



TUGAS AKHIR - SS141501

**KLASIFIKASI KELAS RISIKO PASIEN PNEUMONIA
MENGUNAKAN REGRESI LOGISTIK ORDINAL,
HYBRID REGRESI LOGISTIK ORDINAL - ALGORITMA
GENETIKA DAN *NAÏVE BAYES CLASSIFICATION***

**RATNA JUWITA
NRP 1314 100 091**

**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 REGRESI LOGISTIK ORDINAL,
HYBRID REGRESI LOGISTIK ORDINAL - ALGORITMA
GENETIKA DAN *NAÏVE BAYES CLASSIFICATION***

**RATNA JUWITA
NRP 1314 100 091**

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

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



FINAL PROJECT - SS141501

**CLASSIFICATION IN PNEUMONIA RISK CLASSES USING
ORDINAL LOGISTIC REGRESSION, HYBRID ORDINAL
LOGISTIC REGRESSION-GENETIC ALGORITHM AND
NAÏVE BAYES CLASSIFICATION**

**RATNA JUWITA
NRP 1314 100 091**

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

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

LEMBAR PENGESAHAN

**KLASIFIKASI KELAS RISIKO PASIEN PNEUMONIA
MENGUNAKAN REGRESI LOGISTIK ORDINAL,
HYBRID REGRESI LOGISTIK ORDINAL -
ALGORITMA GENETIKA DAN NAÏVE BAYES
CLASSIFICATION**

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 :

Ratna Juwita
NRP. 1314 100 091

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

()

Mengetahui,
Kepala Departemen


Dr. Suhartono
NIP. 19710929 199512 1 001

SURABAYA, 24 JANUARI 2018

**KLASIFIKASI KELAS RISIKO PASIEN PNEUMONIA
MENGUNAKAN REGRESI LOGISTIK ORDINAL,
HYBRID REGRESI LOGISTIK ORDINAL –
ALGORITMA GENETIKA DAN NAÏVE BAYES
CLASSIFICATION**

Nama Mahasiswa : Ratna Juwita
NRP : 1314 100 091
Departemen : Statistika
Dosen Pembimbing : Irhamah, M.Si., Ph.D

Abstrak

Pneumonia merupakan salah satu infeksi saluran napas akut yang menjadi penyebab tingginya angka kematian di Indonesia. Pada umumnya tim medis akan melakukan diagnosis awal terhadap pasien kemudian menentukan derajat keparahan pneumonia yang diderita pasien sebagai upaya untuk mengetahui penanganan yang tepat bagi pasien. Penilaian ini menggunakan sistem skor Pneumonia Severty Index (PSI) dan sistem skor CURB-65, namun belum ada acuan yang pasti untuk menentukan sistem penilaian yang akan digunakan. Oleh karena itu perlu dilakukan seleksi variabel untuk menentukan variabel-variabel yang berpengaruh dalam penentuan kelas risiko pneumonia. Pada penelitian ini digunakan seleksi variabel menggunakan metode backward, forward dan stepwise. Selain itu digunakan pendekatan algoritma genetika untuk seleksi variabel dan optimasi parameter pada Regresi Logistik Ordinal dan Naive Bayes Classification untuk mengelompokkan kelas risiko pasien pneumonia di RSUD Dr. Soetomo. Hasil klasifikasi menunjukkan bahwa metode Hybrid Regresi Logistik Ordinal-Algoritma Genetika adalah metode yang paling baik karena menghasilkan nilai ketepatan klasifikasi tertinggi.

Kata Kunci : *Algoritma Genetika, Ketepatan Klasifikasi, Naïve Bayes, Pneumonia, Regresi Logistik Ordinal*

(Halaman ini sengaja dikosongkan)

***CLASSIFICATION IN PNEUMONIA RISK CLASSES
USING ORDINAL LOGISTIC REGRESSION, HYBRID
ORDINAL LOGISTIC REGRESSION - GENETIC
ALGORITHM AND NAÏVE BAYES CLASSIFICATION***

Student Name : Ratna Juwita
Student Number : 1314 100 091
Department : Statistics
Supervisor : Irhamah, M.Si, Ph.D

Abstract

Pneumonia is one of the acute respiratory infections that cause the high mortality rate in Indonesia. In general, the medical team will make an initial diagnosis of the patient then the medical team will assess the degree of severity of pneumonia suffered by the patient in an effort to determine the appropriate treatment for the patient. This assessment uses the Pneumonia Severity Index (PSI) and the CURB-65 score system but there is no definitive reference for determining the system assessment to be used. Therefore, variable selection is needed to determine the variables that influence the determination of pneumonia risk class. This research applies Ordinal Logistic Regression method with selection of forward selection, backward elimination and stepwise method. In addition, genetic algorithm approach is used for variable selection and parameter optimization for Ordinal Logistic Regression and Naïve Bayes Classification to classify risk class of pneumonia patient in RSUD Dr. Soetomo. The classification results show that Hybrid Ordinal Logistic Regression - Genetic Algorithm is the best method because it produces the highest accuracy.

Keywords : Accuracy, Genetic Algorithm, Naïve Bayes, Pneumonia, Ordinal Logistic Regression

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Puji syukur kehadiran Allah SWT tak pernah berhenti terucap atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul “**Klasifikasi Kelas Risiko Pasien Pneumonia Menggunakan Regresi Logistik Ordinal, Hybrid Regresi Logistik Ordinal-Algoritma Genetika dan Naïve Bayes Classification**” dengan lancar.

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

1. Ibu Irhamah, M.Si, Ph.D selaku dosen pembimbing yang telah sabar dalam memberikan bimbingan, saran, dan dukungan selama penyusunan Tugas Akhir
2. Ibu Dr. Dra. Kartika Fithriasari, M.Si dan Ibu Pratnya Paramitha Oktaviana, S.Si, M.Si selaku dosen penguji yang telah memberikan banyak bantuan dan masukan untuk kesempurnaan Tugas Akhir ini.
3. Bapak Sumali, Ibu Komasri dan Setyawan Basuki selaku keluarga penulis yang selalu memberikan dukungan dan kasih sayang, yang menjadi motivasi bagi penulis.
4. Bapak Dr. Suhartono selaku Kepala Departemen Statistika ITS dan Bapak Dr. Sutikno, S.Si, M.Si selaku Ketua Program Strudi Sarjana Statistika ITS yang selalu memberi nasihat dan dukungan selama ini.
5. Dr.rer.pol. Hari Kuswanto, S.Si, M.Si selaku dosen wali yang selalu memberi dukungan dan motivasi selama ini
6. Seluruh dosen Departemen Statistika ITS yang telah memberikan ilmu dan pengetahuan yang tak ternilai harganya, serta segenap karyawan Departemen Statistika ITS.

Penulis berharap hasil Tugas Akhir ini dapat bermanfaat bagi kita semua.

Surabaya, Januari 2018

Penulis

(Halaman ini sengaja dikosongkan)

DAFTAR ISI

HALAMAN JUDUL	i
COVER PAGE	iii
HALAMAN PENGESAHAN	v
Abstrak	vi
Abstract	viii
KATA PENGANTAR	x
DAFTAR ISI	xii
DAFTAR GAMBAR	xiv
DAFTAR TABEL	xvi
DAFTAR LAMPIRAN	xviii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah.....	6
1.3 Tujuan Penelitian.....	6
1.4 Manfaat Penelitian.....	6
1.5 Batasan Penelitian.....	7
BAB II TINJAUAN PUSTAKA	9
2.1 Uji ANOVA Satu Arah.....	9
2.2 Regresi Logistik Ordinal.....	10
2.3 Estimasi Parameter.....	12
2.4 Uji Signifikansi Parameter.....	13
2.5 Algoritma Genetika.....	15
2.6 <i>Naïve Bayes Classification</i>	22
2.7 <i>K-fold Cross Validation</i>	24
2.8 Evaluasi Kinerja Klasifikasi.....	24
2.9 Pneumonia.....	25
BAB III METODOLOGI PENELITIAN	29
3.1 Sumber Data.....	29
3.2 Variabel Penelitian.....	29
3.3 Definisi Operasional.....	30
3.4 Struktur Data Penelitian.....	31
3.5 Langkah Analisis.....	32
BAB IV ANALISIS DAN PEMBAHASAN	41

4.1	Karakteristik Data Pasien Pneumonia di RSUD Dr. Soetomo	41
4.2	Korelasi Antar Variabel Independen	54
4.3	Regresi Logistik Ordinal dengan <i>Backward Elimination</i>	56
4.4	Regresi Logistik Ordinal dengan <i>Forward Selection</i>	60
4.5	Regresi Logistik Ordinal dengan <i>Stepwise Method</i>	60
4.6	Regresi Logistik Ordinal dengan Seleksi Variabel Algoritma Genetika.....	60
4.6	Optimasi Parameter Regresi Logistik Ordinal Hasil Seleksi <i>Backward</i> Menggunakan Algoritma Genetika	69
4.7	Optimasi Parameter Regresi Logistik Ordinal Hasil Seleksi Algoritma Genetika	74
4.8	<i>Naïve Bayes Classification</i> Pasien Pneumonia	77
4.9	Perbandingan Hasil Ketepatan Klasifikasi Regresi Logis-tik Ordinal, <i>Hybrid</i> Regresi Logistik Ordinal- Algoritma Genetika dan <i>Naïve Bayes Classification</i>	80
4.10	Perbandingan Hasil Ketepatan Klasifikasi dengan Metode Sebelumnya	83
BAB V KESIMPULAN DAN SARAN		85
5.1	Kesimpulan	85
5.2	Saran	85
DAFTAR PUSTAKA		87
LAMPIRAN.....		91

DAFTAR GAMBAR

Gambar 2.1	Istilah dalam Algoritma Genetika	16
Gambar 2.2	Ilustrasi Proses Pindah Silang	20
Gambar 2.3	Ilustrasi Proses Mutasi	21
Gambar 3.1	Diagram Alir Penelitian	36
Gambar 3.2	Diagram Alir Klasifikasi menggunakan Regresi Logistik Ordinal	37
Gambar 3.3	Diagram Alir Klasifikasi menggunakan Algoritma Genetika	38
Gambar 3.4	Diagram Alir Klasifikasi menggunakan <i>Naïve</i> <i>Bayes</i>	39
Gambar 4.1	Proporsi Jumlah Pasien pada Tiap Kelas Risiko Pneumonia	41
Gambar 4.2	Proporsi Jumlah Pasien Pneumonia pada Tiap Kelas Risiko berdasarkan Jenis Kelamin	44
Gambar 4.3	Perbedaan Rata-rata Variabel Usia antar Kelas ...	45
Gambar 4.4	Perbedaan Rata-rata Tekanan Darah Sistolik antar Kelas	46
Gambar 4.5	Perbedaan Rata-rata Tekanan Darah Diastolik antar Kelas	47
Gambar 4.6	Perbedaan Rata-rata Denyut Nadi antar Kelas	47
Gambar 4.7	Perbedaan Rata-rata Frekuensi Napas Pasien antar Kelas	48
Gambar 4.8	Perbedaan Rata-rata Suhu Tubuh antar Kelas	49
Gambar 4.9	Perbedaan Rata-rata Analisis Gas Darah Arteri antar Kelas	49
Gambar 4.10	Perbedaan Rata-rata Kadar Natrium dalam Darah antar Kelas	50
Gambar 4.11	Perbedaan Rata-rata Kadar Glukosa dalam Darah antar Kelas	51
Gambar 4.12	Perbedaan Rata-rata Hematokrit antar Kelas	51
Gambar 4.13	Perbedaan Rata-rata Kadar PO ₂ antar Kelas	52
Gambar 4.14	Perbedaan Rata-rata BUN antar Kelas	53

Gambar 4.15	Representasi Kromosom Awal pada Seleksi Variabel	61
Gambar 4.16	Ilustrasi <i>Single Point Crossover</i> pada Seleksi Variabel	63
Gambar 4.17	Ilustrasi Proses Mutasi pada Seleksi Variabel	64
Gambar 4.18	Representasi Kromosom Awal Estimasi Parameter.....	70
Gambar 4.19	Ilustrasi <i>Single Point Crossover</i> pada Estimasi Parameter.....	72
Gambar 4.20	Ilustrasi Proses Mutasi pada Estimasi Parameter.....	73
Gambar 4.21	Representasi Kromosom Awal Estimasi Parameter Hasil Seleksi Algoritma Genetika	74

DAFTAR TABEL

Tabel 2.1	Uji ANOVA Satu Arah.....	9
Tabel 2.2	<i>Confusion Matrix</i>	24
Tabel 2.3	Skor Prediksi CURB-65.....	26
Tabel 2.4	Kelas Risiko dengan Skor CURB-65.....	27
Tabel 2.5	Skor Prediksi PSI.....	27
Tabel 2.6	Kelas Risiko dengan Skor PSI.....	28
Tabel 3.1	Variabel Penelitian.....	29
Tabel 3.2	Struktur Data.....	31
Tabel 4.1	Karakteristik Variabel Independen	42
Tabel 4.2	Hasil Pengujian Beda Rata-rata	42
Tabel 4.3	Korelasi antar Variabel Independen.....	55
Tabel 4.4	Uji Parsial Seleksi <i>Backward</i>	56
Tabel 4.5	Fungsi Logit Seleksi <i>Backward</i>	57
Tabel 4.6	<i>Odds Ratio</i> Seleksi <i>Backward</i>	58
Tabel 4.7	<i>Confusion Matriks</i> Klasifikasi Regresi Logistik Ordinal Hasil Seleksi <i>Backward</i>	59
Tabel 4.8	Ilustrasi Kromosom Dalam Populasi Awal pada Seleksi Variabel.....	61
Tabel 4.9	Ilustrasi Nilai <i>Fitness</i> Setiap Kromosom Populasi Awal Data pada Seleksi Variabel	62
Tabel 4.10	Ilustrasi Proses RWS pada Seleksi Variabel.....	62
Tabel 4.11	Kromosom dan Nilai <i>Fitness</i> Algoritma Genetika ...	65
Tabel 4.12	Uji Parsial Seleksi Algoritma Genetika	66
Tabel 4.13	Fungsi Logit Seleksi Algoritma Genetika.....	66
Tabel 4.14	<i>Odds Ratio</i> Seleksi Algoritma Genetika.....	67
Tabel 4.15	<i>Confusion Matriks</i> Klasifikasi Regresi Logistik Ordinal Hasil Seleksi <i>Backward</i>	69
Tabel 4.16	Ilustrasi Kromosom Dalam Populasi Awal pada Estimasi Parameter	70
Tabel 4.17	Ilustrasi Nilai <i>Fitness</i> Setiap Kromosom Populasi Awal pada Estimasi Parameter	70
Tabel 4.18	Ilustrasi Proses RWS pada Estimasi Parameter	71
Tabel 4.19	Kromosom dan Nilai <i>Fitness</i> Algoritma Genetika ...	73

Tabel 4.20	Fungsi Logit Estimasi Parameter Regresi Logistik Ordinal Hasil Seleksi <i>Backward</i> menggunakan Algoritma Genetika berdasarkan Nilai AUC.....	73
Tabel 4.21	Fungsi Logit Estimasi Parameter Regresi Logistik Ordinal Hasil Seleksi <i>Backward</i> menggunakan Algoritma Genetika berdasarkan Nilai G_mean	74
Tabel 4.22	Ilustrasi Kromosom Dalam Populasi Awal pada Estimasi Parameter Hasil Seleksi Algoritma Genetika.....	75
Tabel 4.23	Kromosom dan Nilai <i>Fitness</i> Generasi Terakhir Hasil Seleksi Algoritma Genetika.....	75
Tabel 4.24	Fungsi Logit Estimasi Parameter Regresi Logistik Ordinal Hasil Seleksi Algoritma Genetika berdasarkan Nilai AUC.....	76
Tabel 4.25	Fungsi Logit Estimasi Parameter Regresi Logistik Ordinal Hasil Seleksi Algoritma Genetika berdasarkan Nilai G_mean	76
Tabel 4.26	<i>Confusion Matriks</i> Klasifikasi Regresi Logistik Ordinal Hasil Seleksi Algoritma Genetika menggunakan AUC.....	76
Tabel 4.27	Nilai Probabilitas <i>Prior</i> Data Training	77
Tabel 4.28	Nilai Peluang Bersyarat Variabel Jenis Kelamin....	78
Tabel 4.29	Rata-rata dan Standar Deviasi Tiap Kategori Kelas Risiko Data <i>Training</i>	78
Tabel 4.30	Peluang Parsial pada Tiap Kategori Data <i>Testing</i> Pertama	78
Tabel 4.31	<i>Confusion Matriks</i> Klasifikasi <i>Naive Bayes</i> Data <i>Testing</i>	80
Tabel 4.32	Ketepatan Klasifikasi Masing-masing Metode.....	82
Tabel 4.33	Perbandingan Ketepatan Klasifikasi dengan Metode Sebelumnya.....	83

DAFTAR LAMPIRAN

Lampiran 1.	Data Pasien Pneumonia	91
Lampiran 2.	<i>Boxplot</i> Data Training pada Seleksi <i>Backward</i> , <i>Forward</i> , dan <i>Stepwise</i>	92
Lampiran 3.	<i>Boxplot</i> Data Training pada Seleksi Algoritma Genetika	94
Lampiran 4.	Tahapan Seleksi Variabel Metode <i>Backward</i>	96
Lampiran 5.	Tahapan Seleksi Variabel Metode <i>Forward</i>	97
Lampiran 6.	Tahapan Seleksi Variabel Metode <i>Stepwise</i>	98
Lampiran 7.	<i>Syntax Cross Validation</i> dan Regresi Logistik Ordinal.....	99
Lampiran 8.	<i>Syntax</i> Seleksi <i>Backward</i>	100
Lampiran 9.	<i>Syntax</i> Seleksi <i>Forward</i>	101
Lampiran 10.	<i>Syntax</i> Seleksi <i>Stepwise</i>	101
Lampiran 11.	<i>Syntax</i> Seleksi Algoritma Genetika.....	103
Lampiran 12.	<i>Syntax</i> Optimasi Algoritma Genetika.....	105
Lampiran 13.	<i>Syntax</i> Memanggil <i>Function</i> Optimasi Algoritma Genetika	107
Lampiran 14.	<i>Syntax Naive Bayes Classification</i>	108
Lampiran 15.	Surat Pernyataan Pengambilan Data	109

(Halaman ini sengaja dikosongkan)

BAB I

PENDAHULUAN

1.1 Latar Belakang

Salah satu masalah utama dalam bidang kesehatan yang sering dijumpai di lingkungan masyarakat adalah Infeksi Saluran Pernapasan Akut (ISPA). ISPA didefinisikan sebagai penyakit saluran pernapasan akut yang disebabkan oleh agen infeksius yang ditularkan dari manusia ke manusia (WHO, 2007). Berdasarkan letak infeksi yang diderita pasien penyakit ini dibedakan menjadi infeksi saluran napas atas dan infeksi saluran napas bawah. Infeksi saluran napas atas meliputi *rhinitis*, *sinusitis*, *faringitis*, *laringitis*, *epiglottitis*, *tonsilitis*, dan *otitis*. Sedangkan pada infeksi saluran napas bawah terdiri atas infeksi pada *bronchus*, *alveoli* seperti *bronkhitis*, *bronkiolitis* dan pneumonia. Infeksi saluran napas yang sering ditemukan pada negara berkembang salah satunya adalah pneumonia. Pneumonia merupakan infeksi di ujung *bronkiol* dan *alveoli* yang disebabkan oleh berbagai patogen seperti bakteri, jamur, virus dan parasit. Pneumonia disebabkan oleh mikroorganisme, yaitu bakteri, virus, *mycoplasma*, *chlamydia* dan jamur. Pneumonia yang disebabkan oleh virus banyak dijumpai pada pasien *immunocompromised*, bayi dan anak. Virus-virus yang menginfeksi adalah virus saluran napas seperti RSV, *Influenza type A*, *parainfluenza*, dan *adenovirus* (Depkes, 2005).

Laporan WHO 1999 menyebutkan bahwa penyebab kematian tertinggi adalah infeksi saluran napas akut salah satunya pneumonia. Berdasarkan *SEAMIC Health Statistic 2001* pneumonia menjadi penyebab kematian nomor 6 di Indonesia. Pada Survei Kesehatan Rumah Tangga Depkes 2001, penyakit infeksi saluran napas bawah menempati urutan kedua sebagai penyebab kematian di Indonesia. Hasil Riskesdas 2013 menyebutkan bahwa *period prevalence* pneumonia berdasarkan diagnosis selama satu bulan sebelum wawancara sebesar 0.2%. Sedangkan berdasarkan diagnosis/ gejala sebesar 1.8%. Jika hasil

ini dibandingkan dengan Riskesdas 2007 yaitu sebesar 2,13% maka *period prevalence* pneumonia pada tahun 2013 mengalami penurunan menjadi 1,8%. Pada balita, *period prevalence* berdasarkan diagnosis sebesar 2.4 per 1000 balita dan berdasarkan diagnosis/ gejala sebesar 18.5 per 1000 balita. Sebagian besar provinsi di Indonesia mengalami penurunan *period prevalence* pneumonia pada tahun 2013 dibandingkan tahun 2007. Tercatat, 11 provinsi di Indonesia (33.3%) yang mengalami kenaikan *period prevalence* pneumonia pada tahun 2013. Menurut umur, *period prevalence* pneumonia tertinggi terjadi pada kelompok umur balita terutama <1 tahun. Menurut daerah tempat tinggal, *period prevalence* pneumonia di pedesaan (2.0%) lebih tinggi dibandingkan dengan daerah perkotaan (1.6%). Menurut status ekonomi dengan menggunakan indeks kepemilikan, semakin rendah status ekonomi maka semakin tinggi *period prevalence* pneumonia (Kemenkes, 2014). Beberapa rumah sakit besar di Indonesia mencatat bahwa kematian yang disebabkan pneumonia masih cukup tinggi. Salah satunya adalah RSUD Dr. Soetomo sebagai rumah sakit terlengkap di Jawa Timur. RSUD Dr. Soetomo sering dijadikan sebagai rumah sakit rujukan untuk beberapa kasus tertentu yang tidak mampu ditangani oleh rumah sakit lainnya di wilayah Jawa Timur maupun Indonesia bagian timur lainnya. Berdasarkan Hasil Survei Kesehatan Rumah Tangga 2001 tercatat 180 kasus pneumonia komunitas ditangani di RSUD Dr. Soetomo dengan angka kematian antara 20-35%. Pneumonia komunitas menduduki peringkat keempat dari sepuluh penyakit terbanyak yang dirawat tiap tahunnya (PDPI, 2003).

