,1)
  bobot=matrix(0,4,1)

  for (i in 1:4)
  {
    sum=sum(Tabel[,i])
    r[i]=Tabel[i,i]/sum
    bobot[i]=sum(Tabel[,i])/sum(Tabel[,i])
  }
  hasilauc=sum(r*bobot)*100
}

for (i in 1:cv)
{
  ctrl<- tune.control(sampling="cross", cross=5)
  tune_par<-tune(svm,
    kelas~,
    data=dataa,
    ranges=list(cost=range_cost, gamma=range_gamma),
    scale=FALSE,
    tunecontrol=ctrl)
  model = svm(kelas~,data=dataa,cost = tune_par$best.parameters$cost ,
    gamma = tune_par$best.parameters$gamma )
}

```

**Lampiran 3** Syntax Tuning Parameter SVM (lanjutan)

```

#parameter terbaik
hasil[i,]=c(tune_par$best.parameters$cost,
            tune_par$best.parameters$gamma, auc(model),
            1-tune_par$best.performance)
}

akurasi_average=mean(hasil[,3])
cost.opt=hasil[which.max(hasil[,3]),1]
gamma.opt=hasil[which.max(hasil[,3]),2]
hasil

akurasi_average

proc.time()-ptm

```

**Lampiran 4** Estimasi Parameter SVM

```

set.seed(123)

auc= function(model,dataa)
{
  pred = predict(model,dataa,type="class")
  Tabel=table(pred=pred,dataa$kelas)
  r=matrix(0,4,1)
  bobot=matrix(0,4,1)

  for (i in 1:4)
  {
    sum=sum(Tabel[,i])
    r[i]=Tabel[i,i]/sum
    bobot[i]=sum(Tabel[,i])/sum(Tabel[,i])
  }
  hasilauc=sum(r*bobot)*100
  hasilnya=cbind(r,hasilauc)
  hasilnya
}

```

**Lampiran 4** Estimasi Parameter SVM (lanjutan)

```
xtrain = subset(dataa, select=-kelas)
ytrain = subset(dataa, select=kelas)

xtest = subset(datatest, select=-kelas)
ytest = subset(datatest, select=kelas)

svm.fit.train <- svm(kelas~.,data=dataa,cost = cost.opt , gamma = gamma.opt )
summary(svm.fit.train)

pred.train = predict(svm.fit.train,xtrain,type="class")
pred.test = predict(svm.fit.train,xtest,type="class")

table(ytrain[,-100],pred=pred.train)
table(ytest[,-100],pred=pred.test)

auc(svm.fit.train,dataa)
auc(svm.fit.train,datatest)
```

**Lampiran 5** Seleksi Variabel GA

```
library(car)
library(GA)
library(e1071)

Y_Training=dataa[,14]
X_Training=dataa[,1:13]

Y_Test=datatest[,14]
X_Test=datatest[,1:13]

fitness <- function(string) {
  inc <- which(string == 1)
  X_Training <- X_Training[,inc]
  X_Test<- X_Test[,inc]

  b=cbind(Y_Training,X_Training)

  model<- svm(Y_Training~.,data=b,cost = cost.opt , gamma = gamma.opt )
```

```

pred.test = predict(model,X_Test,type="class")
Tabel.test = table(pred.test,Y_Test)

r.test=matrix(0,4,1)
bobot.test=matrix(0,4,1)
for (i in 1:4)
{
  sum.test=sum(Tabel.test[,i])
  r.test[i]=Tabel.test[,i]/sum.test
  bobot.test[i]=sum(Tabel.test[,i])/sum(Tabel.test[,i])
}

auc=sum(r.test)*100/4
return(auc)
}

gaControl("binary"=list(selection="ga_rwSelection"))
GA <- ga("binary", fitness = fitness, nBits = ncol(X_Training),
  names = colnames(X_Training),
  selection=gaControl("binary")$selection,
  popSize=100, pcrossover=0.8,
  pmutation=0.1,
  elitism=base::max(1,round(100*0.2)),
  monitor = plot,maxiter = 30)

plot(GA)
summary(GA)

f=GA@solution
f=f[1,]
ff=c(1:13)
fff=f*ff
f1=sort(fff)
f2=f1[(14-sum(f)):13]
f2

```

**Lampiran 6** SVM-GA dengan *Fitness AUC*

```

library(e1071)
library(GA)

kolom=ncol(dataa)
Y_Training=dataa[,kolom]
X_Training=dataa[,1:kolom-1]