Tim Medis membutuhkan waktu beberapa hari untuk mengidentifikasi penyebab dari pneumonia. Hal ini karena penyebab pneumonia cukup sulit untuk diidentifikasi. Apabila tidak segera ditangani, pneumonia dapat menyebabkan kematian pada pasien. Di Amerika, meskipun menggunakan cara invasif penyebab pneumonia hanya bisa ditemukan sebesar 50% (PDPI, 2003). Penyebaran penyakit pneumonia tergolong cepat karena

melalui udara dan mempunyai gejala yang mirip ketika seseorang terserang batuk sehingga menghambat deteksi dini pneumonia pada pasien. Pada umumnya, diagnosis pneumonia komuniti didapatkan dari hasil anamnesis, gejala klinis pemeriksaan fisis, foto toraks dan laboratorium. Setelah dilakukan diagnosis tim medis akan melakukan penilaian terhadap derajat keparahan pneumonia yang diderita pasien sebagai upaya untuk mengetahui penanganan yang tepat bagi pasien. Penilaian ini digunakan untuk mengklasifikasikan pasien berdasarkan tingkat keparahannya. Klasifikasi kelas risiko pneumonia dapat dilakukan menggunakan sistem skor *Pneumonia Severity Index* (PSI) menurut hasil penelitian *Pneumonia Patient Outcome Research Team* (PORT) dan sistem skor CURB-65 yang direkomendasikan oleh *British Thoracic Society* (BTS). Kedua metode tersebut memiliki performansi yang sama dalam mengelompokkan kelas risiko pneumonia. Namun belum dilakukan pengujian terhadap kedua metode tersebut untuk pengelompokkan pada sampel kasus pneumonia yang besar. Beberapa penelitian tentang penyakit pneumonia pernah dilakukan sebelumnya. Santoso, Wulandari dan Purhadi (2012) melakukan penelitian tentang faktor-faktor yang mempengaruhi pneumonia pada balita di Jawa Timur dengan pendekatan *Geographically Weighted Regression*. Noviana, Wulandari dan Purhadi (2013) melakukan permodelan risiko penyakit pneumonia pada balita di Jawa Timur menggunakan Regresi Logistik Biner Stratifikasi.

Salah satu metode statistika yang dapat digunakan untuk mengklasifikasikan kelas risiko pneumonia adalah Regresi Logistik Ordinal. Regresi Logistik Ordinal adalah suatu metode analisis regresi dimana variabel responnya bersifat ordinal dengan variabel prediktor berupa data kategorik maupun kontinu (Hosmer & Lemeshow, 2000). Pada penelitian yang dilakukan oleh Johnson dkk (2014) mengenai prediksi dari perkembangan penyakit *Alzheimer* menggunakan algoritma genetika dengan regresi logistik menunjukkan bahwa hasil akurasi menggunakan algoritma genetika lebih tinggi dibandingkan regresi logistik.

Gradianta dan Irhamah (2014) melakukan penelitian tentang klasifikasi pasien diabetes melitus tipe II menggunakan metode *hybrid* analisis diskriminan–algoritma genetika pada 130 sampel pasien dari tahun 2012 sampai 2013. Hasil yang diperoleh menyimpulkan bahwa ketepatan klasifikasi pada *hybrid* analisis diskriminan–algoritma genetika lebih tinggi dibandingkan metode analisis diskriminan. Selain regresi logistik, metode lain yang dapat digunakan untuk klasifikasi kelas risiko pneumonia adalah *Naïve Bayes*. Metode *Naïve Bayes* merupakan algoritma yang memanfaatkan teori probabilitas yaitu dengan memprediksi probabilitas di masa depan berdasarkan pengalaman sebelumnya (Kusrini, 2009). Fridayanthie (2015) melakukan analisis prediksi penyakit hepatitis menggunakan *Naïve Bayes* dan *Support Vector Machine* (SVM) menyimpulkan bahwa metode *Naïve Bayes* memiliki akurasi yang lebih tinggi dibandingkan SVM. Firdausanti, Irhamah dan Wulandari (2017) melakukan penelitian untuk klasifikasi kelas risiko pneumonia di RSUD Dr. Soetomo menggunakan metode *Hybrid Analisis Diskriminan-Particle Swarm Optimization* (ADL-PSO) dan *Naïve Bayes*. Hasil dari penelitian tersebut menyatakan bahwa metode klasifikasi yang menghasilkan nilai akurasi tertinggi berdasarkan kriteria G_mean dan AUC adalah Analisis Diskriminan dengan *Particle Swarm Optimization* (PSO). Penelitian serupa juga dilakukan oleh Rahimatin, Irhamah, dan Wulandari (2017) menggunakan metode *Hybrid Analisis Diskriminan-Algorithm Genetika* dengan kesimpulan nilai akurasi tertinggi berdasarkan kriteria G_mean dan AUC adalah Analisis Diskriminan dengan Algoritma Genetika.

Pada penelitian ini akan dilakukan klasifikasi kelas risiko pasien pneumonia menggunakan data Tugas Akhir Firdausanti (2017) ditambah dengan variabel jenis kelamin. Pada penelitian tersebut metode yang digunakan adalah analisis diskriminan, *Hybrid Analisis Diskriminan-Particle Swarm Optimization* (PSO) dan *Naïve Bayes Classification*. Hasil dari penelitian tersebut menunjukkan bahwa metode klasifikasi yang menghasilkan nilai

akurasi tertinggi berdasarkan kriteria G_mean dan AUC adalah *Hybrid Analisis Diskriminan-Particle Swarm Optimization* (PSO) yaitu sebesar 67.36% dan 71.63%. Oleh karena itu pada penelitian ini diusulkan beberapa metode lain untuk dibandingkan dengan metode tersebut. Metode yang digunakan adalah regresi logistik ordinal karena belum terdapat penelitian sebelumnya tentang klasifikasi kelas risiko pasien pneumonia menggunakan regresi logistik ordinal. Regresi Logistik Ordinal juga mampu mengakomodir variabel prediktor yang bersifat kontinu dan kategorik sehingga dengan penambahan variabel jenis kelamin diharapkan dapat meningkatkan nilai ketepatan klasifikasi. Regresi logistik merupakan salah satu metode yang mudah diimplementasikan namun seringkali diperlukan seleksi variabel jika variabel prediktor yang digunakan cukup banyak. Hal ini dapat menyebabkan variabel prediktor yang berkorelasi tinggi dengan variabel respon terbuang. Seiring berkembangnya ilmu pengetahuan banyak algoritma optimasi yang dapat dikombinasikan dengan regresi logistik salah satunya adalah algoritma genetika. Kelebihan dari metode algoritma genetika dibandingkan algoritma optimasi lainnya, yaitu kemampuan untuk menangani permasalahan kompleks dan *parallel*. Algoritma genetika 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). Metode lain yang digunakan adalah *Naïve Bayes* karena memiliki beberapa kelebihan di antaranya adalah desain proses yang sederhana karena tidak memerlukan asumsi tertentu yang harus dipenuhi, mudah digunakan untuk mengolah data yang memiliki *range* yang luas dan secara proses komputasi metode *Naïve Bayes* lebih cepat dalam membuat keputusan (Ashari, Paryudi, & Tjoa, 2013). Berdasarkan uraian tersebut, maka pada penelitian ini digunakan pendekatan algoritma genetika untuk seleksi variabel dan optimasi parameter pada Regresi Logistik Ordinal dan

metode *Naïve Bayes Classification* untuk mengelompokkan kelas risiko pasien pneumonia di RSUD Dr. Soetomo.

1.2 Perumusan Masalah

Pneumonia merupakan salah satu masalah kesehatan utama di Indonesia. Penyebab terjadinya pneumonia sulit ditemukan dan memerlukan waktu beberapa hari untuk mendapatkan hasilnya. Jika telat dalam penanganan maka nyawa pasien akan terancam. Berdasarkan hal tersebut maka perlu dilakukan diagnosis terhadap keparahan penyakit pneumonia pada pasien dan tindakan yang akan dilakukan oleh tim medis. Permasalahan utama yang dibahas pada penelitian ini adalah penentuan klasifikasi pasien pneumonia di RSUD Dr. Soetomo pada tahun 2015 menggunakan metode Regresi Logistik Ordinal, *Hybrid* Regresi Logistik Ordinal-Algoritma Genetika dan *Naïve Bayes* serta membandingkan akurasi dari ketiga metode tersebut.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah tersebut, tujuan yang ingin dicapai pada penelitian ini adalah sebagai berikut.

1. Menentukan klasifikasi pasien pneumonia di RSUD Dr. Soetomo pada tahun 2015 menggunakan metode Regresi Logistik Ordinal, *Hybrid* Regresi Logistik Ordinal-Algoritma Genetika dan *Naïve Bayes Classification*.
2. Menentukan dan membandingkan ketepatan klasifikasi pasien pneumonia di RSUD Dr. Soetomo pada tahun 2015 menggunakan metode Regresi Logistik Ordinal, *Hybrid* Regresi Logistik Ordinal-Algoritma Genetika dan *Naïve Bayes Classification* serta membandingkan dengan penelitian sebelumnya.

1.4 Manfaat Penelitian

Hasil penelitian ini diharapkan dapat bermanfaat dalam pengembangan ilmu pengetahuan khususnya Algoritma Genetika dalam bidang kesehatan. Analisis ini diharapkan dapat memberikan tambahan informasi metode yang paling tepat untuk

mengklasifikasikan pasien pneumonia sehingga membantu mempercepat proses klasifikasi pasien dengan tepat dan akurasi yang tinggi.

1.5 Batasan Penelitian

Data yang digunakan merupakan data pasien pneumonia di RSUD Dr. Soetomo pada tahun 2015 dengan kelas risiko pasien pneumonia terdiri atas kelas II, III, IV dan V tanpa kelas I karena tidak ada pasien dengan kelas I yang dirawat pada rentang waktu tersebut. Metode yang digunakan adalah Regresi Logistik Ordinal, *Hybrid* Regresi Logistik Ordinal-Algoritma Genetika dan *Naïve Bayes*. Metode algoritma genetika dilakukan untuk seleksi variabel dan estimasi parameter dengan harapan dapat meningkatkan akurasi klasifikasi. Metode regresi logistik ordinal dan *Naïve Bayes* digunakan untuk membandingkan ketiga metode tersebut karena metode yang kompleks belum tentu menghasilkan akurasi klasifikasi yang tinggi.

(Halaman ini sengaja dikosongkan)

BAB II

TINJAUAN PUSTAKA

2.1 Uji ANOVA Satu Arah

Analysis of Variance (ANOVA) digunakan untuk menguji signifikansi perbedaan rata-rata hitung yang hanya mencakup satu variabel independen. Analisis varians berangkat dari adanya sejumlah variabilitas yang terdapat dalam data kelompok sampel yang akan diuji. Berikut ini hipotesis uji ANOVA satu arah.

$$H_0: \mu_1 = \mu_2 = \dots = \mu_J$$

H_1 : Minimal ada satu $\mu_j \neq \mu_{j'}; j \neq j'; j, j' = 1, 2, 3, \dots, J$
 statistik uji yang digunakan adalah uji F dengan rincian sebagai berikut.

Tabel 2.1 Uji ANOVA Satu Arah

<i>Source of Variation</i>	<i>df</i>	<i>Sum Square</i>	<i>Mean Square</i>	F_0
<i>Treatments</i>	$J-1$	$SSA = n \sum_{j=1}^J \sum_{i=1}^n (\bar{y}_{j.} - \bar{y}_{..})^2$	$s_1^2 = \frac{SSA}{J-1}$	$\frac{s_1^2}{s^2}$
<i>Error</i>	$J(n-1)$	$SSE = \sum_{j=1}^J \sum_{i=1}^n (y_{ji} - \bar{y}_{j.})^2$	$s^2 = \frac{SSE}{J(n-1)}$	
<i>Total</i>	$Jn-1$	$SST = \sum_{j=1}^J \sum_{i=1}^n (y_{ji} - \bar{y}_{..})^2$		

di mana n merupakan banyaknya observasi dan J merupakan banyaknya kelompok. y_{ji} menunjukkan observasi ke- i pada kelas ke- j , $\bar{y}_{j.}$ merupakan rata-rata dari semua observasi pada sampel dari kelas ke- j , dan $\bar{y}_{..}$ merupakan rata-rata dari seluruh observasi pada sampel tersebut Tolak H_0 jika $F_0 > F_{(\alpha, v_1, v_2)}$ (Walpole, Myers, Myers, & Ye, 2011).

2.2 Regresi Logistik Ordinal

Regresi logistik ordinal merupakan salah satu metode statistika untuk menganalisis variabel dependen yang mempunyai skala data ordinal. Variabel independen yang dapat disertakan dalam model bisa berupa data kategori atau kontinu yang berjumlah dua variabel atau lebih. Model yang dapat dipakai untuk regresi logistik adalah *cumulative logit models*. Pada model logit ini sifat ordinal dari respon Y dituangkan dalam peluang kumulatif sehingga *cumulative logit models* merupakan model yang didapatkan dengan membandingkan peluang kumulatif yaitu peluang kurang dari atau sama dengan kategori respon ke- j pada p variabel prediktor yang dinyatakan dalam vektor \mathbf{x} , $P(Y \leq j|\mathbf{x})$ dengan peluang lebih besar dari kategori respon ke- j , $P(Y > j|\mathbf{x})$ (Hosmer & Lemeshow, 2000).

Pada kasus klasifikasi, *cumulative logit model* merupakan fungsi klasifikasi. Jika terdapat J kategori respon maka fungsi klasifikasi yang terbentuk adalah sebanyak $J-1$. Jika $\pi_j(\mathbf{x}) = P(Y = j|\mathbf{x})$ menyatakan peluang kategori respon ke- j pada p variabel prediktor yang dinyatakan dalam vektor \mathbf{x} dan $P(Y \leq j|\mathbf{x})$ menyatakan peluang kumulatif pada p variabel prediktor yang dinyatakan dalam vektor \mathbf{x} maka nilai $\pi_j(\mathbf{x})$ didapatkan dengan persamaan sebagai berikut.

$$P(Y \leq j|\mathbf{x}) = \pi_1(\mathbf{x}) + \pi_2(\mathbf{x}) + \dots + \pi_j(\mathbf{x}) \quad (2.1)$$

di mana $j = 1, 2, \dots, J$ adalah kategori respon. Sesuai dengan definisi *cumulative logit model*, maka didapatkan model sebagai berikut.

$$\begin{aligned} \text{Logit } P(Y \leq j|\mathbf{x}) &= \log \left(\frac{P(Y \leq j|\mathbf{x})}{P(Y > j|\mathbf{x})} \right) \\ &= \log \frac{\pi_1(\mathbf{x}) + \pi_2(\mathbf{x}) + \dots + \pi_j(\mathbf{x})}{\pi_{j+1}(\mathbf{x}) + \dots + \pi_J(\mathbf{x})} \end{aligned} \quad (2.2)$$

Setiap *cumulative logit model* menggunakan semua J kategori respon. Sehingga model simultan yang didapatkan dengan

menggunakan semua *cumulative logit model* adalah sebagai berikut.

$$\begin{aligned} \text{Logit } P(Y \leq j|\mathbf{x}) &= \log \left(\frac{P(Y \leq j|\mathbf{x})}{1-P(Y \leq j|\mathbf{x})} \right) \\ &= \alpha_j + \sum_{k=1}^p \beta_k x_k \end{aligned} \quad (2.3)$$

Fungsi $\pi(x)$ adalah fungsi nonlinier sehingga perlu dilakukan transformasi logit untuk memperoleh fungsi linier yang dapat digunakan untuk melihat hubungan antara variabel respon (y) dengan variabel prediktor (x). Bentuk logit $\pi(x)$ apabila ditransformasikan menghasilkan fungsi $g(x)$ sebagai berikut.

$$g(x) = \ln \left[\frac{\pi_j(x)}{1-\pi_j(x)} \right] = \alpha_j + \beta_1 x_1 + \dots + \beta_p x_p \quad (2.4)$$

Jika terdapat empat kategori respon dimana $j=1,2,3,4$ dengan $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ merupakan nilai pengamatan ke- i , $i=1,2,\dots,n$ dari setiap p variabel prediktor maka nilai dari peluang kategori respon ke- j diperoleh dengan persamaan sebagai berikut.

$$\pi_1(\mathbf{x}_i) = P(Y \leq 1|\mathbf{x}_i) = \frac{\exp(\alpha_1 + \boldsymbol{\beta}^T \mathbf{x}_i)}{1 + \exp(\alpha_1 + \boldsymbol{\beta}^T \mathbf{x}_i)} \quad (2.5)$$

$$\begin{aligned} \pi_2(\mathbf{x}_i) &= P(Y \leq 2|\mathbf{x}_i) - \pi_1(\mathbf{x}_i) \\ &= \frac{\exp(\alpha_2 + \boldsymbol{\beta}^T \mathbf{x}_i)}{1 + \exp(\alpha_2 + \boldsymbol{\beta}^T \mathbf{x}_i)} - \frac{\exp(\alpha_1 + \boldsymbol{\beta}^T \mathbf{x}_i)}{1 + \exp(\alpha_1 + \boldsymbol{\beta}^T \mathbf{x}_i)} \end{aligned} \quad (2.6)$$

$$\begin{aligned} \pi_3(\mathbf{x}_i) &= P(Y \leq 3|\mathbf{x}_i) - \pi_2(\mathbf{x}_i) - \pi_1(\mathbf{x}_i) \\ &= \frac{\exp(\alpha_3 + \boldsymbol{\beta}^T \mathbf{x}_i)}{1 + \exp(\alpha_3 + \boldsymbol{\beta}^T \mathbf{x}_i)} - \frac{\exp(\alpha_2 + \boldsymbol{\beta}^T \mathbf{x}_i)}{1 + \exp(\alpha_2 + \boldsymbol{\beta}^T \mathbf{x}_i)} \end{aligned} \quad (2.7)$$

$$\pi_4(\mathbf{x}_i) = 1 - P(Y \leq 3|\mathbf{x}_i) = 1 - \frac{\exp(\alpha_3 + \boldsymbol{\beta}^T \mathbf{x}_i)}{1 + \exp(\alpha_3 + \boldsymbol{\beta}^T \mathbf{x}_i)} \quad (2.8)$$

Nilai pada Persamaan 2.5-2.8 akan dijadikan pedoman pengklasifikasian. Suatu pengamatan masuk dalam respon kategori- j berdasarkan nilai $\pi_j(\mathbf{x}_i)$ yang terbesar (Agresti, 2002).

2.3 Estimasi Parameter

Metode yang digunakan untuk menaksir α dan β pada regresi logistik ordinal adalah metode *Maksimum Likelihood Estimation* (MLE). Metode ini memperoleh dugaan *maksimum likelihood* bagi α dan β dengan iterasi *Newton Raphson*. Penduga parameter maksimum merupakan penduga yang konsisten dan efisien untuk ukuran sampel yang besar. Bentuk umum dari fungsi *likelihood* untuk sampel dengan n pengamatan bebas, $i = 1, 2, \dots, n$ adalah sebagai berikut.

$$\begin{aligned} l(\theta) &= \prod_{i=1}^n f(y_i) \\ &= \prod_{i=1}^n [\pi_1(\mathbf{x}_i)^{y_{1i}} \pi_2(\mathbf{x}_i)^{y_{2i}} \pi_3(\mathbf{x}_i)^{y_{3i}} \\ &\quad (1 - \pi_1(\mathbf{x}_i) - \pi_2(\mathbf{x}_i) - \pi_3(\mathbf{x}_i))^{1-y_{1i}-y_{2i}-y_{3i}}] \end{aligned} \quad (2.9)$$

Selanjutnya fungsi *log likelihood* yang didapatkan adalah sebagai berikut.

$$\begin{aligned} L(\theta) &= \sum_{i=1}^n \{y_{1i} \log[\pi_1(\mathbf{x}_i)] + y_{2i} \log[\pi_2(\mathbf{x}_i)] + \\ &\quad y_{3i} \log[\pi_3(\mathbf{x}_i)] + (1 - y_{1i} - y_{2i} - y_{3i}) \log[1 - \\ &\quad \pi_1(\mathbf{x}_i) - \pi_2(\mathbf{x}_i) - \pi_3(\mathbf{x}_i)]\} \end{aligned} \quad (2.10)$$

Estimasi dari parameter regresi logistik ordinal didapatkan dengan menurunkan fungsi *log likelihood* terhadap parameter yang akan diestimasi dan disama dengankan nol. Persamaan $\frac{\partial L(\theta)}{\partial \beta_k} = 0$ digunakan untuk estimasi parameter β_k di mana $k=1, 2, \dots, p$ dan $\frac{\partial L(\theta)}{\partial \alpha_j} = 0$ digunakan untuk estimasi parameter α_j di mana $j = 1, 2, \dots, J - 1$. Hasil dari kedua persamaan tersebut merupakan persamaan non linier sehingga untuk menyelesaikannya harus dilakukan iterasi hingga diperoleh α dan β yang konvergen, kemudian penyelesaian dilakukan menggunakan iterasi *Newton Raphson*. Iterasi *Newton Raphson* digunakan untuk menyelesaikan persamaan yang non linear.

$$\boldsymbol{\theta}^{(z+1)} = \boldsymbol{\theta}^{(z)} - [\mathbf{H}(\boldsymbol{\theta}^{(z)})]^{-1} \mathbf{g}(\boldsymbol{\theta}^{(z)}), z = 0, 1, 2, \dots \quad (2.11)$$

Langkah-langkah iterasi *Newton Raphson* adalah sebagai berikut

1. Menentukan nilai awal estimasi parameter $\hat{\boldsymbol{\theta}}^{(0)}$
2. Membentuk vektor gradien \mathbf{g} dan matriks *Hessian* \mathbf{H}

3. Memasukan nilai $\hat{\theta}^{(0)}$ pada elemen \mathbf{g} dan \mathbf{H} sehingga diperoleh $\mathbf{g}\hat{\theta}^{(0)}$ dan $\mathbf{H}\hat{\theta}^{(0)}$
4. Iterasi mulai $z = 0$ menggunakan Persamaan (2.11). Nilai $\hat{\theta}^{(z)}$ merupakan sekumpulan penaksir parameter yang konvergen pada iterasi ke z
5. Apabila belum diperoleh estimasi parameter yang konvergen, maka langkah 3 diulang kembali hingga nilai $\|\hat{\theta}^{(z+1)} - \hat{\theta}^{(z)}\| \leq \varepsilon$, dengan ε merupakan bilangan yang sangat kecil. Hasil estimasi yang diperoleh adalah $\hat{\theta}^{(z+1)}$ pada iterasi terakhir.

Adapun vektor gradien $\mathbf{g}(\theta)$ adalah vektor dengan elemen turunan parsial pertama dari fungsi *log likelihood* terhadap parameter yang akan diestimasi. Matriks *Hessian* $\mathbf{H}(\theta)$ merupakan matriks non singular dengan elemen-elemen matriksnya adalah turunan parsial kedua dari fungsi *log likelihood* terhadap parameter yang akan diestimasi, dan z adalah banyaknya iterasi dengan susunan sebagai berikut.

$$\mathbf{g} = \begin{bmatrix} \frac{\partial L(\theta)}{\partial \alpha_1} \\ \frac{\partial L(\theta)}{\partial \alpha_2} \\ \vdots \\ \frac{\partial L(\theta)}{\partial \beta_p} \end{bmatrix}; \mathbf{H} = \begin{bmatrix} \frac{\partial^2 L(\theta)}{\partial \alpha_1^2} & \frac{\partial^2 L(\theta)}{\partial \alpha_1 \partial \alpha_2} & \cdots & \frac{\partial^2 L(\theta)}{\partial \alpha_1 \partial \beta_p} \\ \frac{\partial^2 L(\theta)}{\partial \alpha_2 \partial \alpha_1} & \frac{\partial^2 L(\theta)}{\partial \alpha_2^2} & \cdots & \frac{\partial^2 L(\theta)}{\partial \alpha_2 \partial \beta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 L(\theta)}{\partial \beta_p \partial \alpha_1} & \frac{\partial^2 L(\theta)}{\partial \beta_p \partial \alpha_2} & \cdots & \frac{\partial^2 L(\theta)}{\partial \beta_p^2} \end{bmatrix}$$

2.4 Uji Signifikansi Parameter

Model regresi logistik ordinal yang telah terbentuk perlu diuji signifikansi dari parameternya. Pengujian dilakukan dengan uji statistik untuk memastikan adanya hubungan yang nyata antara variabel-variabel prediktor dengan variabel respon. Pengujian yang dilakukan adalah meliputi uji serentak dan uji parsial.

Uji Serentak

Pengujian ini dilakukan untuk mengetahui signifikansi parameter β terhadap variabel respon secara bersama-sama dengan menggunakan statistik uji G atau *likelihood ratio test*. Dimisalkan terdapat empat kategori variabel respon dengan variabel prediktor sebanyak p .

Hipotesis

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

H_1 : minimal ada satu $\beta_k \neq 0$ dengan $k=1,2,\dots,p$

Statistik uji :

$$G = -2 \ln \frac{\left[\frac{n_1}{n}\right]^{\pi_1} \left[\frac{n_2}{n}\right]^{\pi_2} \left[\frac{n_3}{n}\right]^{\pi_3} \left[\frac{n_4}{n}\right]^{\pi_4}}{\prod_{i=1}^n \hat{\pi}_{1i}^{y_{1i}} \hat{\pi}_{2i}^{y_{2i}} \hat{\pi}_{3i}^{y_{3i}} (1 - \hat{\pi}_{1i} - \hat{\pi}_{2i} - \hat{\pi}_{3i})^{1-y_{1i}-y_{2i}-y_{3i}}} \quad (2.12)$$

Keterangan:

n_1 = Banyaknya observasi pada kategori 1

n_2 = Banyaknya observasi pada kategori 2

n_3 = Banyaknya observasi pada kategori 3

n_4 = Banyaknya observasi pada kategori 4

$$n = n_1 + n_2 + n_3 + n_4$$

Kriteria penolakan H_0 apabila $G > \chi^2_{(p,\alpha)}$ atau $p\text{-value} < \alpha$

Uji Parsial

Pengujian ini dilakukan untuk mengetahui signifikansi parameter β terhadap variabel respon secara parsial dengan menggunakan statistik uji *Wald*.

Hipotesis :

$$H_0: \beta_k = 0$$

$$H_1: \beta_k \neq 0$$

Dengan $k=1,2,\dots,p$

Statistik uji :

$$W = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)} \quad (2.13)$$

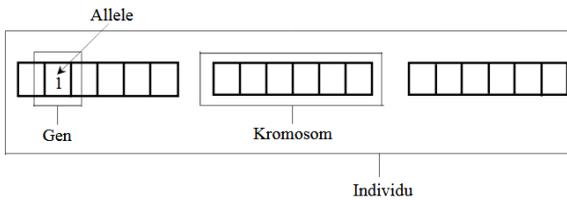
Di mana $\hat{\beta}_k$ merupakan penaksir parameter β_k dan $SE(\hat{\beta}_k)$ adalah taksiran standar *error*. Statistik uji W mengikuti distribusi normal sehingga akan diperoleh keputusan tolak H_0 apabila $|W| > Z_{\frac{\alpha}{2}}$ atau jika *p-value* $< \alpha$ (Hosmer & Lemeshow, 2000).

2.5 Algoritma Genetika

Algoritma genetika ditemukan oleh John Holland dan kawan-kawannya pada tahun 1975 di Universitas Michigan. Konsep dari algoritma genetika terinspirasi dari teori evolusi yang dikemukakan oleh Darwin. Konsep yang dikembangkan dalam algoritma ini adalah menemukan cara dalam melakukan teknik pencarian untuk mendapatkan solusi aproksimasi berdasarkan pada proses evolusi alam. Pada proses evolusi akan didapatkan individu-individu harus mampu menyesuaikan diri dengan lingkungannya agar mampu bertahan hidup. Individu-individu ini telah secara berulang kali mengalami perubahan gen untuk dapat menyesuaikan diri dengan lingkungan hidupnya. Perubahan gen ini terjadi melalui proses perkembangbiakan, dalam algoritma genetika, proses perkembangbiakan ini merupakan dasar pemikiran dalam mendapatkan anak yang lebih baik. Konsep yang dibuat John Holland adalah ketahanan, dan juga keseimbangan antara efisiensi dan kemenangan sistem untuk bertahan hidup dalam setiap kondisi (Gunawan, dkk., 2012).

Algoritma genetika memiliki beberapa kelebihan, salah satunya yaitu mampu mengatasi berbagai jenis fungsi obyektif dari berbagai kromosom. Algoritma genetika juga adaptif dan mudah dikombinasikan dengan metode lain (Gen & Cheng, 1997). Kelebihan lain dari algoritma genetika dibandingkan algoritma optimasi lainnya, yaitu kemampuan untuk menangani permasalahan kompleks dan parallel. Algoritma genetika 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 algoritma genetika yang didasarkan pada ilmu genetik menyebabkan istilah-istilah yang digunakan dalam algoritma genetika banyak diadaptasi dari ilmu tersebut. Beberapa istilah dalam algoritma genetika disajikan pada Gambar 2.1. Individu dalam populasi disebut string atau genotip. Pada algoritma genetika, masing-masing individu hanya ada pada kromosom. Kromosom terdiri atas gen / karakter / *decoder* yang tersusun linier. Posisi yang ditempati oleh gen pada kromosom disebut lokus. Nilai yang terkandung dalam gen disebut *allele*. Tipe data *allele* bisa berupa biner, *floating point*, atau integer tergantung representasi genetik yang digunakan. Sementara gabungan *allele* bisa memberi nilai pada kromosom yang disebut fenotip (Gunawan, dkk., 2012).



Gambar 2.1 Istilah dalam Algoritma Genetika

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

1. Menentukan Representasi Genetik

Bentuk kromosom harus ditentukan paling awal karena setiap *allele* yang bergabung harus disediakan tempat penampungannya. Sebagai contoh, 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

Terdapat beberapa parameter yang perlu diberi nilai, antara lain sebagai berikut.

a. Ukuran populasi

Ukuran populasi adalah banyaknya individu yang terdapat di dalam populasi. 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

Jumlah generasi juga mempunyai andil yang besar dalam menemukan individu yang lebih baik, karena 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 identik dengan *parent*.

Pada metode algoritma genetika, 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
- ii. *Real number encoding*: nilai gen berada dalam interval $[0, R]$, dengan R adalah bilangan real positif dan biasanya $R=1$.
- iii. *Discrete desimal 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 bertahan hidup dan sebaliknya individu yang memiliki nilai *fitness* rendah tidak dapat bertahan hidup. Fungsi *fitness* yang digunakan dalam penelitian ini adalah ketepatan klasifikasi, 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 menurut (Jadaan, Rajamani, & Rao, 2005-2008) adalah sebagai berikut ini.

- i. Menghitung nilai *fitness* masing-masing kromosom v_i .
- ii. Menghitung total nilai *fitness* dalam populasi.

$$F = \sum_{h=1}^{n_{pop}} f(h) \quad (2.14)$$

- iii. Menghitung proporsi masing-masing kromosom.

$$P_h = \frac{f(h)}{F} \quad (2.15)$$

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

$$S_h = \sum_{q=1}^h P_q \quad (2.16)$$

- 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 5 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_q : nilai *fitness* kumulatif kromosom q ke- h

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]

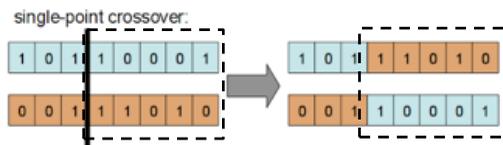
Orang tua 2 : [1 0 1 0 0 0 1 1]

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 algoritma genetika yang sudah pernah dilakukan sebaiknya nilai probabilitas pindah silang tinggi, yaitu antara 0.8-0.9 agar memberikan hasil yang baik. Ilustrasi terjadinya pindah silang dapat dilihat pada Gambar 2.2.

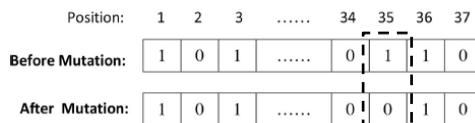


Gambar 2.2 Ilustrasi Proses Pindah Silang

5. Mutasi

Mutasi digunakan untuk mencegah algoritma yang terjebak pada solusi lokal 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 lokal 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, 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 2.3.



Gambar 2.3 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 algoritma genetika 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 $N \times G$ 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. Tidak menutup kemungkinan penghapusan individu dilakukan pada semua individu dalam populasi tersebut.

2.6 Naïve Bayes Classification

Naïve Bayes adalah salah satu metode klasifikasi probabibilistik paling sederhana yang didasarkan pada teorema *Bayes*. *Naïve Bayes* merupakan metode klasifikasi populer dan masuk dalam sepuluh algoritma terbaik dalam data mining (Wu & Kumar, 2009). Diketahui A_1, A_2, \dots, A_p merupakan kejadian independen di dalam ruang sampel Ω sehingga $\sum_{k=1}^p A_k = \Omega$. Kejadian A_1, A_2, \dots, A_p merupakan partisi dalam Ω , dan B merupakan suatu kejadian acak dari peristiwa $A_1 \cap B, A_2 \cap B, \dots, A_p \cap B$ yang membentuk partisi dalam B .

$$P(B) = \sum_{k=1}^p P(A_k \cap B) \quad (2.17)$$

Jika $P(A_i) > 0$ pada $k = 1, 2, \dots, p$ maka $P(A_k \cap B) = P(B|A_k)P(A_k)$

$$P\{B\} = \sum_{k=1}^p P\{B|A_k\}P\{A_k\} \quad (2.18)$$

Kejadian acak B dan A_1, A_2, \dots, A_p merupakan partisi dari ruang sampel Ω . Apabila $P\{B\} > 0$ dan $P\{A_k\} > 0$ untuk $k = 1, 2, \dots, p$ maka

$$P\{A_k|B\} = \frac{P\{B|A_k\}P\{A_k\}}{\sum_{k=1}^p P\{B|A_k\}P\{A_k\}} \quad (2.19)$$

$P\{A_k|B\}$ disebut *posterior probability* karena nilainya bergantung pada nilai B , $P\{A_k\}$ merupakan *prior probability* karena nilainya tidak bergantung pada B , dan $P\{B|A_k\}$ merupakan fungsi *likelihood* dan $P\{B\}$ merupakan keterangan (*evidence*).

Diberikan $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p\}$ merupakan variabel yang digunakan untuk menentukan kelas \mathbf{y} . Perhitungan *posterior probability* untuk setiap kelas \mathbf{y}_j menggunakan teorema Bayes adalah sebagai berikut.

$$P(\mathbf{y}_j|\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p) = \frac{P(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p|\mathbf{y}_j) \cdot P(\mathbf{y}_j)}{P(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)} \quad (2.20)$$

Kelas yang terpilih adalah kelas yang memaksimalkan nilai $P(\mathbf{y}_j|\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)$ atau memaksimalkan nilai dari peluang $P(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p|\mathbf{y}_j) \cdot P(\mathbf{y}_j)$. Berdasarkan Persamaan 2.20 maka diperlukan perhitungan $P(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p|\mathbf{y}_j)$. Setiap variabel diasumsikan saling bebas untuk kelas \mathbf{y} .

$$P(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p|\mathbf{y}_j) = P(\mathbf{x}_1|\mathbf{y}_j) \cdot P(\mathbf{x}_2|\mathbf{y}_j) \dots P(\mathbf{x}_p|\mathbf{y}_j) \quad (2.21)$$

Jika terdapat variabel yang bersifat kuantitatif atau kontinu, maka $P(\mathbf{x}_k|\mathbf{y}_j)$ dihitung menggunakan pendekatan distribusi normal.

$$P(\mathbf{x}_k|\mathbf{y}_j) = \frac{1}{\sigma_{kj}\sqrt{2\pi}} \exp\left(-\frac{(\mathbf{x}_k - \mu_{kj})^2}{2\sigma_{kj}^2}\right) \quad (2.22)$$

Estimasi peluang $P(\mathbf{x}_k|\mathbf{y}_j)$ dapat dihitung untuk setiap variabel \mathbf{x}_k dan kelas \mathbf{y}_j sehingga data baru akan dapat diklasifikasikan ke dalam kelas \mathbf{y}_j apabila peluangnya lebih besar dibandingkan yang lainnya. Pada *Naïve Bayes* digunakan *Hypothesis*

Maximum A Posterior (HMAP) untuk memaksimalkan nilai probabilitas dari masing-masing kelas dengan rumus sebagai berikut (Gorunescu,2011).

$$H_{MAP} = \arg \max \frac{P(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p | \mathbf{y}_j) \cdot P(\mathbf{y}_j)}{P(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)} \quad (2.23)$$

2.7 K-fold Cross Validation

Pada *k-fold cross validation* data akan dipartisi secara acak menjadi k bagian atau *folds* yaitu D_1, D_2, \dots, D_k dengan masing-masing ukuran yang hampir sama. Validasi menggunakan *training* dan *testing* dilakukan sebanyak k kali. Pada iterasi ke- i , partisi D_i akan diatur sebagai data *testing* dan partisi yang tersisa lainnya akan digunakan sebagai data *training* untuk memperoleh model. Artinya, pada iterasi yang pertama, partisi D_1, D_2, \dots, D_k akan menjadi data *training* untuk mendapatkan model yang pertama yang akan diuji dengan data pada partisi D_1 . Pada iterasi kedua partisi D_1, D_2, \dots, D_k akan menjadi data *training* kemudian D_2 akan menjadi data *testing*, begitu seterusnya (Han, Kamber, & Pei, 2012).

2.8 Evaluasi Kinerja Klasifikasi

Evaluasi kinerja klasifikasi adalah suatu evaluasi yang melihat peluang kesalahan klasifikasi berdasarkan kriteria atau ukuran yang digunakan. Pada klasifikasi empat kelompok, *confusion matrix* yang terbentuk adalah seperti pada Tabel 2.2.

Tabel 2.2 *Confusion Matrix*

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}

Ketepatan klasifikasi pada data *unbalance* dapat dihitung menggunakan *Geometric mean* (G_mean). Pada pengelompokkan

dengan lebih dari dua kelas menggunakan *Expanding G_mean* (Bekkar, Djemaa, & Alitouche, 2013). Selain *G_mean*, ketepatan klasifikasi dapat dihitung menggunakan *Area Under ROC (Receive Operating Characteristic) Curve* (AUC). Kurva ini menggambarkan probabilitas dengan variabel sensitivitas dan kekhususan (*specifity*) dengan batas nilai 0 sampai 1. Berikut ini adalah rumus untuk menghitung *G_mean* dan AUC.

$$G_mean = \left(\prod_{j=1}^J R_j \right)^{1/J} \quad (2.24)$$

$$AUC = \frac{1}{J} \sum_{j=1}^J R_j \quad (2.25)$$

$$\text{Dengan } R_j = \frac{n_{jj}}{n_j}; j = 1, 2, \dots, J \quad (2.26)$$

Keterangan:

n_j : jumlah observasi pada kelas j

n_{jj} : jumlah prediksi kelas j yang tepat diklasifikasikan ke kelas j

2.9 Pneumonia

Pneumonia merupakan infeksi di ujung bronkiol dan alveoli yang disebabkan oleh berbagai patogen seperti bakteri, jamur, virus dan parasit. Pneumonia disebabkan oleh mikroorganisme, yaitu bakteri, virus, *mycoplasma*, *chlamydia* dan jamur. Pneumonia yang disebabkan oleh virus banyak dijumpai pada pasien *immunocompromised*, bayi dan anak. Virus-virus yang menginfeksi adalah virus saluran napas seperti RSV, *Influenza type A*, *parainfluenza*, dan *adenovirus* (Depkes, 2005).

Pneumonia menjadi penyebab kematian tertinggi pada balita dan bayi serta menjadi penyebab penyakit umum terbanyak. Pneumonia dapat terjadi sepanjang tahun dan dapat menyerang semua usia. Manifestasi klinik menjadi sangat berat pada pasien dengan usia sangat muda, manula serta pada pasien dengan kondisi kritis (Depkes, 2005). Berdasarkan hasil Riskesdas 2007,

pneumonia merupakan penyakit penyebab kematian kedua tertinggi setelah diare. Hal ini mengindikasikan bahwa pneumonia merupakan masalah utama kesehatan masyarakat yang berkontribusi terhadap tingginya kematian balita di Indonesia.

Pada saat melakukan diagnosis terhadap pasien pneumonia maka perlu dilakukan anamnesis yaitu gambaran klinis dari pasien dan pemeriksaan fisik. Selain itu, dilakukan pemeriksaan penunjang yang meliputi gambaran radiologis dan pemeriksaan laboratorium. Pemeriksaan penunjang dilakukan untuk memperkuat diagnosis awal dari tim medis guna memastikan keadaan pasien. Hasil diagnosis akan digunakan dalam penentuan perawatan yang tepat bagi pasien. Klasifikasi pasien yang dilakukan pihak rumah sakit biasanya menggunakan sistem skor. Sistem skor digunakan untuk menentukan level risiko pasien penyakit pneumonia. Beberapa sistem skor yang digunakan dalam dunia medis adalah sistem *Pneumonia Severity Index* (PSI) yang dikembangkan oleh *Pneumonia Patient Outcome Research Team* (PORT) dan sistem CURB-65 yang merupakan rekomendasi dari *British Thoracic Society* (BTS).

Sistem CURB-65 merupakan sistem yang mudah dan praktis digunakan untuk penilaian. Sistem skor ini sudah divalidasi meskipun menggunakan sampel yang lebih sedikit dibandingkan PSI. Kelebihan dari sistem ini adalah penggunaannya yang mudah dirancang untuk menilai keparahan penyakit dibandingkan menilai pasien pneumonia dengan risiko kematian. Variabel yang terdapat dalam penentuan skor prediksi menggunakan sistem CURB-65 dan jumlah poin tiap variabel disajikan dalam Tabel 2.3.

Tabel 2.3 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

Penjumlahan poin dari CURB-65 akan menghasilkan skor bagi pasien dan digunakan untuk mengelompokkan pasien berdasarkan tingkat keparahan seperti disajikan pada Tabel 2.4.

Tabel 2.4 Kelas Risiko dengan Skor 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

Tabel 2.5 Skor Prediksi 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 kongestif	+10
Penyakit serebrovaskuler	+10
Penyakit ginjal	+10
Pemeriksaan fisis	
Perubahan status mental	+20
Pernafasan > 30 kali per menit	+20
Tekanan darah sistolik \leq 90 mmHg	+20
Suhu tubuh < 35 ^o atau > 40 ^o	+15
Nadi \geq 125 kali per menit	+10
Hasil laboratorium atau radiologi	
Analisis gas darah arteri : pH < 7.35	+30
BUN (<i>Blood Urea Nitrogen</i>) > 30 mgdL	+20
Natrium < 130 mEq/liter	+20
Glukosa > 250 mgdL	+10
Hematokrit < 30%	+10
PO ₂ \leq 60 mmHg	+10
Efusi pleura	+10

Sistem PSI ditetapkan oleh Persatuan Dokter Paru Indonesia (PDPI) menjadi acuan dalam diagnosis pneumonia. Variabel yang terdapat dalam penentuan skor prediksi menggunakan sistem PSI dan jumlah poin tiap variabel disajikan dalam Tabel 2.5. Langkah selanjutnya adalah menjumlahkan poin-poin pada hasil PORT. Hasil penjumlahan tersebut kemudian dikategorikan menurut kelas risikonya seperti pada Tabel 2.5.

Tabel 2.6 Kelas Risiko dengan Skor PSI

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

BAB III

METODOLOGI PENELITIAN

3.1 Sumber Data

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari Tugas Akhir Firdausanti (2017) tentang Klasifikasi Kelas Risiko Pneumonia menggunakan Metode *Hybrid Analisis Diskriminan-Particle Swarm Optimization* dan *Naïve Bayes Classification*. Data merupakan data rekam medis pasien pneumonia pada tahun 2015 di RSUD Dr. Soetomo.

3.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah sebagai berikut.

Tabel 3.1 Variabel Penelitian

Variabel	Keterangan	Skala
Y	Kelas risiko pneumonia (II/III/IV/V)	Ordinal
X ₁	Jenis kelamin 1 = perempuan 2 = laki-laki	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 napas 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	Presentase sel darah merah terhadap volume darah/ Hematokrit (%)	Rasio
X ₁₂	Tekanan oksigen dalam darah/ kadar PO ₂ (mmHg)	Rasio
X ₁₃	Urea Nitrogen darah (mg/dL)	Rasio

3.3 Definisi Operasional

Adapun penjelasan dari variabel-variabel yang digunakan dalam penelitian ini adalah sebagai berikut.