Y_Test=datatest[,kolom]
X_Test=datatest[,1:kolom-1]

fitnessFunc = function(x)
{
  par_cost = x[1]
  par_gamma = x[2]
  b=cbind(Y_Training,X_Training)

  model<- svm(Y_Training~.,data=b,cost = par_cost , gamma = par_gamma ,
cross = 5)

  pred.test = predict(model,X_Test,type="class")
  Tabel.test = table(pred.test,Y_Test)

  r.test=matrix(0,4,1)
  bobot.test=matrix(0,4,1)
  for (i in 1:4)
  {
    sum.test=sum(Tabel.test[,i])
    r.test[i]=Tabel.test[,i]/sum.test
    bobot.test[i]=sum(Tabel.test[,i])/sum(Tabel.test[,i])
  }

  auc=sum(r.test)*100/4
  return(auc)
}

theta_min = c(p_cost = 7, p_gamma = 0)
theta_max = c(p_cost = 9, p_gamma = 1.125)
gaControl("real-valued"=list(selection="ga_rwSelection",
crossover="gareal_laCrossover",
mutation="gareal_raMutation"))

```

**Lampiran 6** SVM-GA dengan *Fitness AUC* (lanjutan)

```
fitnesvalue = c()
solutions = c()
for(i in 1:5)
{
  result = ga(type="real-valued",
             fitness=fitnessFunc,
             names=names(theta_min),
             min=theta_min,
             max=theta_max,
             selection = gaControl("real-valued")$selection,
             crossover=gaControl("real-valued")$crossover,
             mutation=gaControl("real-valued")$mutation,
             popSize=100,
             maxiter=100,
             run=5,
             maxFitness=100,
             pcrossover=0.8,
             pmutation=0.1,
             monitor=plot)
  summary(result)
  solutions=c(solutions,summary(result)[11])
  fitnesvalue=c(fitnesvalue,summary(result)[10])
}
plot(result)
solutions
fitnesvalue
proc.time()-ptm
```



**Lampiran 7** SVM-GA dengan *Fitness* akurasi (lanjutan)

```
fitnesvalue = c()
solutions = c()
for(i in 1:5)
{
  result = ga(type="real-valued",
             fitness=fitnessFunc,
             names=names(theta_min),
             min=theta_min,
             max=theta_max,
             selection = gaControl("real-valued")$selection,
             crossover=gaControl("real-valued")$crossover,
             mutation=gaControl("real-valued")$mutation,
             popSize=100,
             maxiter=100,
             run=5,
             maxFitness=100,
             pcrossover=0.8,
             pmutation=0.1,
             monitor=plot)
  summary(result)
  solutions=c(solutions,summary(result)[11])
  fitnesvalue=c(fitnesvalue,summary(result)[10])
}
plot(result)
solutions
fitnesvalue
proc.time()-ptm
```

**Lampiran 8 Surat Pernyataan****SURAT PERNYATAAN**

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

Nama : Rahajeng Rindrasari

NRP : 062114 4000 0117

menyatakan bahwa data yang digunakan dalam Tugas Akhir/ Thesis ini merupakan data sekunder yang diambil dari penelitian / buku/ Tugas Akhir/ Thesis/ publikasi lainnya yaitu:

Sumber : Tugas Akhir Neni Alya Firdausanti Mahasiswa Program Studi Sarjana Departemen Statistika ITS Angkatan 2013

Keterangan : Data rekam medis pasien pneumonia di RSUD Dr. Soetomo tahun 2015

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

Mengetahui  
Pembimbing Tugas Akhir

  
(Irhamah, M.Si, Ph.D)  
NIP. 19780406 200112 2 002

Surabaya, 27 Desember 2017

  
(Rahajeng Rindrasari)  
NRP. 062114 4000 0117

\*(coret yang tidak perlu)

*(Halaman ini sengaja dikosongkan)*

## BIODATA PENULIS



Rahajeng Rindrasari atau sering dipanggil Ajeng adalah anak ke-4 dari empat bersaudara, lahir di Semarang, 22 Maret 1996. Penulis telah menempuh pendidikan formal di SD PL Don Bosko Semarang (2002-2008), SMP PL Domenico Savio Semarang (2008-2011), dan SMA Kolese Loyola Semarang (2011-2014) yang melanjutkan studi sarjana di Departemen Statistika Fakultas Matematika, Komputasi, dan Sains Data (FMKSD) Institut Teknologi Sepuluh Nopember (ITS) melalui jalur Kemitraan dan Mandiri. Selama 3.5 tahun masa perkuliahan, penulis merupakan aktifis dalam beberapa organisasi dan kepanitiaan yaitu HIMASTA-ITS Departemen Minat Bakat dan PSDM, UKM Bola Basket Departemen PSDM dan Sekretaris, PKMBK PMK ITS Sie Dana, dan SCC. Selain itu, penulis juga mewakili Departemen, Fakultas, dan Institut dalam kompetisi basket. Segala kritik dan saran serta diskusi lebih lanjut mengenai Tugas Akhir ini dapat dikirimkan melalui surat elektronik (*e-mail*) ke [ajengrindrasari@gmail.com](mailto:ajengrindrasari@gmail.com).