- a. Tekanan darah adalah tekanan yang ditimbulkan pada dinding arteri. Tekanan puncak terjadi saat ventrikel berkontraksi dan disebut tekanan sistolik. Tekanan diastolik adalah tekanan terendah yang terjadi saat jantung beristirahat. Tekanan darah biasanya digambarkan sebagai rasio tekanan sistolik terhadap tekanan diastolik, dengan nilai dewasa normalnya berkisar dari 90/60 sampai 140/90. Rata-rata tekanan darah normal biasanya 120/80 (Smeltzer & Bare, 2001).
- b. Frekuensi denyut nadi merupakan banyaknya denyut nadi yang terjadi setiap menit. Denyut nadi normal orang dewasa adalah 60-100 kali/menit.
- c. Frekuensi napas menunjukkan tingkat keseringan pasien dalam bernafas tiap menit dengan frekuensi normal yaitu berada pada interval 12-20 kali/menit.
- d. Suhu tubuh adalah ukuran dari kemampuan tubuh dalam menghasilkan dan menyingkirkan hawa panas. Nilai normal untuk suhu oral manusia adalah 37°C (98.6°F), tetapi pada sebuah penelitian terhadap orang-orang muda normal, rata-rata suhu oral pagi hari adalah 36.7°C .
- e. Analisa gas darah dilakukan untuk mengevaluasi status oksigen dan karbondioksida di dalam darah arteri dan mengukur nilai pH pasien. pH normal bagi seseorang adalah 7.35.
- f. Natrium memiliki fungsi utama membantu mengendalikan kadar air dalam tubuh. Kadar natrium normal dalam tubuh adalah antara 135-145 *milliequivalents per liter* (mEq/L).
- g. Gula darah adalah istilah yang mengacu kepada tingkat glukosa di dalam darah. Umumnya tingkat gula darah bertahan pada batas-batas yang sempit sepanjang hari 70-150 mg/dl.
- h. Hematokrit adalah jumlah sel darah merah dalam darah sehingga dengan melakukan pemeriksaan hematokrit maka akan didapatkan hasil perbandingan jumlah sel darah merah (eritrosit) terhadap volume darah dalam satuan persen. Nilai

normal hematokrit pada pria dewasa berkisar pada nilai 38.8% - 50%, sedangkan pada wanita dewasa adalah 34.9% - 44.5%.

- i. PO_2 adalah tekanan gas O_2 dalam darah dengan kadar normal sebesar 80-100 mmHg.
- j. Ureum adalah salah satu produk dari pemecahan protein dalam tubuh yang disintesis di hati dan 95% dibuang oleh ginjal dan sisanya 5% dalam feses. Secara normal kadar ureum dalam darah adalah 7 – 25 mg dalam 100 mililiter darah. Kadar ureum di luar negeri sering disebut sebagai *Blood Urea Nitrogen* (BUN) (Nursalam, 2006).

3.4 Struktur Data Penelitian

Struktur data yang digunakan dalam penelitian ini disajikan pada Tabel 3.2.

Tabel 3.2 Struktur Data

Kelas	Sampel	Respon	Variabel Prediktor			
		Y	X_1	X_2	...	X_{13}
2	1	$y_{2.1}$	$x_{2.1.1}$	$x_{2.1.2}$...	$x_{2.1.13}$
	2	$y_{2.2}$	$x_{2.2.1}$	$x_{2.2.2}$...	$x_{2.2.13}$
	⋮	⋮	⋮	⋮	⋮	⋮
	n_2	$y_{2.n_2}$	$x_{2.n_2.1}$	$x_{2.n_2.2}$...	$x_{2.n_2.13}$
3	1	$y_{3.1}$	$x_{3.1.1}$	$x_{3.1.2}$...	$x_{3.1.13}$
	2	$y_{3.2}$	$x_{3.2.1}$	$x_{3.2.2}$...	$x_{3.2.13}$
	⋮	⋮	⋮	⋮	⋮	⋮
	n_3	$y_{3.n_3}$	$x_{3.n_3.1}$	$x_{3.n_3.2}$...	$x_{3.n_3.13}$
4	1	$y_{4.1}$	$x_{4.1.1}$	$x_{4.1.2}$...	$x_{4.1.13}$
	2	$y_{4.2}$	$x_{4.2.1}$	$x_{4.2.2}$...	$x_{4.2.13}$
	⋮	⋮	⋮	⋮	⋮	⋮
	n_4	$y_{4.n_4}$	$x_{4.n_4.1}$	$x_{4.n_4.2}$...	$x_{4.n_4.13}$
5	1	$y_{5.1}$	$x_{5.1.1}$	$x_{5.1.1}$...	$x_{5.1.13}$
	2	$y_{5.2}$	$x_{5.2.1}$	$x_{5.2.2}$...	$x_{5.2.13}$
	⋮	⋮	⋮	⋮	⋮	⋮
	n_5	$y_{5.n_5}$	$x_{5.n_5.1}$	$x_{5.n_5.2}$...	$x_{5.n_5.13}$

Keterangan:

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

3.5 Langkah Analisis

Berikut ini adalah langkah analisis yang digunakan dalam klasifikasi pasien pneumonia di RSUD Dr. Soetomo.

1. Mendeskripsikan karakteristik pasien pneumonia di RSUD Dr. Soetomo.
2. Membentuk data dengan *5-fold Cross Validation*. Observasi pada kelas V hanya berjumlah 8 observasi sehingga jumlah *fold* yang mungkin untuk digunakan adalah 5.
3. Melakukan klasifikasi menggunakan regresi logistik ordinal dengan metode seleksi variabel menggunakan metode *forward selection*, *backward elimination*, *stepwise* dan algoritma genetika.
 - a. *Backward elimination*, dengan memasukkan semua variabel independen ke dalam model kemudian melakukan eliminasi satu persatu variabel yang tidak signifikan berdasarkan nilai AIC.
 - b. *Forward selection* dilakukan dengan memasukkan satu persatu variabel independen ke dalam model secara bertahap.
 - c. *Stepwise method* dilakukan dengan memasukkan variabel independen menggunakan metode *forward selection*, dan akan dieliminasi menggunakan *backward elimination*.
 - d. Langkah-langkah seleksi variabel menggunakan algoritma genetika adalah sebagai berikut.
 1. 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 mem-

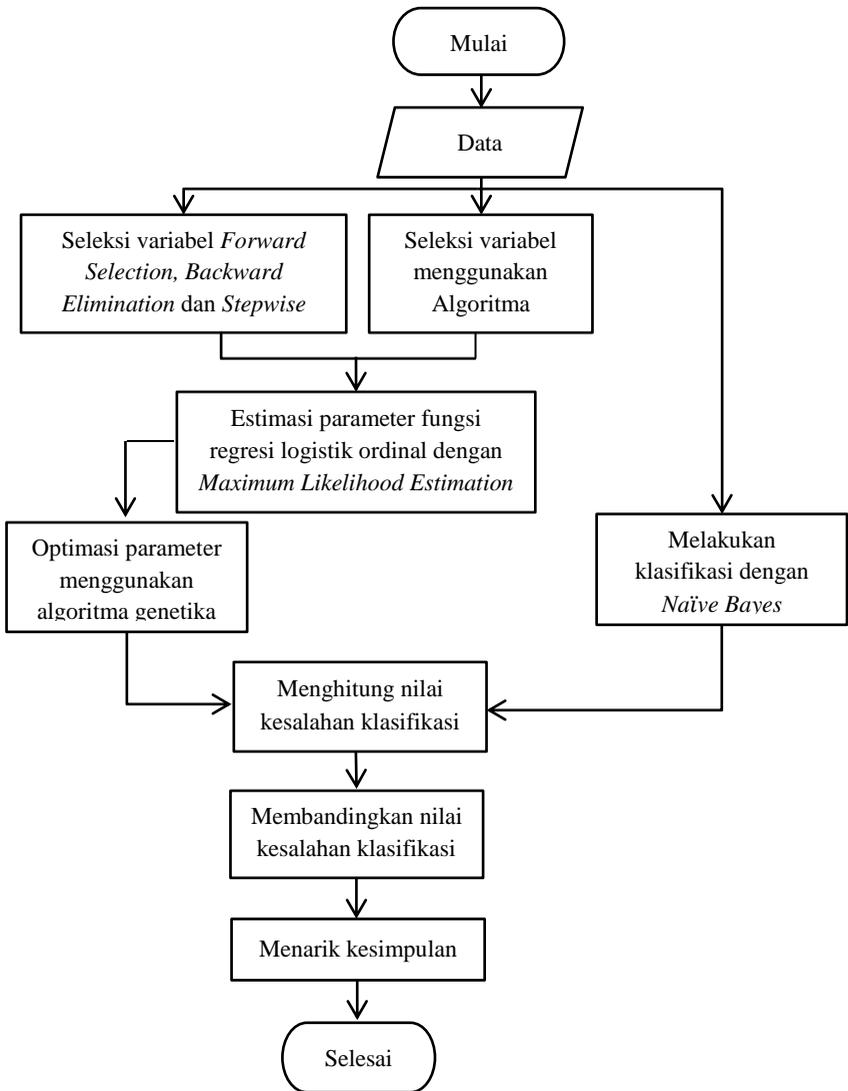
- bangkitkan 100 individu secara *random*. Salah satu individu diisi dengan parameter yang telah diperoleh dari fungsi regresi logistik.
2. Mengevaluasi masing-masing kromosom menggunakan nilai *fitness*. Pada penelitian ini yang menjadi nilai *fitness* adalah nilai ketepatan klasifikasi.
 3. Melakukan proses seleksi sebanyak 100 kromosom dari sejumlah 100 induk yang berasal dari populasi menggunakan seleksi *roulette wheel*.
 4. Melakukan proses pindah silang jika nilai bilangan *random* r antara $[0,1]$ yang dibangkitkan kurang dari probabilitas pindah silang $(P_s) = 0.8$.
 5. Melakukan proses mutasi jika nilai bilangan *random* r antara $[0,1]$ yang dibangkitkan kurang dari nilai probabilitas proses mutasi $(P_m) = 0.1$.
 6. Melakukan proses elitisme dimana dua kromosom dengan nilai *fitness* terbaik akan bertahan ke generasi selanjutnya.
 7. Melakukan pergantian populasi lama dengan generasi baru dengan cara memilih kromosom terbaik berdasarkan nilai *fitness* yang telah melalui proses seleksi, pindah silang dan elitisme.
 8. 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 algoritma genetika diulang dari langkah 3-6.
4. Melakukan analisis menggunakan metode regresi logistik ordinal dengan langkah-langkah sebagai berikut.
1. Menghitung estimasi parameter kemudian membuat model regresi logistik ordinal.
 2. Melakukan uji signifikansi parameter secara serentak dan parsial menggunakan Persamaan (2.12) dan Persamaan (2.13).

3. Melakukan interpretasi model kemudian menghitung ketepatan klasifikasi.
5. Memilih parameter regresi logistik ordinal menggunakan *hybrid* regresi logistik ordinal-algoritma genetika yang menghasilkan ketepatan klasifikasi tertinggi.
 1. Merepresentasikan model dari regresi logistik ordinal ke dalam kromosom dan menentukan nilai inisialisasi. Inisialisasi dilakukan dengan membangkitkan 100 individu secara *random* dimana salah satu individu diisi dengan parameter yang diperoleh dari model regresi logistik ordinal. Selain itu inisialisasi kromosom awal juga menggunakan nilai bilangan *random*.
 2. Menentukan nilai *fitness* untuk setiap kromosom yang didapatkan dari nilai ketepatan klasifikasi. Pada penelitian ini yang menjadi nilai objektif merupakan parameter pada model regresi logistik ordinal.
 3. Melakukan proses seleksi sebanyak 100 kromosom dari sejumlah 100 induk yang berasal dari populasi dengan seleksi *roulette wheel*.
 4. Melakukan proses pindah silang jika nilai bilangan *random r* antara $[0,1]$ yang dibangkitkan kurang dari probabilitas pindah silang (P_s) = 0.8.
 5. Melakukan proses mutasi jika nilai bilangan *random r* antara $[0,1]$ yang dibangkitkan kurang dari probabilitas proses mutasi (P_m) = 0.1.
 6. Melakukan proses elitisme dimana dua kromosom dengan nilai *fitness* terbaik akan bertahan ke generasi selanjutnya.
 7. Melakukan pergantian populasi lama dengan generasi baru dengan cara memilih kromosom terbaik berdasarkan nilai *fitness*-nya yang telah melalui proses seleksi, pindah silang dan elitisme.
 8. Melakukan pengecekan terhadap solusi yang telah didapatkan. Solusi dikatakan telah mencapai kriteria

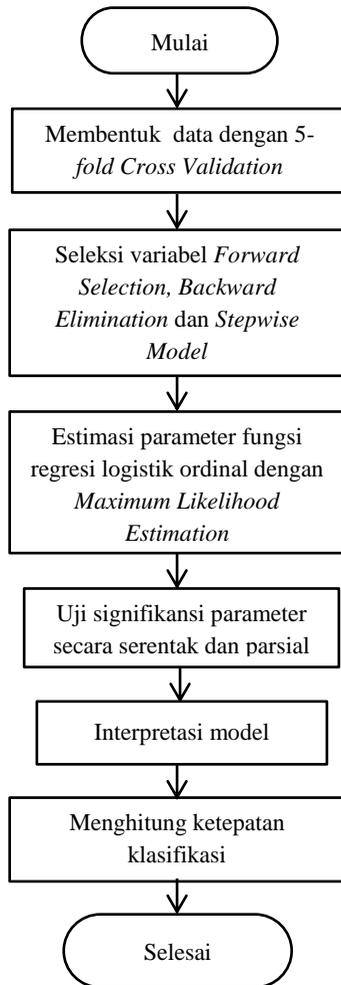
apabila nilai *fitness* terbaik telah konvergen, apabila kondisi ini belum terpenuhi maka proses algoritma genetika diulang dari langkah 3-6.

9. Menghitung ketepatan klasifikasi.
6. Melakukan klasifikasi data menggunakan metode *Naïve Bayes Classification* dengan langkah-langkah analisis sebagai berikut.
 1. Menghitung nilai probabilitas untuk setiap variabel k , $k = 1, 2, \dots, p$. Apabila data berupa data numerik (kontinu), maka akan dihitung nilai rata-rata, varian dan deviasi standar dari masing-masing variabel pada setiap kategori. Selanjutnya adalah menentukan nilai probabilitas menggunakan pendekatan distribusi normal.
 2. Menentukan probabilitas akhir dari semua parameter terhadap setiap kategori.
 3. Menentukan kelompok kategori berdasarkan nilai probabilitas yang tertinggi.
 4. Menghitung ketepatan klasifikasi.
 5. Membandingkan variabel terpilih dari hasil seleksi variabel dan nilai ketepatan klasifikasi dari hasil analisis metode regresi logistik ordinal dan *hybrid* regresi logistik ordinal-algoritma genetika.
7. Membandingkan ketepatan klasifikasi antara ketiga metode yaitu regresi logistik ordinal, *hybrid* regresi logistik ordinal-algoritma genetika, dan *Naïve Bayes Classification* serta membandingkan ketepatan klasifikasi dengan metode sebelumnya.
8. Menarik kesimpulan dan membuat saran untuk penelitian selanjutnya.

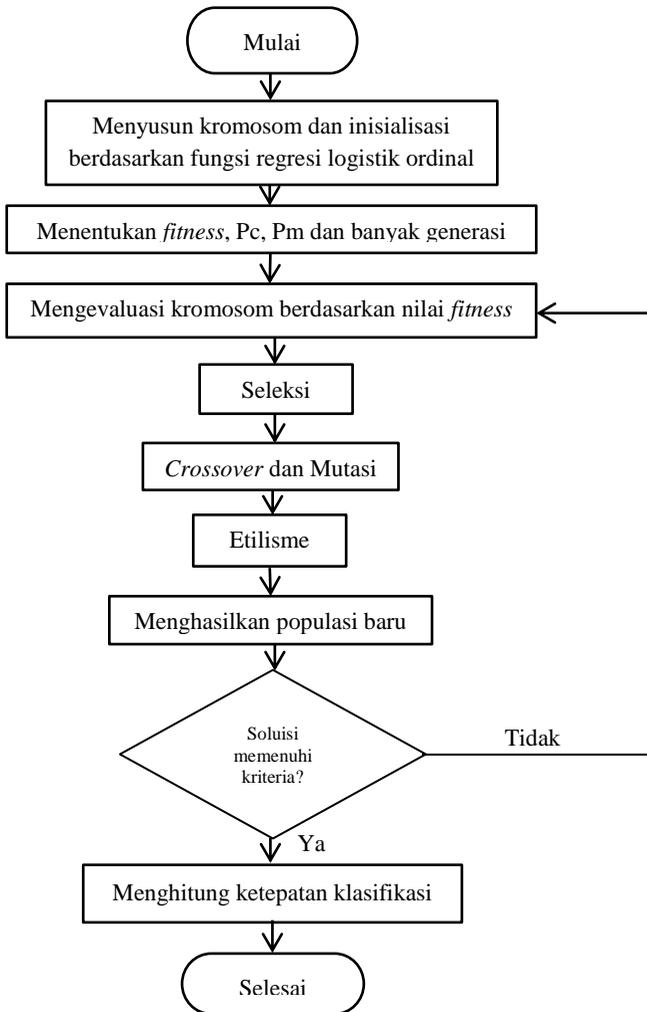
Langkah-langkah analisis tersebut dapat digambarkan dalam diagram alir sebagaimana yang ditampilkan pada Gambar 3.1-3.4.



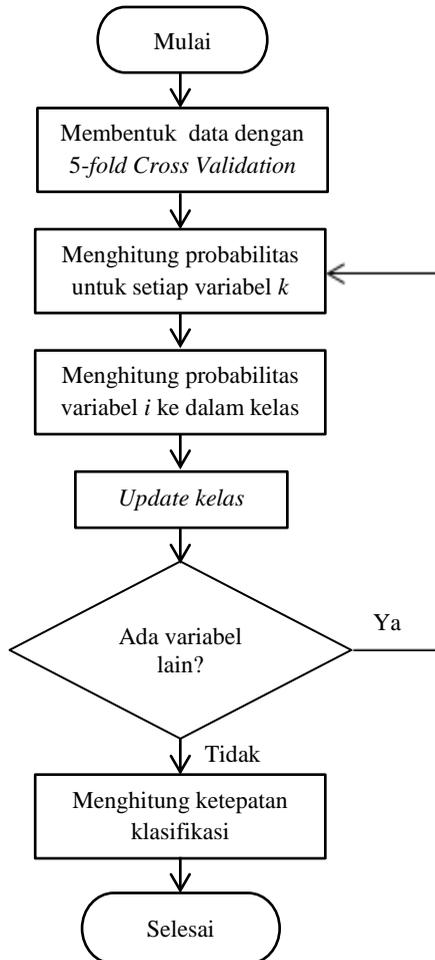
Gambar 3.1 Diagram Alir Penelitian



Gambar 3.2 Diagram Alir Klasifikasi menggunakan Regresi Logistik Ordinal



Gambar 3. 3 Diagram Alir Klasifikasi menggunakan Algoritma Genetika



Gambar 3.4 Diagram Alir Klasifikasi menggunakan *Naive Bayes*

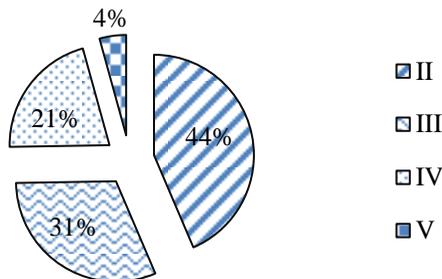
(Halaman ini sengaja dikosongkan)

BAB IV ANALISIS DAN PEMBAHASAN

Pada bab ini dibahas mengenai karakteristik pasien pneumonia di RSUD Dr. Soetomo dan hasil klasifikasi menggunakan metode Regresi Logistik Ordinal, *Hybrid* Regresi Logistik Ordinal – Algoritma Genetika dan *Naïve Bayes Classification*.

4.1 Karakteristik Data Pasien Pneumonia di RSUD Dr. Soetomo

Data rekam medis pasien pneumonia terdiri atas 12 variabel independen dan 4 kategori pada variabel dependen. Berikut ini merupakan proporsi dari jumlah pasien pneumonia pada masing-masing kelas.



Gambar 4.1 Proporsi Jumlah Pasien pada Tiap Kelas Risiko Pneumonia

Gambar 4.1 menunjukkan bahwa terdapat empat kategori kelas risiko pasien pneumonia yaitu kelas II, III, IV dan V dan memiliki proporsi yang berbeda (*unbalanced*). Pasien pneumonia dengan jumlah terbanyak adalah pasien pada kelas risiko II yaitu sebesar 44%. Kemudian jumlah pasien terbanyak kedua adalah pada kelas risiko III yaitu sebesar 31%. Proporsi pasien pneumonia pada kelas IV sebesar 21% dan yang paling sedikit adalah pasien pneumonia pada kelas risiko V yaitu hanya sebanyak 4%.

Perhitungan rata-rata, deviasi standar, nilai minimum, dan nilai maksimum digunakan untuk mengetahui karakteristik data rekam medis pasien pneumonia seperti disajikan pada Tabel 4.1.

Tabel 4.1 Karakteristik Variabel Independen

Variabel	Mean	StDev	Min	Max
Usia (tahun)	55.28	14.26	16.00	89.00
Sistolik (mmHg)	119.88	20.82	80.00	190.00
Diastolik (mmHg)	74.50	12.05	40.00	119.00
Nadi (kali/menit)	105.17	16.45	70.00	160.00
Napas (kali/menit)	28.02	6.51	16.00	52.00
Suhu (°C)	37.08	0.93	32.10	39.50
pH	7.42	0.10	7.02	7.77
Natrium (mEq/L)	136.68	7.02	115.90	166.00
Glukosa (mg/dL)	156.78	84.20	32.00	503.00
Hematokrit (%)	38.03	6.49	21,80	54.90
PO ₂ (mmHg)	107.84	55.57	33.40	349.00
BUN (mg/dL)	19.59	14.56	3.00	88.00

Tabel 4.1 menunjukkan bahwa rata-rata pasien yang pernah dirawat di RSUD Dr. Soetomo karena terserang pneumonia berusia 55.28 tahun dengan pasien termuda berusia 16 tahun dan yang tertua adalah 89 tahun. Rata-rata tekanan darah sistolik pasien adalah sebesar 119.88 mmHg dengan tekanan terendah adalah 80 mmHg dan yang tertinggi sebesar 190 mmHg. Tekanan sistolik tersebut dikatakan sangat tinggi akan tetapi tidak dianggap berbahaya bagi penderita pneumonia Sedangkan untuk tekanan darah diastolik pasien pneumonia berkisar antara 40 hingga 119 mmHg dengan rata-rata 74.50 mmHg. Rata-rata tekanan darah tersebut masih tergolong normal karena sudah lebih dari 60 mmHg.

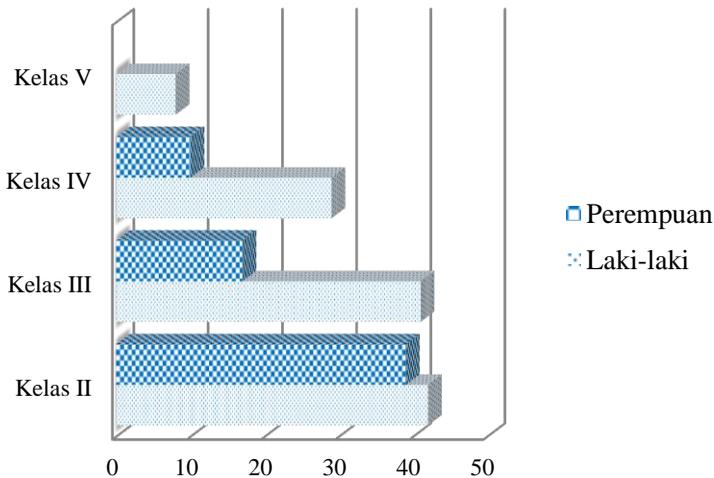
Rata-rata denyut nadi pasien pneumonia sebesar 105.17 denyut/menit, sedikit berada di atas batas normal yaitu 100 denyut/menit. Denyut nadi pasien pneumonia terendah adalah 60 dan tertinggi adalah 160 denyut/menit sehingga angka tersebut

dikatakan tinggi dan berbahaya bagi pasien pneumonia. Rata-rata frekuensi napas pasien pneumonia dikatakan tinggi yaitu sebesar 28.02 kali tiap menit. Frekuensi napas terendah pasien pneumonia adalah 16 kali dan yang tertinggi adalah 90 kali tiap menit, semakin tinggi frekuensi napas merupakan salah satu indikasi bahwa pneumonia pasien tersebut sudah cukup parah. Rata-rata suhu tubuh pasien pneumonia yaitu 37.07 °C sehingga masih dikatakan normal karena suhu tubuh pasien berada di sekitar nilai 37°C. Suhu terendah berada di bawah batas normal yaitu hanya sebesar 32 °C, begitu juga dengan suhu tubuh tertinggi pasien yaitu sebesar 39.5 °C yang berada di atas suhu tubuh normal.

Analisis gas darah arteri pasien (dalam pH) memiliki rata-rata yang masih normal, yaitu sebesar 7.42, dengan analisis gas darah (dalam pH) terendah sebesar 7.02, dan yang tertinggi tertinggi sebesar 7.77. Kadar natrium darah pasien pneumonia memiliki rata-rata yang masih dikatakan normal yaitu sebesar 136.68 mEq/liter. Kadar natrium pasien tertinggi sebesar 166 mEq/liter sehingga berada di atas batas atas normal yaitu 145 mEq/liter. Kadar natrium terendah sebesar 115.9, kadar natrium tersebut dikatakan sangat rendah karena berada di bawah batas normal 135 mEq/liter. Rata-rata kadar gula dalam darah pasien pneumonia sebesar 156,78 mg/dL, sedikit lebih tinggi dari batas normal yaitu 150 mg/dL. Kadar glukosa darah terendah yaitu sebesar 32 mg/dL, angka ini cukup kecil akan tetapi pada kasus penyakit pneumonia tidak terlalu menjadi perhatian. Sedangkan kadar gula darah pasien yang tertinggi yaitu sebesar 503 mg/liter, angka tersebut dikatakan sangat tinggi karena lebih dari tiga kali lipat batas normal kadar gula pada darah orang normal. Hal ini merupakan salah satu hal yang menjadi indikasi tingkat keparahan penyakit pneumonia.

Persentase sel darah merah dibandingkan dengan volume darah atau biasa disebut hematokrit pada pasien pneumonia memiliki rata-rata sebesar 38.03%. Persentase hematokrit tertinggi sebesar 54.90% dan persentase hematokrit terendah adalah 21.8%, persentase tersebut dinilai cukup rendah dan

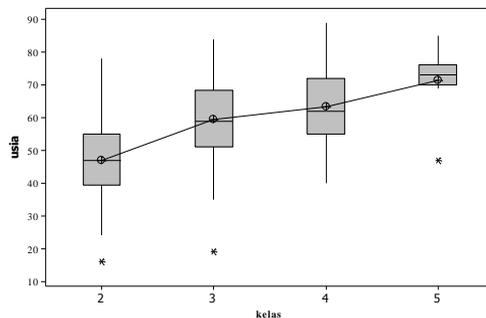
dianggap berbahaya bagi pasien. Tekanan oksigen dalam darah atau kadar PO_2 pasien pneumonia memiliki rata-rata sebesar 107.8 mmHg, sedikit lebih tinggi dari batas normal sehingga kadar PO_2 tersebut masih dikatakan normal. Kadar PO_2 pasien terendah adalah 33.40 mmHg, kadar tersebut berada di bawah normal dan menjadi salah satu pertimbangan penentuan keparahan pneumonia. Sedangkan kadar PO_2 tertinggi sebesar 349 mmHg, walaupun angka tersebut cukup tinggi tapi tidak menjadi pertimbangan dalam penentuan keparahan pneumonia. Rata-rata BUN (*Blood Urea Nitrogen*) pada pasien pneumonia adalah sebesar 19.59 mg/dL, rata-rata tersebut masih dikatakan normal karena tidak lebih dari 25 mg/dL. Nilai BUN pasien terendah yaitu 3 mg/dL dan BUN pasien tertinggi yaitu sebesar 114 mm/dL, angka tersebut sangat jauh di atas normal karena bernilai lebih dari empat kali lipat dari batas yang seharusnya dan dianggap cukup berbahaya bagi pasien pneumonia.



Gambar 4.2 Proporsi Jumlah Pasien Pneumonia pada Tiap Kelas Risiko berdasarkan Jenis Kelamin

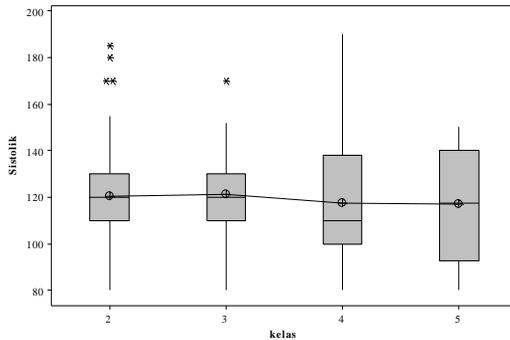
Proporsi jumlah pasien pneumonia laki-laki dan perempuan pada setiap kelas ditunjukkan dengan diagram batang yang terdapat pada Gambar 4.2. Diagram batang tersebut menunjukkan bahwa jumlah pasien laki-laki lebih banyak dibandingkan jumlah pasien pneumonia perempuan. Hal ini ditunjukkan dengan ukuran diagram yang lebih panjang pada bagian atasnya. Pada kelas risiko II proporsi laki-laki dengan perempuan hampir sama. Pada kelas risiko III jumlah pasien pneumonia laki-laki hampir 3 kali lipat dari jumlah pasien pneumonia perempuan begitu pula dengan proporsi pasien pada kelas risiko IV. Pada kelas risiko V semua pasien pneumonia berjenis kelamin laki-laki. Berdasarkan uraian tersebut dapat disimpulkan bahwa pneumonia lebih banyak menyerang pada pasien yang berjenis kelamin laki-laki.

Rata-rata variabel independen antar kelas risiko pneumonia dapat digunakan untuk mengetahui karakteristik tertentu pada masing-masing kelas risiko. Perbedaan rata-rata masing-masing variabel independen antar kelas risiko dapat dilihat menggunakan *boxplot*. *Boxplot* tersebut dapat menunjukkan adanya perbedaan dari setiap kelas risiko pneumonia. Berikut ini adalah *boxplot* serta uraian untuk masing-masing variabel independen yang bersifat kontinu antar kelas risiko. Gambar 4.3 menunjukkan *boxplot* perbandingan rata-rata usia pasien antar kelas risiko pneumonia.



Gambar 4.3 Perbedaan Rata-rata Variabel Usia antar Kelas

Rata-rata usia pasien pneumonia berbeda pada tiap kelas risiko dengan ditunjukkan oleh garis *mean* pada *boxplot* yang tidak sejajar. Pada kelas risiko II,III dan V menunjukkan distribusi yang simetris, sedangkan pada kelas risiko IV lebih banyak pasien yang berusia tua dibanding usia muda. Hal ini ditunjukkan dengan *boxplot* yang berbentuk *skewness* kanan seperti disajikan pada Gambar 4.3. Rata-rata usia pasien pneumonia pada kelas III, IV dan V adalah pasien yang berusia di atas 50 tahun. Salah satu alasan pneumonia banyak menyerang para lansia adalah karena sistem kekebalan tubuh manusia akan menurun setelah usia 50 tahun.

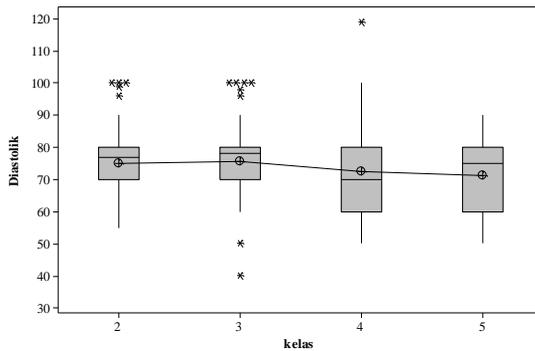


Gambar 4.4 Perbedaan Rata-rata Tekanan Darah Sistolik antar Kelas

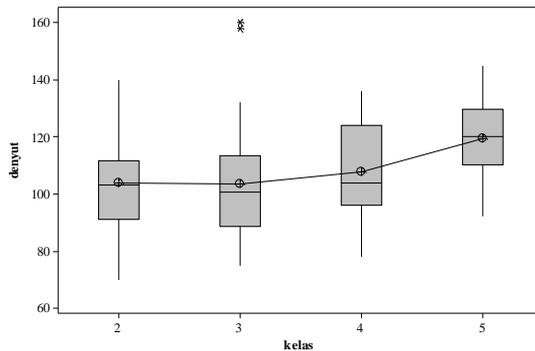
Gambar 4.4 menunjukkan bahwa rata-rata tekanan darah sistolik pasien pneumonia pada masing-masing kelas risiko tidak berbeda secara signifikan. Hal tersebut dapat dilihat dari garis *mean* pada *boxplot* yang hampir sejajar. Nilai rata-rata dan median tekanan darah sistolik pasien antar kelas risiko hampir sama kecuali pada kelas risiko IV yang sedikit lebih rendah dibandingkan kelas yang lainnya.

Rata-rata tekanan darah diastolik pasien pneumonia pada tiap kelas risiko tidak berbeda signifikan seperti disajikan pada Gambar 4.5. Hal ini ditunjukkan dengan garis *mean* pada *boxplot* yang tidak sejajar. Nilai mean dan median tekanan darah diastolik

pada setiap kategori kelas risiko tidak sama namun tidak jauh berbeda. Pada kelas IV distribusi tekanan darah diastolik menunjukkan distribusi yang simetris, sedangkan pada kelas II, III, dan V distribusi tekanan darah diastolik tidak simetris. Hal ini dikarenakan pasien yang memiliki tekanan darah yang rendah lebih banyak dari pasien yang memiliki tekanan darah diastolik yang tinggi.



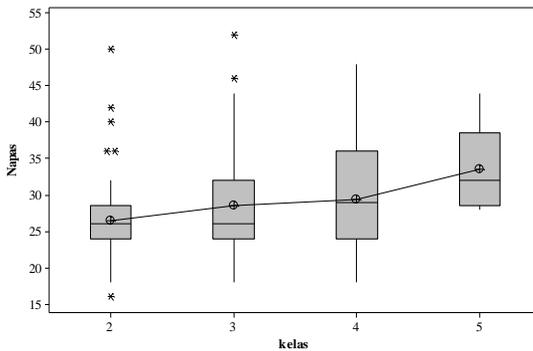
Gambar 4.5 Perbedaan Rata-rata Tekanan Darah Diastolik antar Kelas



Gambar 4.6 Perbedaan Rata-rata Denyut Nadi antar Kelas

Gambar 4.6 menunjukkan *boxplot* denyut nadi pasien pneumonia pada setiap kelas risiko. Berdasarkan *boxplot* tersebut

terlihat bahwa semakin parah pneumonia seseorang maka rata-rata denyut nadinya juga semakin tinggi. Pada kelas risiko II dan V nilai rata-rata dan median sama sedangkan pada kelas risiko III dan IV rata-rata denyut nadi pasien lebih tinggi dibandingkan mediannya. Hal ini disebabkan adanya denyut nadi pasien yang berbeda cukup jauh dibandingkan pasien yang lainnya pada kelas risiko III dan IV. Gambar 4.7 menunjukkan perbandingan rata-rata frekuensi napas dari pasien pneumonia antar kelas risiko.

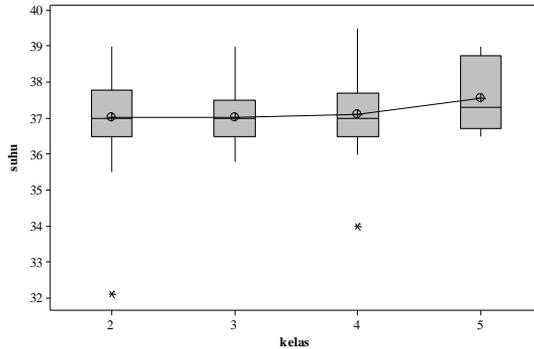


Gambar 4.7 Perbedaan Rata-rata Frekuensi Napas Pasien antar Kelas

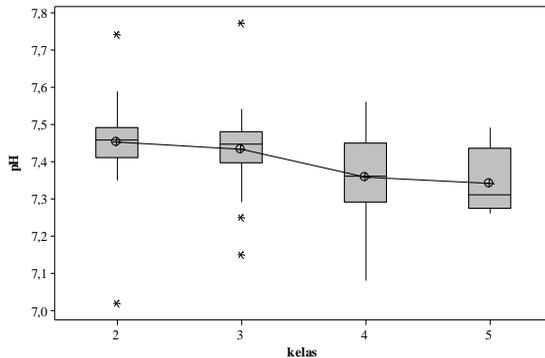
Boxplot pada Gambar 4.7 menunjukkan bahwa semakin parah pneumonia yang diderita pasien maka frekuensi napasnya semakin cepat. Hal ini dapat dilihat dari garis *mean* pada *boxplot* yang tidak sejajar dan cenderung naik pada setiap kelas risiko. Nilai median dan rata-rata frekuensi napas pada kelas risiko III dan V berbeda karena terdapat pasien dengan frekuensi napas yang berbeda jauh dari pasien lainnya. Sedangkan pada kelas II dan IV nilai median dan rata-rata frekuensi napas pasien hampir sama.

Gambar 4.8 menunjukkan *boxplot* suhu tubuh pasien pneumonia pada tiap kelas risiko. Berdasarkan *boxplot* tersebut terlihat bahwa semakin parah pneumonia dari pasien, rata-rata suhu tubuhnya semakin tinggi. Rata-rata suhu tubuh pasien yang

berada pada kelas risiko II dan kelas risiko III hampir sama, kemudian mulai naik pada kelas risiko IV hingga kelas risiko V.



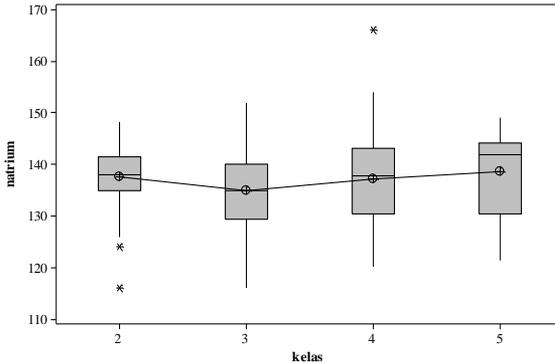
Gambar 4.8 Perbedaan Rata-rata Suhu Tubuh antar Kelas



Gambar 4.9 Perbedaan Rata-rata Analisis Gas Darah Arteri (dalam pH) antar Kelas

Gambar 4.9 menunjukkan bahwa semakin parah pneumonia yang diderita pasien, nilai analisis gas darah arteri (dalam pH) pasien pneumonia semakin rendah. Hal ini ditunjukkan dengan nilai rata-rata pada *boxplot* yang berbeda pada masing-masing kelas risiko dan semakin menurun. Pada kelas risiko II dan III

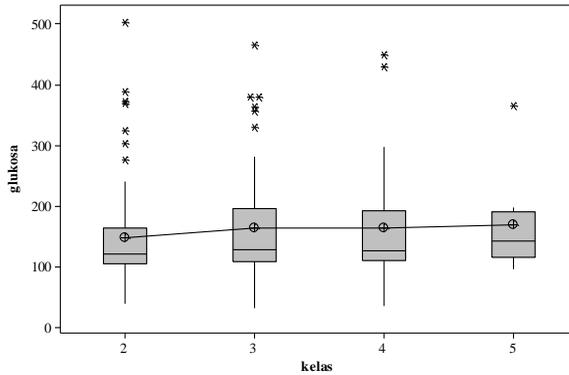
rata-rata analisis gas darah arteri (dalam pH) hampir sama kemudian mulai menurun pada kelas IV hingga kelas risiko V.



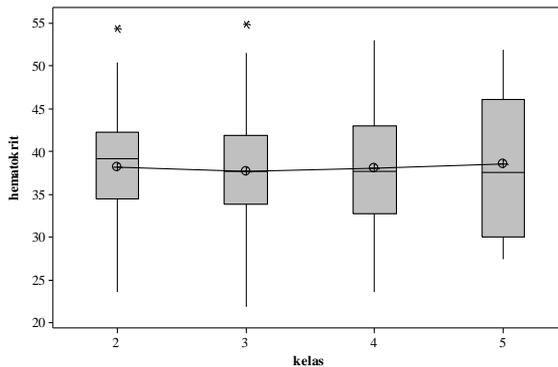
Gambar 4.10 Perbedaan Rata-rata Kadar Natrium dalam Darah antar Kelas

Gambar 4.10 merupakan *boxplot* kadar natrium dalam darah pasien pneumonia tiap kelas risiko. Rata-rata kadar natrium dalam darah tiap kelas risiko berbeda walaupun tidak secara signifikan. Hal ini ditunjukkan dengan garis *mean* yang menghubungkan *boxplot* setiap kelas yang tidak sejajar. Nilai median dan rata-rata kadar natrium darah hampir sama pada setiap kelas risiko kecuali pada kelas risiko V. Rata-rata kadar natrium dalam darah pasien pneumonia pada kelas risiko V lebih rendah dibandingkan nilai mediannya. Hal ini disebabkan karena terdapat pasien dengan kadar natrium dalam darah yang berbeda dengan pasien yang lain.

Rata-rata kadar glukosa dalam darah pasien pneumonia pada setiap kelas risiko tidak berbeda secara signifikan seperti disajikan pada Gambar 4.11. Rata-rata kadar glukosa dalam darah pada pasien yang berada di kelas risiko III sama dengan pasien yang berada pada kelas risiko IV. Nilai median dan rata-rata kadar glukosa dalam darah pasien pada masing-masing kelas risiko berbeda. Perbedaan ini disebabkan adanya pasien yang memiliki kadar glukosa dalam darah yang berbeda cukup jauh dengan pasien lainnya atau bisa dikatakan ekstrem.



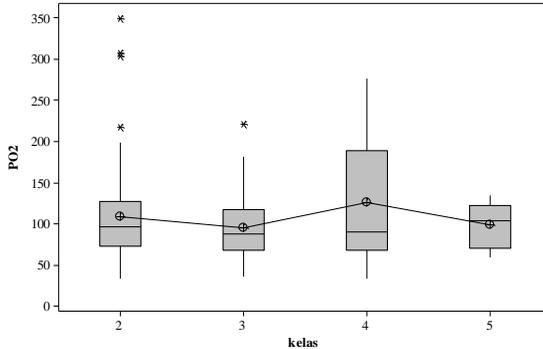
Gambar 4.11 Perbedaan Rata-rata Kadar Glukosa dalam Darah antar Kelas



Gambar 4.12 Perbedaan Rata-rata Hematokrit antar Kelas

Gambar 4.12 menunjukkan *boxplot* persentase hematokrit yaitu perbandingan sel darah merah terhadap volume darah pada setiap kelas risiko pneumonia. Berdasarkan *boxplot* tersebut terlihat bahwa kadar hematokrit pasien antar kelas risiko memiliki nilai yang hampir sama. Hal ini ditandai dengan garis *mean* yang menghubungkan *boxplot* pada masing-masing kelas risiko hampir sejajar. Pada kelas risiko III dan IV nilai rata-rata dan median hematokrit pasien hampir sama. Rata-rata presentase hematokrit

pasien pada kelas risiko II lebih rendah dibandingkan nilai mediannya. Sedangkan pada kelas risiko V rata-rata presentase hematokrit pasien lebih tinggi dibandingkan dengan nilai mediannya. Hal ini menunjukkan bahwa pada kelas tersebut terdapat pasien yang memiliki kadar hematokrit berbeda dengan pasien lainnya pada.

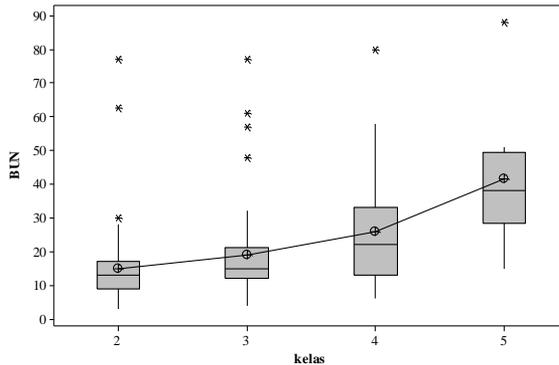


Gambar 4.13 Perbedaan Rata-rata Kadar PO_2 antar Kelas

Boxplot pada Gambar 4.13 menunjukkan bahwa garis *mean* yang menghubungkan *boxplot* antar kelas risiko tidak sejajar seperti disajikan. Hal ini menunjukkan bahwa rata-rata kadar PO_2 tiap kelas risiko berbeda. Nilai rata-rata dan median PO_2 pada setiap kelas risiko hampir sama kecuali pada kategori kelas risiko II dan kelas risiko IV. Rata-rata kadar PO_2 pada kelas II dan IV lebih tinggi dibandingkan nilai mediannya. Hal ini menunjukkan bahwa terdapat pasien pneumonia dengan kadar PO_2 yang berbeda dengan pasien pneumonia yang lainnya pada kelas tersebut.

Rata-rata BUN pada setiap kelas risiko berbeda secara signifikan yang ditandai dengan garis *mean* yang tidak sejajar dan cenderung meningkat seperti disajikan pada Gambar 4.14. Hal ini menunjukkan bahwa semakin parah pneumonia yang diderita oleh pasien maka semakin tinggi nilai BUN pasien tersebut. Nilai rata-rata BUN pasien pada setiap kelas risiko berbeda dengan nilai

mediannya. Perbedaan ini disebabkan adanya pasien dengan nilai BUN yang berbeda dengan pasien lainnya pada kelas tersebut.



Gambar 4.14 Perbedaan Rata-rata BUN antar Kelas

Sebagian besar variabel independen memiliki perbedaan rata-rata yang cukup signifikan secara visual sehingga diperlukan pengujian beda rata-rata antar kelas risiko untuk memperkuat hal tersebut. Adapun pengujian rata-rata untuk lebih dari dua kelas dapat menggunakan ANOVA dengan hasil sebagai berikut.

Tabel 4.2 Hasil Pengujian Beda Rata-rata

Variabel	F_{hitung}	p -value
Usia (X_2)	26.05	0.000*
Sistolik (X_3)	0.32	0.809
Diastolik (X_4)	0.69	0.559
Denyut (X_5)	2.88	0.037*
Napas (X_6)	4.51	0.004*
Suhu (X_7)	0.86	0.461
pH (X_8)	11.57	0.000*
Natrium (X_9)	1.88	0.135
Glukosa (X_{10})	0.56	0.644
Hematokrit (X_{11})	0.09	0.967
PO2 (X_{12})	2.53	0.059
BUN (X_{13})	13.57	0.000*

Ket: * = signifikan (P -value < α (0.05))

Hasil pengujian beda rata-rata antar kelas risiko pada masing-masing variabel disajikan pada Tabel 4.2. Keputusan yang diambil adalah Tolak H_0 apabila nilai $p\text{-value} < \alpha$ (0.05). Hasil tersebut menunjukkan bahwa terdapat perbedaan usia, frekuensi denyut per menit, frekuensi napas per menit, nilai analisis gas darah (dalam pH) dan BUN yang signifikan antar kelas risiko pasien pneumonia. Sedangkan untuk variabel yang lain yaitu tekanan darah sistolik, tekanan darah diastolik, suhu tubuh, kadar natrium dan glukosa dalam darah, hematokrit dan kadar oksigen dalam darah (PO_2) jika dilihat dari *boxplot* memang terdapat perbedaan namun hasil pengujian menunjukkan bahwa perbedaan tersebut tidak signifikan.

4.2 Korelasi Antar Variabel Independen

Hubungan antar variabel dapat diketahui berdasarkan nilai korelasi. Seluruh variabel independen selain jenis kelamin yang digunakan dalam mengidentifikasi kelas risiko pasien pneumonia bersifat kontinu sehingga korelasi yang digunakan adalah korelasi *Pearson*. Korelasi digunakan dalam menentukan hubungan antar variabel independen. Nilai korelasi *Pearson* disajikan pada Tabel 4.2. Beberapa hasil korelasi antar variabel independen pada Tabel 4.2 menunjukkan nilai $p\text{-value} < \alpha$ (0.05) yang menunjukkan bahwa terdapat hubungan yang signifikan di antara variabel independen. Korelasi yang tinggi mengindikasikan terjadinya kasus multikolinieritas di antara variabel independen. Proses selanjutnya adalah melakukan seleksi variabel independen sebagai upaya mengatasi adanya kasus multikolinieritas. Seleksi variabel independen dapat dilakukan menggunakan metode *backward*, *forward*, *stepwise* dan algoritma genetika.

Tabel 4.3 Korelasi antar Variabel Independen

	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}
X_3	0.171 0.019*										
X_4	0.010 0.892	0.686 0.000*									
X_5	-0.131 0.075	-0.134 0.067	0.016 0.828								
X_6	-0.133 0.071	-0.119 0.105	-0.004 0.956	0.446 0.000*							
X_7	0.109 0.138	-0.004 0.956	-0.064 0.383	0.188 0.010*	0.105 0.154						
X_8	-0.194 0.008*	0.060 0.414	0.072 0.326	0.010 0.896	0.118 0.109	0.084 0.255					
X_9	0.116 0.116	0.120 0.102	0.071 0.333	0.050 0.497	0.020 0.785	0.004 0.956	-0.262 0.000*				
X_{10}	-0.038 0.607	0.067 0.366	0.084 0.252	0.031 0.679	-0.068 0.353	0.030 0.684	0.020 0.782	-0.096 0.192			
X_{11}	0.012 0.866	0.070 0.344	0.126 0.086	-0.105 0.153	-0.047 0.525	-0.133 0.070	-0.391 0.000*	0.200 0.006*	-0.031 0.676		
X_{12}	-0.064 0.385	-0.13 0.860	-0.080 0.280	0.065 0.381	-0.005 0.950	0.000 0.996	-0.237 0.001*	0.111 0.131	-0.042 0.565	0.011 0.885	
X_{13}	0.249 0.001*	-0.007 0.920	-0.012 0.875	0.013 0.865	0.057 0.443	0.021 0.776	-0.150 0.040*	0.248 0.001*	0.016 0.823	0.017 0.815	0.072 0.328

Ket: * = signifikan (P -value < α (0.05), X_2 : usia, X_3 : sistolik, X_4 : diastolik, X_5 : denyut, X_6 : napas, X_7 : suhu, X_8 : pH, X_9 : natrium, X_{10} : glukosa, X_{11} : hematokrit, X_{12} : PO₂, X_{13} : BUN

4.3 Regresi Logistik Ordinal dengan *Backward Elimination*

Seleksi variabel menggunakan metode *backward* dilakukan dengan cara memasukkan semua variabel prediktor ke dalam model terlebih dahulu. Seleksi variabel dilakukan berdasarkan nilai AIC, model terbaik yang dipilih yaitu model dengan AIC terkecil. Tahapan seleksi variabel menggunakan metode *backward* ditampilkan pada Lampiran 4. Variabel yang terpilih menggunakan metode *backward* adalah jenis kelamin (X_1), usia (X_2), denyut (X_5), napas (X_6), pH (X_8), natrium (X_9), glukosa (X_{10}), hematokrit (X_{11}) dan BUN (X_{13}) sehingga sembilan variabel tersebut masuk dalam model. Pada *boxplot* yang disajikan pada Lampiran 2 terlihat bahwa terdapat perbedaan rata-rata antar kelas pada masing-masing variabel yang terpilih menggunakan seleksi *backward*. Jika dilihat tanda pada masing-masing parameter tanda tersebut sudah sesuai dengan *boxplot* dan telah sesuai dengan kenyataan pada dunia medis.

Tabel 4.4 Uji Parsial Seleksi *Backward*

Variabel	<i>B</i>	<i>SE Error</i>	<i>Wald</i>	<i>p-value</i>
<i>Intercept</i> 1 (α_1)	-0.420	1.764	-0.238	0.810
<i>Intercept</i> 2 (α_2)	2.596	1.784	1.455	0.142
<i>Intercept</i> 3 (α_3)	6.475	1.893	3.420	0.001
JK: Laki-laki ($X_1 = 2$)	1.340	0.437	3.069	0.002
Usia (X_2)	9.524	1.489	6.397	0.000
Denyut (X_5)	2.889	1.244	2.322	0.021
Napas (X_6)	5.892	1.292	4.562	0.000
pH (X_8)	-12.446	2.132	-5.838	0.000
Natrium (X_9)	-6.602	1.578	-4.183	0.000
Glukosa (X_{10})	2.389	1.055	2.265	0.025
Hematokrit (X_{11})	-2.384	1.161	-2.054	0.042
BUN (X_{13})	5.284	1.243	4.250	0.000

Tahap selanjutnya adalah melakukan uji serentak untuk mengetahui adanya variabel independen yang berpengaruh signifikan terhadap variabel respon secara serentak. Hasil uji serentak menghasilkan nilai *p-value* sebesar 0.000. Nilai *p-value* ($0.000 < \alpha (0.05)$) sehingga dapat disimpulkan bahwa secara

serentak minimal ada satu variabel independen yang berpengaruh signifikan dalam model. Proses selanjutnya adalah melakukan uji parsial untuk mengetahui variabel yang berpengaruh signifikan secara parsial dalam model yang terbentuk. Tabel 4.4 menunjukkan bahwa seluruh variabel independen berpengaruh signifikan terhadap penentuan kelas risiko pasien pneumonia yang ditandai dengan nilai $p\text{-value} < \alpha$ (0.05). Fungsi logit yang terbentuk berdasarkan nilai estimasi parameter pada uji parsial disajikan pada Tabel 4.5.

Tabel 4.5 Fungsi Logit Seleksi *Backward*

Fungsi Logit
$\hat{g}_1(x) = -0.420 + 1.340 X_{1(2)} - 9.524 X_2 + 2.889X_5 + 5.892X_6 - 12.446X_8 - 6.602X_9 + 2.389X_{10} - 2.384X_{11} + 5.284X_{13}$
$\hat{g}_2(x) = 2.596 + 1.340 X_{1(2)} - 9.524 X_2 + 2.889X_5 + 5.892X_6 - 12.446X_8 - 6.602X_9 + 2.389X_{10} - 2.384X_{11} + 5.284X_{13}$
$\hat{g}_3(x) = 6.475 + 1.340 X_{1(2)} - 9.524 X_2 + 2.889X_5 + 5.892X_6 - 12.446X_8 - 6.602X_9 + 2.389X_{10} - 2.384X_{11} + 5.284X_{13}$

Fungsi logit yang didapatkan kemudian digunakan untuk membuat fungsi peluang setiap kategori kelas risiko yang digunakan untuk mengklasifikasikan pasien pada data *testing*. Berdasarkan Tabel 4.4 terdapat empat kemungkinan nilai dari Y yaitu:

$$Y_i = 2 \text{ jika } Y * _i \leq -0.420$$

$$Y_i = 3 \text{ jika } -0.420 \leq Y * _i \leq 2.596$$

$$Y_i = 4 \text{ jika } 2.596 \leq Y * _i \leq 6.475$$

$$Y_i = 5 \text{ jika } Y * _i \geq 6.475$$

Rumus $Y * _i$ untuk mengelompokkan pasien ke dalam kelas risiko (Y) adalah sebagai berikut:

$$Y * _i = 1.340 X_{1(2)} - 9.524 X_2 + 2.889X_5 + 5.892X_6 - 12.446X_8 - 6.602X_9 + 2.389X_{10} - 2.384X_{11} + 5.284X_{13}$$

Adapun model peluang pada masing-masing kelas risiko yang didapatkan adalah sebagai berikut.

$$P(Y = 2) = \frac{1}{1 - e^{(Y * _i + 0.420)}}$$

$$P(Y = 3) = \frac{1}{1 - e^{(Y^*_i - 2.596)}} - P(Y = 2)$$

$$P(Y = 4) = \frac{1}{1 - e^{(Y^*_i - 6.475)}} - P(Y = 3)$$

$$P(Y = 5) = 1 - \frac{1}{1 - e^{(Y^*_i - 6.475)}}$$

Hasil uji kesesuaian model menghasilkan nilai *p-value* sebesar 1.000 sehingga dapat disimpulkan bahwa model sesuai artinya tidak ada perbedaan yang nyata antara hasil observasi dengan kemungkinan hasil prediksi model. Interpretasi model dilakukan dengan melakukan interpretasi nilai *odds ratio*. Nilai *odds ratio* disajikan pada Tabel 4.6.

Tabel 4.6 *Odds Ratio* Seleksi *Backward*

Variabel	B	Odds Ratio
Laki-laki ($X_1 = 2$)	1.340	3.819
Usia (X_2)	9.524	13684.239
Denyut (X_5)	2.889	17.975
Napas (X_6)	5.892	362.129
pH (X_8)	-12.446	0.000
Natrium (X_9)	-6.602	0.001
Glukosa (X_{10})	2.389	10.903
Hematokrit (X_{11})	-2.384	0.092
BUN (X_{13})	5.284	197.157

Tabel 4.6 menunjukkan bahwa pasien laki-laki berisiko menderita pneumonia sebesar 3.819 kali lebih tinggi dibandingkan pasien perempuan. Setiap kenaikan usia pasien sebesar 1 tahun maka pasien berisiko masuk ke dalam kelas risiko V 13684.239 kali lebih tinggi dibandingkan masuk ke dalam kelas risiko selain V. Setiap kenaikan satu frekuensi denyut nadi pasien per menit maka pasien tersebut berisiko masuk ke dalam kelas risiko V sebesar 17.975 kali lebih tinggi dibandingkan masuk ke dalam kelas risiko selain V. Setiap kenaikan satu frekuensi napas pasien per menit maka pasien tersebut berisiko masuk ke dalam kelas V 362.129 kali lebih tinggi dibandingkan masuk ke dalam kelas risiko selain V. Setiap kenaikan kadar natrium dalam darah sebesar 1 mEq/liter maka pasien tersebut berisiko masuk ke kelas selain V 1000 kali lebih tinggi

dibandingkan masuk ke kelas V. Setiap kenaikan 1 mg/dL kadar glukosa dalam darah akan meningkatkan risiko pasien tersebut masuk ke dalam kelas V 10.903 kali dibandingkan masuk ke dalam kelas yang lainnya. Kenaikan kadar hematokrit dalam darah sebesar 1% akan meningkatkan risiko pasien masuk ke dalam kelas selain kelas V sebesar $1/0.092=10.870$ kali dibandingkan masuk ke dalam kelas V. Setiap kenaikan kadar urea nitrogen dalam darah (BUN) sebesar 1 mg/dL maka pasien tersebut berisiko masuk ke dalam kelas V 197.157 kali lebih tinggi dibandingkan masuk ke dalam kelas risiko selain V.

Hasil klasifikasi pada data *testing* diperoleh dengan menghitung dan membandingkan peluang dari masing-masing kelas risiko. Pasien akan diklasifikasikan pada kelas yang memiliki peluang terbesar. Hasil klasifikasi pasien pneumonia disajikan pada Tabel 4.7.

Tabel 4.7 *Confusion Matriks* Klasifikasi Regresi Logistik Ordinal Hasil Seleksi *Backward*

Aktual	Prediksi			
	II	III	IV	V
II	11	4	1	0
III	3	8	1	0
IV	0	3	5	0
V	0	0	0	1

Nilai *recall* (R) untuk setiap kelas didapatkan dari perhitungan nilai yang tepat diklasifikasikan dibagi dengan total observasi pada setiap kelas. Masing-masing nilai R adalah sebagai berikut.

$$R_1 = \frac{11}{16} = 0.688; R_2 = 0.667; R_3 = 0.625; R_4 = 1$$

Ukuran ketepatan klasifikasi yang digunakan adalah *Area Under ROC Curve* (AUC) dan *Geometric mean* yang menghasilkan ketepatan klasifikasi sebesar 0.7448 dan 0.7316 dengan perhitungan sebagai berikut.

$$AUC = \frac{1}{4}(0.69 + 0.67 + 0.63 + 1) = 0.7448$$

$$G_mean = (0.69 \times 0.67 \times 0.63 \times 1)^{\frac{1}{4}} = 0.7316$$

4.4 Regresi Logistik Ordinal dengan *Forward Selection*

Seleksi variabel menggunakan metode *forward* dimulai dengan memodelkan model dengan *intercept* saja kemudian dilanjutkan dengan memasukkan satu persatu variabel independen ke dalam model. Hasil seleksi variabel menggunakan metode *forward* ditampilkan pada Lampiran 2. Terdapat sembilan variabel independen yang masuk ke dalam model dengan seleksi menggunakan metode *forward*. Variabel yang terpilih adalah jenis kelamin (X_1), usia (X_2), denyut (X_5), napas (X_6), pH (X_8), natrium (X_9), glukosa (X_{10}), hematokrit (X_{11}) dan BUN (X_{13}). Variabel yang terpilih pada seleksi ini sama dengan metode *backward* sehingga model yang terbentuk dan langkah selanjutnya sama dengan metode seleksi sebelumnya.

4.5 Regresi Logistik Ordinal dengan *Stepwise Method*

Metode seleksi variabel yang ketiga adalah menggunakan *stepwise* yang merupakan kombinasi dari dua metode sebelumnya yaitu *backward* dan *forward*. Tahapan seleksi variabel menggunakan metode *stepwise* disajikan pada Lampiran 3. Hasil seleksi variabel menggunakan metode *stepwise* menghasilkan variabel yang sama dengan dua metode sebelumnya. Variabel tersebut adalah jenis kelamin (X_1), usia (X_2), denyut (X_5), napas (X_6), pH (X_8), natrium (X_9), glukosa (X_{10}), hematokrit (X_{11}) dan BUN (X_{13}). Model yang terbentuk serta tahapan selanjutnya sama dengan seleksi variabel menggunakan metode *backward* dan *forward*.

4.6 Regresi Logistik Ordinal dengan Seleksi Variabel Algoritma Genetika

Seleksi variabel menggunakan metode algoritma genetika dilakukan dengan memilih kombinasi variabel yang menghasilkan nilai ketepatan klasifikasi tertinggi. Pada proses seleksi vari-

abel menggunakan algoritma genetika akan melalui beberapa tahapan yaitu proses seleksi orang tua, pindah silang, mutasi dan elitisme. Variabel diberi label 1 untuk yang masuk ke dalam model sedangkan label 0 untuk variabel yang tidak masuk model. Berikut merupakan salah satu kromosom yang digunakan dalam representasi populasi awal.

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}
1	1	0	0	1	1	0	1	1	1	1	0	1

Gambar 4.15 Representasi Kromosom Awal pada Seleksi Variabel

Gambar 4.15 merupakan salah satu nilai inisialisasi pada kromosom pertama. Kromosom yang digunakan berisi 13 gen sesuai dengan jumlah variabel independen. Setelah dilakukan inisialisasi, langkah selanjutnya adalah membentuk 100 kromosom awal yang merupakan bangkitan bilangan 0 dan 1 yang akan digunakan sebagai populasi awal yang disajikan pada Tabel 4.8.

Tabel 4.8 Ilustrasi Kromosom Dalam Populasi Awal pada Seleksi Variabel

Populasi	Kromosom				
	X_1	X_2	X_3	...	X_{13}
1	1	1	0	...	1
2	1	0	1	...	1
3	1	0	0	...	0
\vdots	\vdots	\vdots	\vdots	...	\vdots
99	1	1	0	...	1
100	1	0	0	...	1

Setelah diperoleh populasi awal proses selanjutnya adalah melakukan evaluasi pada masing-masing kromosom yang terbentuk menggunakan fungsi *fitness* yang telah ditentukan. Fungsi *fitness* yang digunakan pada seleksi variabel adalah ketepatan klasifikasi pada data *training*. Nilai *fitness* untuk masing-masing kromosom pada populasi awal ditampilkan pada Tabel 4.9. Setelah mendapatkan nilai *fitness* dari masing-masing kromosom, langkah berikutnya adalah melakukan pembentukan kromosom orang tua dengan menggunakan metode seleksi *Roulette Wheel Selection* (RWS).

Tabel 4.9 Ilustrasi Nilai *Fitness* Setiap Kromosom Populasi Awal Data pada Seleksi Variabel

Populasi	Kromosom					Nilai
	X_1	X_2	X_3	...	X_{13}	<i>Fitness</i>
1	1	1	0	...	1	0.602
2	0	0	0	...	0	0.000
3	1	1	0	...	0	0.352
⋮	⋮	⋮	⋮	...	⋮	⋮
99	0	1	0	...	0	0.479
100	0	0	1	...	1	0.427

Langkah awal pada proses RWS adalah menghitung proporsi nilai *fitness* yaitu dengan membagi nilai *fitness* pada masing-masing kromosom dengan jumlah *fitness*. Kemudian menghitung nilai kumulatif dari proporsi nilai *fitness* tersebut. Langkah selanjutnya adalah membangkitkan 100 bilangan *random* seperti disajikan pada Tabel 4.10.

Tabel 4.10 Ilustrasi Proses RWS pada Seleksi Variabel

Populasi	<i>Fitness</i>	Proporsi	Nilai <i>Fitness</i>	<i>Random Number</i>
		Nilai <i>Fitness</i>	Kumulatif	
1	0.602	0.018	0.018	0.023
2	0.000	0.000	0.018	0.061
3	0.352	0.010	0.028	0.827
⋮	⋮	⋮	⋮	⋮
99	0.479	0.014	0.987	0.156
100	0.427	0.013	1.000	0.549

Bilangan *random* pertama yaitu sebesar 0.023 seperti ditunjukkan pada Tabel 4.10, maka kromosom orangtua 1 didapatkan dari nilai *fitness* kumulatif yang lebih besar dari 0.023 yang pertama yaitu kromosom 3, maka kromosom tersebut menjadi kromosom 1 pada populasi baru. Bilangan *random* diperoleh dengan membangkitkan 100 data berdistribusi *uniform* dengan batas bawah 0 dan batas atas 1. Tahapan tersebut berhenti ketika telah diperoleh 100 kromosom orang tua berdasarkan bilangan *random* yang telah dibangkitkan.

Setelah mendapatkan 100 kromosom orang tua, dilakukan proses pindah silang atau *crossover*. Proses pindah silang meng-

hasilkan kromosom baru dari hasil perpaduan 2 kromosom orang tua. Metode pindah silang yang digunakan adalah *single point crossover*, yaitu dua kromosom dipotong sekali pada titik-titik yang sesuai dan bagian setelah pemotongan dipertukarkan. Pada penelitian ini peluang pindah silang yang digunakan yaitu 0.8.

Orangtua	X_1	X_2	...	X_5		X_6	...	X_{13}	random
2	0	1	...	1		0	...	0	0.44
3	1	1	...	0		1	...	1	0.07
Anak	Pindah silang								
2	0	1	...	1		1	...	1	
3	1	1	...	0		0	...	0	

Gambar 4.16 Ilustrasi *Single Point Crossover* pada Seleksi Variabel

Kromosom orang tua yang mengalami pindah silang adalah kromosom yang memiliki bilangan *random* kurang dari probabilitas pindah silang yaitu sebesar 0.8. Salah satu contoh kromosom yang mengalami pindah silang dapat dilihat pada Gambar 4.16. Bilangan *random* pada kromosom orangtua ke-2 dan ke-3 kurang dari 0.8 sehingga akan dilakukan pindah silang pada kedua kromosom tersebut. Langkah pertama adalah menentukan posisi pindah silang. Cara menentukan posisi pindah silang adalah dengan membangkitkan bilangan *random* antara 1 sampai 13. Pada ilustrasi tersebut bilangan *random* yang terpilih adalah 5 yang ditunjukkan dengan garis putus-putus sehingga kromosom induk akan menghasilkan anak yang telah berpindah silang antara kedua kromosom induknya pada titik ke-5. Proses pindah silang ini dilanjutkan untuk semua pasang kromosom yang memiliki bilangan *random* kurang dari peluang pindah silang.

Proses selanjutnya setelah melakukan proses pindah silang adalah melakukan mutasi gen. Metode mutasi gen yang digunakan adalah mutasi *uniform* yaitu dengan memberikan kesempatan yang sama pada setiap gen untuk dilakukan proses mutasi. Tahapan awal mutasi adalah dengan membangkitkan bilangan *random* pada setiap gen. Bilangan *random* tersebut kemudian dibandingkan dengan peluang mutasi yaitu sebesar 0.1. Apabila

nilai bilangan *random* pada suatu gen kurang dari peluang mutasi, maka proses mutasi akan dilakukan pada gen tersebut. Proses mutasi dilakukan dengan mengubah kode 1 menjadi 0 atau 0 menjadi 1 pada gen yang mengalami proses mutasi. Salah satu contoh kromosom yang mengalami proses mutasi terdapat pada Gambar 4.17.

Kromosom	X_1	X_2	...	X_{12}	X_{13}
Bil. <i>random</i>	0.97	0.89	...	0.04	0.01
1	1	1	...	0	1
Proses Mutasi					
1	1	1	...	1	0

Gambar 4.17 Ilustrasi Proses Mutasi pada Seleksi Variabel

Gambar 4.17 menjelaskan terjadinya proses mutasi pada kromosom 1. Gen yang mengalami proses mutasi adalah gen yang memiliki bilangan *random* kurang dari peluang mutasi 0.1. Gen tersebut adalah gen ke-12 dan gen ke-13. Gen ke-12 yang awalnya memiliki kode 0 berubah menjadi 1 dan gen ke-13 yang pada kromosom awal memiliki kode 1 berubah menjadi 0. Proses ini dilakukan pada semua gen yang memiliki bilangan *random* kurang dari peluang mutasi 0.1. Proses selanjutnya setelah dilakukan mutasi adalah elitisme yaitu proses pengopian individu untuk mempertahankan kromosom terbaik dalam populasi. Kromosom yang dipertahankan sebesar 20% dari total kromosom dalam populasi. Proses seleksi variabel menggunakan algoritma genetika dilanjutkan hingga mendapatkan nilai *fitness* yang konvergen.

Variabel terpilih yang diperoleh setelah dilakukan seleksi variabel menggunakan metode algoritma genetika dapat dilihat pada Tabel 4.11. Kromosom yang menghasilkan nilai *fitness* terbaik digunakan untuk menentukan variabel yang akan masuk ke dalam model. Tabel 4.13 menyajikan 100 kromosom algoritma genetika dengan hasil nilai *fitness* terbaik yaitu 0.6206 dengan variabel independen terpilih yaitu variabel jenis kelamin (X_1), usia (X_2), sistolik (X_3), diastolik (X_4), denyut (X_5), napas (X_6), pH

(X_8), natrium (X_9), glukosa (X_{10}), PO_2 (X_{12}), dan BUN (X_{13}). Jika dilihat berdasarkan *boxplot* pada Lampiran 3 terlihat bahwa terdapat perbedaan rata-rata variabel antar kelas risiko. Pada beberapa variabel perbedaan rata-ratanya tidak signifikan namun sebenarnya jika ditinjau dari dunia medis variabel tersebut menentukan pengklasifikasian pasien ke dalam kelas risiko pneumonia. Hal ini bisa terjadi karena dalam memodelkan variabel tersebut terdapat interaksi dengan variabel yang lain sehingga memungkinkan variabel tersebut dapat terseleksi masuk ke dalam model.

Tabel 4.11 Kromosom dan Nilai *Fitness* Algoritma Genetika

Populasi	Kromosom					Nilai
	X_1	X_2	X_3	...	X_{13}	<i>Fitness</i>
1	1	1	1	...	1	0.6206
2	1	1	1	...	1	0.6206
3	1	1	1	...	1	0.6206
⋮	⋮	⋮	⋮	...	⋮	⋮
99	0	1	0	...	1	0.5655
100	1	0	0	...	0	0.2351

Tahap selanjutnya adalah melakukan uji serentak untuk mengetahui adanya variabel independen yang berpengaruh signifikan terhadap variabel respon secara serentak. Hasil uji serentak menghasilkan nilai *p-value* sebesar 0.000. Nilai *p-value* ($0.000 < \alpha (0,05)$) sehingga dapat disimpulkan bahwa secara serentak minimal terdapat satu variabel independen yang berpengaruh signifikan terhadap penentuan kelas risiko pasien pneumonia dalam model. Proses selanjutnya adalah melakukan uji parsial untuk mengetahui variabel yang berpengaruh signifikan secara parsial terhadap model yang terbentuk.

Tabel 4.12 menunjukkan bahwa seluruh variabel independen berpengaruh signifikan terhadap penentuan kelas risiko pasien pneumonia kecuali variabel sistolik (X_3), diastolik (X_4), denyut (X_5), glukosa (X_{10}) dan PO_2 (X_{12}). Tanda pada masing-masing parameter juga telah sesuai dengan tanda pada dunia medis. Koefisien parameter dengan nilai positif berarti menambah

peluang suatu observasi masuk ke dalam kelas risiko V, sedangkan koefisien parameter yang bernilai negatif akan mengurangi peluang suatu observasi untuk masuk ke dalam kelas V.

Tabel 4.12 Uji Parsial Seleksi Algoritma Genetika

Variabel	<i>B</i>	<i>SE Error</i>	<i>Wald</i>	<i>p-value</i>
<i>Intercept</i> 1 (α_1)	1.372	1.583	0.867	0.364
<i>Intercept</i> 2 (α_2)	4.192	1.624	2.582	0.007
<i>Intercept</i> 3 (α_3)	8.015	1.753	4.572	0.000
JK : Laki-laki ($X_1 = 2$)	1.350	0.431	3.134	0.002
Usia (X_2)	9.202	1.409	6.529	0.000
Sistolik (X_3)	0.914	1.434	0.637	0.532
Diastolik (X_4)	-1.598	1.677	-0.953	0.355
Denyut (X_5)	2.148	1.229	1.747	0.087
Napas (X_6)	7.637	1.437	5.315	0.000
pH (X_8)	-10.176	1.877	-5.422	0.000
Natrium (X_9)	-6.458	1.611	-4.009	0.000
Glukosa (X_{10})	2.047	1.050	1.950	0.055
PO2 (X_{12})	-0.602	1.148	-0.524	0.581
BUN (X_{13})	4.658	1.274	3.655	0.000

Fungsi logit pada regresi logistik ordinal hasil seleksi algoritma genetika yang terbentuk disajikan pada Tabel 4.13.

Tabel 4.13 Fungsi Logit Seleksi Algoritma Genetika

Fungsi Logit
$\hat{g}_1(x) = 1.372 + 1.350X_{1(2)} + 9.202 X_2 + 0.914X_3 - 1.598X_4 + 2.148X_5 + 7.637 X_6 - 10.176X_8 - 6.458X_9 + 2.047X_{10} - 0.602X_{12} + 4.658X_{13}$
$\hat{g}_2(x) = 4.192 + 1.350X_{1(2)} + 9.202 X_2 + 0.914X_3 - 1.598X_4 + 2.148X_5 + 7.637 X_6 - 10.176X_8 - 6.458X_9 + 2.047X_{10} - 0.602X_{12} + 4.658X_{13}$
$\hat{g}_3(x) = 8.015 + 1.350X_{1(2)} + 9.202 X_2 + 0.914X_3 - 1.598X_4 + 2.148X_5 + 7.637 X_6 - 10.176X_8 - 6.458X_9 + 2.047X_{10} - 0.602X_{12} + 4.658X_{13}$

Fungsi logit yang didapatkan kemudian digunakan untuk membuat fungsi peluang setiap kategori kelas risiko yang digunakan untuk mengklasifikasikan pasien pada data *testing*. Berdasarkan Tabel 4.14 terdapat empat kemungkinan nilai dari Y yaitu:

$$Y_i = 2 \text{ jika } Y *_i \leq 1.372$$

$$Y_i = 3 \text{ jika } 1.372 \leq Y *_i \leq 4.192$$

$$Y_i = 4 \text{ jika } 4.192 \leq Y *_i \leq 8.015$$

$Y_i = 5$ jika $Y * _i \geq 8.015$

Rumus $Y * _i$ untuk mengelompokkan pasien ke dalam kelas risiko (Y) adalah sebagai berikut:

$$Y * _i = 1.350X_{1(2)} + 9.202 X_2 + 0.914X_3 - 1.598X_4 + 2.148X_5 + 7.637 X_6 - 10.176X_8 - 6.458X_9 + 2.047X_{10} - 0.602X_{12} + 4.658X_{13}$$

Adapun model peluang pada masing-masing kelas risiko yang didapatkan adalah sebagai berikut.

$$P(Y = 2) = \frac{1}{1 - e^{(Y*_i - 1.372)}}$$

$$P(Y = 3) = \frac{1}{1 - e^{(Y*_i - 4.192)}} - P(Y = 2)$$

$$P(Y = 4) = \frac{1}{1 - e^{(Y*_i - 8.015)}} - P(Y = 3)$$

$$P(Y = 5) = 1 - \frac{1}{1 - e^{(Y*_i - 8.015)}}$$

Hasil uji kesesuaian model menghasilkan nilai p -value sebesar 1.000 sehingga dapat disimpulkan bahwa model telah sesuai artinya tidak ada perbedaan yang nyata antara hasil observasi dengan kemungkinan hasil prediksi model.

Tabel 4.14 Odds Ratio Seleksi Algoritma Genetika

Variabel	B	Odds Ratio
JK : Laki-laki ($X_1 = 2$)	1.350	3.857
Usia (X_2)	9.202	9916.943
Sistolik (X_3)	0.914	2.494
Diastolik (X_4)	-1.598	0.202
Denyut (X_5)	2.148	8.568
Napas (X_6)	7.637	2073.514
pH (X_8)	-10.176	0.000
Natrium (X_9)	-6.458	0.002
Glukosa (X_{10})	2.047	7.745
PO2 (X_{12})	-0.602	0.548
BUN (X_{13})	4.658	105.425

Interpretasi model dilakukan dengan melakukan interpretasi nilai *odds ratio*. Nilai *odds ratio* pada masing-masing variabel independen antara kelas V dibandingkan dengan kelas selain kelas V disajikan pada Tabel 4.14. Pneumonia menyerang pasien laki-laki 3.857 kali lebih tinggi dibandingkan pasien perempuan.

Kenaikan usia pasien sebesar 1 tahun akan meningkatkan risiko pasien masuk ke dalam kelas V sebesar 9916.943 kali dibandingkan masuk ke dalam kelas risiko selain V. Setiap kenaikan tekanan darah sistolik 1 mmHg maka pasien tersebut berisiko masuk ke dalam kelas V sebesar 2.494 kali dibandingkan masuk ke dalam kelas selain V. Setiap kenaikan tekanan darah diastolik sebesar 1 mmHg maka pasien berisiko untuk masuk ke dalam kelas selain kelas V sebesar $1/0.202=4.950$ kali lebih tinggi dibandingkan masuk ke dalam kelas risiko V. Setiap kenaikan satu frekuensi denyut nadi pasien per menit maka pasien tersebut berisiko masuk ke dalam kelas V 8.568 kali lebih tinggi dibandingkan masuk ke dalam kelas yang lainnya. Setiap kenaikan satu frekuensi napas pasien per menit maka pasien tersebut memiliki risiko untuk masuk ke dalam kelas V sebesar 2073.154 kali lebih tinggi dibandingkan masuk ke dalam kelas risiko selain V. Setiap kenaikan kadar natrium dalam daerah sebesar 1 mEq/liter maka pasien tersebut berisiko masuk ke kelas selain V sebesar $1/0.002=500$ kali lebih tinggi dibandingkan masuk ke dalam kelas V. Setiap kenaikan 1 mg/dL kadar glukosa dalam darah akan meningkatkan risiko pasien tersebut masuk ke dalam kelas V sebesar 7.745 kali dibandingkan masuk ke dalam kelas risiko lainnya. Kenaikan tekanan oksigen dalam darah (PO_2) sebesar 1 mmHg akan meningkatkan risiko pasien tersebut masuk ke dalam kelas risiko selain kelas V sebesar $1/0.548=1.823$ kali dibandingkan masuk ke dalam kelas risiko V. Setiap kenaikan kadar urea nitrogen dalam darah (BUN) sebesar 1 mg/dL akan meningkatkan risiko pasien tersebut masuk ke dalam kelas V 105.425 kali lebih tinggi dibandingkan masuk ke dalam kelas risiko selain V.

Hasil klasifikasi pada data *testing* diperoleh dengan membandingkan peluang dari masing-masing kelas risiko. Pasien akan diklasifikasikan pada kelas yang memiliki peluang terbesar.

Tabel 4.15 *Confusion Matriks* Klasifikasi Regresi Logistik Ordinal Hasil Seleksi Algoritma Genetika

Aktual	Prediksi			
	II	III	IV	V
II	13	3	0	0
III	0	8	4	0
IV	0	3	3	1
V	0	0	0	2

Tabel 4.15 menunjukkan hasil klasifikasi pada data *testing*. Nilai *recall* (R) untuk setiap kelas didapatkan dari perhitungan nilai yang tepat diklasifikasikan dibagi dengan total observasi pada setiap kelas. Masing-masing nilai R adalah sebagai berikut.

$$R_1 = \frac{13}{16} = 0.8125; R_2 = 0.6667; R_3 = 0.4286; R_4 = 1$$

Ukuran ketepatan klasifikasi yang digunakan adalah *Geometric mean* dan didapatkan hasil ketepatan klasifikasi sebesar 69.41% dengan perhitungan sebagai berikut.

$$AUC = (0.81 \times 0.67 \times 0.43 \times 1)^{1/4} = 69.41\%$$

Hasil seleksi variabel menggunakan kriteria *fitness* yang lain yaitu *Area Under ROC Curve* (AUC) juga menghasilkan variabel terpilih yang sama pada saat menggunakan data tersebut. Tahapan seleksi variabel menggunakan AUC sama dengan pada saat menggunakan kriteria AUC dengan hasil ketepatan klasifikasi yang didapatkan sebesar 72.69%

4.6 Optimasi Parameter Regresi Logistik Ordinal Hasil Seleksi *Backward* Menggunakan Algoritma Genetika

Langkah pertama yang dilakukan untuk melakukan estimasi parameter menggunakan algoritma genetika adalah merepresentasikan model dari persamaan regresi logistik ordinal hasil dari seleksi *backward* ke dalam kromosom dan menentukan nilai inisialisasi. Kromosom yang digunakan berisi nilai estimasi parameter dari fungsi logit. Panjang kromosom sesuai dengan jumlah estimasi parameter fungsi logit. Pada klasifikasi dengan empat kelas terdapat tiga fungsi logit namun pada regresi logistik

ordinal nilai estimasi parameter variabel independen untuk setiap kelas bernilai sama sehingga representasi kromosom awal yang terbentuk adalah sebagai berikut.

α_1	α_2	α_3	$\beta_{1(2)}$...	β_{13}
-0.420	2.596	6.475	1.340	...	5.284

Gambar 4.18 Representasi Kromosom Awal Estimasi Parameter

Gambar 4.18 merupakan salah satu nilai inisialisasi pada kromosom pertama yang terdiri dari 12 gen. Proses algoritma genetika digunakan untuk mendapatkan nilai optimum dari estimasi parameter fungsi logit yang memaksimalkan ketepatan klasifikasi. Setelah dilakukan inisialisasi kemudian membentuk 100 kromosom awal yang dibangkitkan untuk populasi awal yang disajikan pada Tabel 4.16.

Tabel 4.16 Ilustrasi Kromosom Dalam Populasi Awal pada Estimasi Parameter

Populasi	Kromosom				
	Gen1	Gen 2	Gen 3	...	Gen 12
1	-0.420	2.596	6.475	...	5.284
2	-0.420	2.596	6.475	...	5.284
3	-0.420	2.596	6.475	...	5.284
⋮	⋮	⋮	⋮	⋮	⋮
99	-0.420	2.450	6.475	...	5.284
100	-0.420	2.450	6.475	...	5.284

Tabel 4.17 Ilustrasi Nilai *Fitness* Setiap Kromosom Populasi Awal pada Estimasi Parameter

Populasi	Kromosom					Fitness
	Gen1	Gen 2	Gen 3	...	Gen 12	
1	-0.420	2.596	6.475	...	5.284	0.745
2	-0.420	2.596	6.475	...	5.284	0.745
3	-0.420	2.596	6.475	...	5.284	0.745
⋮	⋮	⋮	⋮	⋮	⋮	⋮
99	-0.420	2.450	6.475	...	5.284	0.423
100	-0.420	2.450	6.475	...	5.284	0.425

Langkah selanjutnya adalah melakukan evaluasi terhadap masing-masing kromosom yang telah dibentuk menggunakan fungsi *fitness* yang telah ditentukan. Fungsi *fitness* yang digunakan adalah *Area Under ROC Curve* (AUC) yang disajikan pada

Tabel 4.17. Setelah mendapatkan nilai *fitness* dari masing-masing kromosom, langkah berikutnya adalah melakukan pembentukan kromosom orang tua dengan menggunakan metode seleksi *Roulette Wheel Selection* (RWS). Langkah awal pada proses RWS adalah menghitung proporsi nilai *fitness* yaitu dengan membagi nilai *fitness* pada masing-masing kromosom dengan jumlah *fitness*. Kemudian menghitung nilai kumulatif dari proporsi nilai *fitness* tersebut. Langkah selanjutnya adalah membangkitkan 100 bilangan *random* seperti disajikan pada Tabel 4.18.

Tabel 4.18 Ilustrasi Proses RWS pada Estimasi Parameter

Populasi	<i>Fitness</i>	Proporsi Nilai <i>Fitness</i>	Nilai <i>Fitness</i> Kumulatif	<i>Random</i> Number
1	0.745	0.014	0.011	0.025
2	0.745	0.014	0.027	0.979
3	0.745	0.014	0.041	0.778
⋮	⋮	⋮	⋮	⋮
99	0.423	0.008	0.992	0.419
100	0.425	0.008	1.000	0.041

Bilangan *random* pertama yaitu sebesar 0.025 seperti disajikan pada Tabel 4.18, maka kromosom orangtua 1 didapatkan dari nilai *fitness* kumulatif yang lebih besar dari 0.025 yang pertama yaitu kromosom 2, maka kromosom tersebut menjadi kromosom 1 pada populasi baru. Bilangan *random* diperoleh dengan membangkitkan 100 data berdistribusi *uniform* dengan batas bawah 0 dan batas atas 1. Tahapan tersebut berhenti ketika telah diperoleh 100 kromosom orang tua berdasarkan bilangan *random* yang telah dibangkitkan.

Setelah mendapatkan 100 kromosom orang tua, dilakukan proses pindah silang atau *crossover*. Proses pindah silang menghasilkan kromosom baru dari hasil perpaduan 2 kromosom orang tua. Metode pindah silang yang digunakan adalah *single point crossover*, yaitu dua kromosom dipotong sekali pada titik-titik yang sesuai dan bagian setelah pemotongan dipertukarkan. Pada penelitian ini peluang pindah silang yang digunakan yaitu 0.8.

Orangtua	Gen 1	Gen 2	Gen 3	Gen 4	...	Gen 12	<i>random</i>
1	-0.420	2.596	6.475	1.340	...	5.284	0.77
2	-0.420	2.596	6.475	0.822	...	5.284	0.23
Anak	Pindah Silang						
1	-0.420	2.596	6.475	0.822	...	5.284	
2	-0.420	2.596	6.475	1.340	...	5.284	

Gambar 4.19 Ilustrasi *Single Point Crossover* pada Estimasi Parameter

Kromosom orang tua yang mengalami pindah silang adalah kromosom yang memiliki bilangan *random* kurang dari probabilitas pindah silang yaitu sebesar 0.8. Salah satu contoh kromosom yang mengalami pindah silang dapat dilihat pada Gambar 4.19. Bilangan *random* pada kromosom orangtua pertama dan ke-2 kurang dari 0.8 sehingga akan dilakukan pindah silang pada kedua kromosom tersebut. Langkah pertama adalah menentukan posisi pindah silang. Cara menentukan posisi pindah silang adalah dengan membangkitkan bilangan *random* antara 1 sampai 12. Pada ilustrasi tersebut bilangan *random* yang terpilih adalah 3 yang ditunjukkan dengan garis putus-putus sehingga kromosom induk akan menghasilkan anak yang telah berpindah silang antara kedua kromosom induknya pada titik ke-3. Proses pindah silang ini dilanjutkan untuk semua pasang kromosom yang memiliki bilangan *random* kurang dari peluang pindah silang.

Proses selanjutnya adalah melakukan mutasi gen menggunakan mutasi *uniform* yaitu dengan memberikan kesempatan yang sama pada setiap gen untuk dilakukan proses mutasi. Tahapan awal mutasi adalah dengan membangkitkan bilangan *random* pada setiap gen. Bilangan *random* tersebut kemudian dibandingkan dengan peluang mutasi yaitu sebesar 0.1. Apabila nilai bilangan *random* pada suatu gen kurang dari peluang mutasi, maka proses mutasi akan dilakukan pada gen tersebut. Proses mutasi dilakukan dengan mengganti nilai estimasi parameter dengan bilangan *random*. Gambar 4.20 menjelaskan terjadinya proses mutasi pada kromosom pertama. Gen yang mengalami mutasi adalah gen yang memiliki bilangan *random* kurang dari peluang mutasi 0.1 yaitu pada gen pertama dan gen ke-3. Gen

pertama diganti menjadi 0.880 dan gen ke-3 berubah menjadi 3.945.

Kromosom	Gen 1	Gen 2	Gen 3	...	Gen 12
Bil. <i>random</i>	0.05	0.91	0.01	...	0.11
1	-0.420	2.596	6.475	...	5.284
Anak	Mutasi				
1	0.880	2.596	3.945	...	5.284

Gambar 4.20 Ilustrasi Proses Mutasi pada Estimasi Parameter

Proses selanjutnya adalah elitisme yaitu proses pengopian individu untuk mempertahankan kromosom terbaik dalam populasi. Kromosom yang dipertahankan sebesar 20% dari total kromosom dalam populasi.

Tabel 4.19 Kromosom dan Nilai *Fitness* Algoritma Genetika

Populasi	Kromosom					Fitness
	Gen1	Gen 2	Gen 3	...	Gen 12	
1	-0.420	0.978	6.475	...	5.284	0.8125
2	-0.420	0.978	6.475	...	5.284	0.8125
3	-0.420	0.978	6.475	...	5.284	0.8125
⋮	⋮	⋮	⋮	⋮	⋮	⋮
99	-0.420	2.450	6.475	...	5.284	0.8125
100	-0.420	2.450	6.475	...	0.437	0.7604

Hasil estimasi parameter regresi logistik ordinal hasil dari seleksi *backward* menggunakan algoritma genetika dilihat berdasarkan nilai *fitness* tertinggi yang disajikan pada Tabel 4.20.

Tabel 4.20 Fungsi Logit Estimasi Parameter Regresi Logistik Ordinal Hasil Seleksi *Backward* menggunakan Algoritma Genetika berdasarkan Nilai AUC

Fungsi Logit	
$\hat{g}_1(x)$	$= -0.420 + 1.340 X_{1(2)} + 9.524 X_2 + 2.889X_5 + 5.892X_6 - 12.446X_8 - 6.602X_9 + 0.472X_{10} - 2.384X_{11} + 5.284X_{13}$
$\hat{g}_2(x)$	$= 0.978 + 1.340 X_{1(2)} + 9.524 X_2 + 2.889X_5 + 5.892X_6 - 12.446X_8 - 6.602X_9 + 0.472X_{10} - 2.384X_{11} + 5.284X_{13}$
$\hat{g}_3(x)$	$= 6.475 + 1.340 X_{1(2)} + 9.524 X_2 + 2.889X_5 + 5.892X_6 - 12.446X_8 - 6.602X_9 + 0.472X_{10} - 2.384X_{11} + 5.284X_{13}$

Nilai *fitness* tertinggi yang didapatkan berdasarkan kriteria ketepatan klasifikasi AUC adalah 0.8125 dengan fungsi logit yang terbentuk disajikan pada Tabel 4.20.

Tabel 4.21 Fungsi Logit Estimasi Parameter Regresi Logistik Ordinal Hasil Seleksi *Backward* menggunakan Algoritma Genetika berdasarkan Nilai *G_mean*

Fungsi Logit	
$\hat{g}_1(x)$	$= 0.550 + 1.340 X_{1(2)} + 9.524 X_2 + 2.889X_5 + 5.892X_6 -$ $12.446X_8 - 6.602X_9 + 0.773X_{10} + 0.632X_{11} + 5.284X_{13}$
$\hat{g}_2(x)$	$= 2.596 + 1.340 X_{1(2)} + 9.524 X_2 + 2.889X_5 + 5.892X_6 -$ $12.446X_8 - 6.602X_9 + 0.773X_{10} + 0.632X_{11} + 5.284X_{13}$
$\hat{g}_3(x)$	$= 6.475 + 1.340 X_{1(2)} + 9.524 X_2 + 2.889X_5 + 5.892X_6 -$ $12.446X_8 - 6.602X_9 + 0.773X_{10} + 0.632X_{11} + 5.284X_{13}$

Estimasi parameter pada algoritma genetika menggunakan fungsi *fitness* yang lain yaitu *Geometric mean* yang didapatkan melalui tahapan proses yang sama. Adapun nilai *Geometric mean* terbaik yang didapatkan adalah sebesar 0.8003 dengan hasil estimasi parameter disajikan pada Tabel 4.21. Variabel yang terpilih pada saat seleksi menggunakan metode *backward*, *forward* dan *stepwise* menghasilkan hasil yang sama sehingga hasil estimasi parameter untuk metode *forward* dan *stepwise* sama dengan hasil pada *backward*.

4.7 Optimasi Parameter Regresi Logistik Ordinal Hasil Seleksi Algoritma Genetika

Langkah pertama yang dilakukan adalah merepresentasikan model dari persamaan regresi logistik ordinal hasil dari seleksi algoritma genetika ke dalam kromosom dan menentukan nilai inisialisasi. Kromosom yang digunakan berisi nilai estimasi parameter dari fungsi logit. Panjang kromosom sesuai dengan jumlah estimasi parameter fungsi logit yaitu sebanyak 14 gen dengan representasi kromosom awal adalah sebagai berikut.

α_1	α_2	α_3	$\beta_{1(2)}$...	β_{13}
1.372	4.192	8.015	1.350	...	4.658

Gambar 4.21 Representasi Kromosom Awal Estimasi Parameter Hasil Seleksi Algoritma Genetika

Gambar 4.21 merupakan salah satu nilai inisialisasi pada kromosom pertama yang terdiri dari 14 gen. Proses algoritma genetika digunakan untuk mendapatkan nilai optimum dari estimasi parameter fungsi logit yang memaksimalkan ketepatan klasifikasi. Setelah dilakukan inisialisasi dilakukan pembentukan 100 kromosom awal yang dibangkitkan untuk populasi awal yang disajikan pada Tabel 4.24.

Tabel 4.22 Ilustrasi Kromosom Dalam Populasi Awal pada Estimasi Parameter Hasil Seleksi Algoritma Genetika

Populasi	Kromosom				
	Gen1	Gen 2	Gen 3	...	Gen 12
1	1.372	4.192	8.015	...	4.658
2	1.372	4.192	8.015	...	4.658
3	1.372	4.192	8.015	...	4.658
⋮	⋮	⋮	⋮	⋮	⋮
99	1.088	4.192	8.015	...	5.284
100	1.372	4.192	9.342	...	4.658

Langkah selanjutnya setelah terbentuk 100 kromosom awal sama dengan proses optimasi pada hasil seleksi *backward*. Kromosom tersebut akan mengalami proses seleksi orang tua menggunakan, pindah silang, mutasi dan elitisme hingga mendapatkan nilai *fitness* yang konvergen.

Tabel 4.23 Kromosom dan Nilai *Fitness* Generasi Terakhir Hasil Seleksi Algoritma Genetika

Populasi	Kromosom					Fitness
	Gen1	Gen 2	Gen 3	...	Gen 12	
1	1.372	4.192	8.015	...	4.658	0.752
2	1.372	4.192	8.015	...	4.658	0.752
3	1.372	4.192	8.015	...	4.658	0.752
⋮	⋮	⋮	⋮	⋮	⋮	⋮
99	1.372	0.111	8.015	...	4.658	0.250
100	1.372	4.192	8.015	...	4.658	0.752

Hasil estimasi parameter regresi logistik ordinal hasil dari seleksi algoritma genetika dilihat berdasarkan nilai *fitness* tertinggi yang disajikan pada Tabel 4.25. Nilai *fitness* tertinggi

yang didapatkan berdasarkan kriteria AUC adalah 0.752 dengan fungsi logit yang terbentuk seperti disajikan pada Tabel 4.26.

Tabel 4.24 Fungsi Logit Estimasi Parameter Regresi Logistik Ordinal Hasil Seleksi Algoritma Genetika berdasarkan Nilai AUC

Fungsi Logit	
$\hat{g}_1(x)$	$= 1.372 + 0.542X_{1(2)} + 9.202 X_2 + 0.268X_3 + 0.662X_4 + 2.148X_5 + 7.637 X_6 - 10.176X_8 - 6.458X_9 + 2.047X_{10} + 0.700X_{12} + 4.658X_{13}$
$\hat{g}_2(x)$	$= 4.192 + 0.542X_{1(2)} + 9.202 X_2 + 0.268X_3 + 0.662X_4 + 2.148X_5 + 7.637 X_6 - 10.176X_8 - 6.458X_9 + 2.047X_{10} + 0.700X_{12} + 4.658X_{13}$
$\hat{g}_3(x)$	$= 8.015 + 0.542X_{1(2)} + 9.202 X_2 + 0.268X_3 + 0.662X_4 + 2.148X_5 + 7.637 X_6 - 10.176X_8 - 6.458X_9 + 2.047X_{10} + 0.700X_{12} + 4.658X_{13}$

Hasil klasifikasi pada regresi logistik ordinal menggunakan seleksi algoritma genetika ditampilkan pada Tabel 4.25.

Tabel 4.25 *Confusion Matriks* Klasifikasi Regresi Logistik Ordinal Hasil Seleksi Algoritma Genetika menggunakan AUC

Aktual	Prediksi			
	II	III	IV	V
II	11	2	3	0
III	3	7	2	0
IV	0	0	8	0
V	0	0	0	1

Hasil ketepatan klasifikasi menggunakan kriteria *Area Under ROC Curve* (AUC) yaitu sebesar 75.22% didapatkan melalui perhitungan sebagai berikut.

$$AUC = \frac{1}{4}(0.69 + 0.58 + 1 + 1) = 0.7522$$

Tabel 4.26 Fungsi Logit Estimasi Parameter Regresi Logistik Ordinal Hasil Seleksi Algoritma Genetika berdasarkan Nilai G_{mean}

Fungsi Logit	
$\hat{g}_1(x)$	$= 1.372 + 0.993X_{1(2)} + 9.202 X_2 + 0.914X_3 - 1.598X_4 + 2.148X_5 + 7.637 X_6 - 10.176X_8 - 6.458X_9 + 2.047X_{10} + 0.644X_{12} + 4.658X_{13}$
$\hat{g}_2(x)$	$= 4.192 + 0.993X_{1(2)} + 9.202 X_2 + 0.914X_3 - 1.598X_4 + 2.148X_5 + 7.637 X_6 - 10.176X_8 - 6.458X_9 + 2.047X_{10} + 0.644X_{12} + 4.658X_{13}$
$\hat{g}_3(x)$	$= 8.015 + 0.993X_{1(2)} + 9.202 X_2 + 0.914X_3 - 1.598X_4 + 2.148X_5 + 7.637 X_6 - 10.176X_8 - 6.458X_9 + 2.047X_{10} + 0.644X_{12} + 4.658X_{13}$

Kriteria nilai *fitness* lain yang digunakan adalah *Geometric mean*. Estimasi parameter pada algoritma genetika menggunakan *Geometric mean* didapatkan melalui tahapan proses yang sama. Adapun nilai *Geometric mean* terbaik yang didapatkan adalah sebesar 0.7887 dengan hasil estimasi parameter disajikan pada Tabel 4.26.

4.8 Naïve Bayes Classification Pasien Pneumonia

Sebelum melakukan klasifikasi menggunakan metode *Naïve Bayes*, langkah pertama yang harus dilakukan adalah membagi data menjadi data *training* dan *testing* menggunakan *5-fold cross validation*. Pada klasifikasi berikut, data yang digunakan merupakan data *training* dan data *testing* yang menghasilkan nilai akurasi tertinggi berdasarkan kriteria *Geometric mean* dan AUC dengan melibatkan seluruh variabel.

Langkah awal yang dilakukan untuk melakukan klasifikasi menggunakan *Naïve Bayes* adalah dengan menghitung nilai probabilitas *prior*. Data *training* yang digunakan sebanyak 148 observasi yang terdiri atas 64 observasi pada kelas II, 47 observasi pada kelas III, 31 observasi pada kelas IV dan enam observasi pada kelas V. Nilai probabilitas *prior* disajikan pada tabel 4.27 yang didapatkan dengan membagi jumlah observasi pada masing-masing kelas dengan keseluruhan observasi pada data *training*.

Tabel 4.27 Nilai Probabilitas *Prior* Data Training

Kelas	Probabilitas
II	0.4362
III	0.3154
IV	0.2081
V	0.0403

Langkah selanjutnya adalah menghitung nilai peluang bersyarat pada tiap kelas risiko untuk setiap variabel independen. Nilai peluang bersyarat pada variabel independen yang bersifat kategorik yaitu jenis kelamin ditampilkan pada Tabel 4.28. Nilai peluang bersyarat untuk pasien perempuan yang berasal dari kelas risiko II didapatkan dengan membagi jumlah observasi perempu-

an pada kelas II dengan jumlah observasi keseluruhan pada kelas II.

Tabel 4.28 Nilai Peluang Bersyarat Variabel Jenis Kelamin

Jenis Kelamin	Kelas II	Kelas III	Kelas IV	Kelas V
Perempuan (1)	0.4769	0.3191	0.2258	0.0000
Laki-laki (2)	0.5231	0.6809	0.7742	1.0000

Pada variabel yang bersifat kontinu nilai peluang bersyarat dihitung menggunakan pendekatan distribusi normal sehingga diperlukan nilai rata-rata dan standar deviasi dari masing-masing variabel independen pada setiap kelas risiko. Nilai rata-rata dan standar deviasi ditampilkan pada Tabel 4.29.

Tabel 4.29 Rata-rata dan Standar Deviasi Tiap Kategori Kelas Risiko Data *Training*

Variabel	Kelas II		Kelas III		Kelas IV		Kelas V	
	<i>mean</i>	<i>std</i>	<i>mean</i>	<i>std</i>	<i>mean</i>	<i>std</i>	<i>mean</i>	<i>std</i>
Usia (X_2)	47.37	12.91	57.96	12.28	64.26	10.97	68.83	11.00
Sist. (X_3)	122.54	20.35	121.21	18.95	116.03	23.40	121.67	27.14
Diast. (X_4)	75.45	10.77	75.87	13.12	71.81	13.31	75.00	12.25
Denyut (X_5)	104.63	15.28	105.28	18.39	108.26	15.72	124.00	13.22
Napas (X_6)	26.29	5.47	29.00	7.45	29.61	7.25	34.67	6.15
Suhu (X_7)	37.09	1.04	37.09	0.83	37.13	1.01	37.70	1.13
pH (X_8)	7.45	0.08	7.43	0.09	7.37	0.12	7.35	0.10
Natr. (X_9)	137.06	5.82	134.79	7.29	136.71	7.77	137.87	11.07
Gluk. (X_{10})	147.89	79.28	168.49	96.95	173.03	90.82	178.67	98.70
HMT (X_{11})	38.08	5.97	37.91	6.56	37.44	6.65	36.78	8.12
PO ₂ (X_{12})	108.94	54.73	94.14	38.90	126.39	70.88	95.38	28.57
BUN (X_{13})	15.07	11.63	18.11	11.10	23.68	14.74	36.00	13.24

Nilai rata-rata dan standar deviasi dari data *training* tiap kategori kelas risiko dalam Tabel 4.29 digunakan untuk menentukan peluang tiap kategori pada data *testing* menggunakan persamaan (2.22). Tabel 4.30 menunjukkan peluang parsial tiap kategori kelas risiko pada data *testing* pertama.

Tabel 4.30 Peluang Parsial pada Tiap Kategori Data *Testing* Pertama

x_{1i}	$P(x_{1i} kls. II)$	$P(x_{1i} kls. III)$	$P(x_{1i} kls. IV)$	$P(x_{1i} kls. V)$	
JK (X_1)	2.0	0.5231	0.6809	0.7742	1.0000
Usia (X_2)	40.0	0.0263	0.0112	0.0032	0.0012
Sist. (X_3)	120.0	0.0195	0.0210	0.0168	0.0147

Tabel 4.31 Peluang Parsial pada Tiap Kategori Data *Testing* Pertama

	x_{1i}	$P(x_{1i} kls. II)$	$P(x_{1i} kls. III)$	$P(x_{1i} kls. IV)$	$P(x_{1i} kls. V)$
Diast. (X_4)	80.0	0.0339	0.0289	0.0248	0.0300
Denyut (X_5)	110.0	0.0246	0.0210	0.0252	0.0172
Napas(X_6)	28.0	0.0694	0.0530	0.0537	0.0361
Suhu (X_7)	37.3	0.3755	0.4669	0.3882	0.3313
pH (X_8)	7.39	3.5818	3.8758	3.1704	3.7193
Natr. (X_9)	136.0	0.0675	0.0540	0.0511	0.0355
Gluk. (X_{10})	101.0	0.0042	0.0032	0.0032	0.0030
HCT (X_{11})	36.4	0.0642	0.0593	0.0592	0.0491
PO ₂ (X_{12})	67.0	0.0054	0.0080	0.0040	0.0085
BUN(X_{13})	8.0	0.0285	0.0237	0.0154	0.0032

Setelah menemukan peluang parsial tiap variabel pada setiap kategori kelas risiko, langkah selanjutnya adalah menghitung peluang posterior dikali dengan peluang setiap kelas yang kemudian digunakan untuk menentukan klasifikasi kelompok kelas risiko pada data *testing*.

$$\begin{aligned} P(X_1, X_2, \dots, X_5|Y_2) \cdot P(Y_2) &= P(x_{1.1}|Y_2) \cdot P(x_{1.2}|Y_2) \dots P(x_{1.5}|Y_2) \cdot P(Y_2) \\ &= 0.5230 \times 0.0263 \times \dots \times 0.0285 \times 0.4362 \\ &= 2.5682 \times 10^{-17} \end{aligned}$$

$$\begin{aligned} P(X_1, X_2, \dots, X_5|Y_3) \cdot P(Y_3) &= P(x_{1.1}|Y_3) \cdot P(x_{1.2}|Y_3) \dots P(x_{1.5}|Y_3) \cdot P(Y_3) \\ &= 0.6809 \times 0.0112 \times \dots \times 0.0237 \times 0.3154 \\ &= 5.7820 \times 10^{-18} \end{aligned}$$

$$\begin{aligned} P(X_1, X_2, \dots, X_5|Y_4) \cdot P(Y_4) &= P(x_{1.1}|Y_4) \cdot P(x_{1.2}|Y_4) \dots P(x_{1.5}|Y_4) \cdot P(Y_4) \\ &= 0.7742 \times 0.0032 \times \dots \times 0.0154 \times 0.2081 \\ &= 2.0858 \times 10^{-19} \end{aligned}$$

$$\begin{aligned} P(X_1, X_2, \dots, X_5|Y_5) \cdot P(Y_5) &= P(x_{1.1}|Y_5) \cdot P(x_{1.2}|Y_5) \dots P(x_{1.5}|Y_5) \cdot P(Y_5) \\ &= 1 \times 0.0012 \times \dots \times 0.0032 \times 0.0403 \\ &= 2.2406 \times 10^{-21} \end{aligned}$$

Hasil perhitungan peluang posterior pada data *testing* pertama menunjukkan bahwa peluang terbesar adalah kategori kelas risiko II maka pasien tersebut akan diklasifikasikan ke dalam kelas risiko II. Perhitungan tersebut dilakukan untuk semua data *testing*. Hasil dari klasifikasi ditunjukkan pada Tabel 4.31. Nilai *recall* (R) untuk setiap kategori didapatkan dari perhitungan

nilai yang tepat diklasifikasikan dibagi dengan total observasi setiap kelas. Masing-masing nilai R adalah sebagai berikut.

$$R_1 = \frac{13}{16} = 0.8125; R_2 = 0.6363; R_3 = 0.5; R_4 = 0.5$$

Tabel 4.32 *Confusion Matrix* Klasifikasi *Naïve Bayes* Data Testing

Aktual	Prediksi			
	II	III	IV	V
II	13	2	1	0
III	2	7	2	0
IV	0	3	4	1
V	0	0	1	1

Perhitungan ketepatan klasifikasi yang digunakan adalah *Area Under ROC Curve* (AUC) dan *Geometric mean* dengan masing-masing nilai adalah 0.5996 dan 0.6122. Berikut ini adalah perhitungan dari AUC dan *Geometric mean*.

$$AUC = \frac{1}{4}(0.81 + 0.63 + 0.50 + 0.50) = 0.6122$$

$$G_mean = (0.81 \times 0.63 \times 0.50 \times 0.50)^{\frac{1}{4}} = 0.5996$$

4.9 Perbandingan Hasil Ketepatan Klasifikasi Regresi Logistik Ordinal, *Hybrid* Regresi Logistik Ordinal-Algoritma Genetika dan *Naïve Bayes Classification*

Setelah melakukan klasifikasi pada data pasien pneumonia menggunakan regresi logistik ordinal dengan seleksi variabel *backward elimination*, *forward selection*, dan *stepwise*, *hybrid* regresi logistik ordinal-algoritma genetika dan *Naïve Bayes Classification*, langkah selanjutnya adalah membandingkan hasil ketepatan klasifikasi dari beberapa metode tersebut. Pemilihan metode terbaik ditinjau dari kesesuaian variabel terpilih dan nilai ketepatan klasifikasinya. Pada regresi logistik ordinal menggunakan seleksi *backward*, *forward* dan *stepwise* terpilih 9 variabel yang sama yaitu jenis kelamin (X_1), usia (X_2), denyut (X_5), napas (X_6), pH (X_8), natrium (X_9), glukosa (X_{10}), hema-

tokrit (X_{11}), dan BUN (X_{13}). Secara visual terdapat perbedaan rata-rata antar kelas risiko pada masing-masing variabel, namun pada variabel natrium, glukosa dan hematokrit perbedaannya tidak signifikan. Hal ini bisa terjadi karena dalam memodelkan variabel tersebut dilakukan secara simultan atau serentak sehingga terdapat interaksi dengan variabel lain yang dapat menyebabkan variabel tersebut terpilih dalam seleksi. Selain itu adanya data *outlier* yang tidak diatasi menyebabkan varians data menjadi besar sehingga secara visual rata-ratanya terlihat tidak berbeda signifikan namun sebenarnya berbeda sehingga masuk ke dalam model saat proses seleksi. Jika ditinjau dari tanda pada masing-masing koefisien parameter nilainya telah sesuai dengan *boxplot* pada Lampiran 2 dan sesuai dengan teori kesehatan.

Pada proses seleksi variabel menggunakan metode algoritma genetika terpilih sebanyak 11 variabel yaitu jenis kelamin (X_1), usia (X_2), sistolik (X_3), diastolik (X_4), denyut (X_5), napas (X_6), pH (X_8), natrium (X_9), glukosa (X_{10}), PO_2 (X_{12}), dan BUN (X_{13}). Berdasarkan *boxplot* pada Lampiran 3 hampir semua variabel memiliki perbedaan rata-rata yang signifikan antar kelasnya kecuali pada variabel tekanan darah sistolik dan PO_2 . Namun karena dalam memodelkan variabel dilakukan secara serentak, kedua variabel tersebut terpilih masuk ke dalam model. Hal ini bisa terjadi karena adanya efek interaksi dengan variabel yang lain. Jika ditinjau dari tanda pada masing-masing koefisien parameter, tanda tersebut telah sesuai dengan *boxplot* dan teori kesehatan. Meskipun secara visual terdapat perbedaan rata-rata antar kelasnya, rata-rata tekanan darah sistolik dan PO_2 pada masing-masing kelas masih dalam rentang yang normal sehingga perbedaannya tidak terlihat signifikan. Berdasarkan uraian tersebut, jika ditinjau dari aspek kesesuaian variabel terpilih keempat metode seleksi menunjukkan performansi yang sama. Selain dilakukan analisis terhadap kesesuaian variabel terpilih, kriteria pemilihan model terbaik dapat dilihat berdasarkan nilai

ketepatan klasifikasinya. Ringkasan hasil klasifikasi dan variabel yang terlibat ditunjukkan dalam Tabel 4.32 di mana nilai ketepatan klasifikasi tersebut merupakan hasil rata-rata dari kelima model yang didapatkan dari pembagian data *training* dan data *testing* dengan *5-fold cross validation*.

Tabel 4.33 Ketepatan Klasifikasi Masing-masing Metode

Metode	Variabel	Ketepatan Klasifikasi	
		AUC	<i>G_mean</i>
Regresi Logistik	Semua variabel	0.5896	0.4014
	<i>Backward, Forward, Stepwise</i> ($X_1, X_2, X_5, X_6, X_8, X_9, X_{10}, X_{11}, X_{13}$)	0.5975	0.4140
	Algoritma Genetika ($X_1, X_2, X_3, X_4, X_5, X_6, X_8, X_9, X_{10}, X_{12}, X_{13}$)	0.6069	0.3994
<i>Hybrid</i> Reglog-AG	Semua variabel	0.7215	0.7067
	<i>Backward, Forward, Stepwise</i> ($X_1, X_2, X_5, X_6, X_8, X_9, X_{10}, X_{11}, X_{13}$)	0.7227	0.7077
	Algoritma Genetika ($X_1, X_2, X_3, X_4, X_5, X_6, X_8, X_9, X_{10}, X_{12}, X_{13}$)	0.7430	0.7091
<i>Naïve</i> <i>Bayes</i>	Semua variabel	0.4671	0.2030
	<i>Backward, Forward, Stepwise</i> ($X_1, X_2, X_5, X_6, X_8, X_9, X_{10}, X_{11}, X_{13}$)	0.4556	0.1285
	Algoritma Genetika ($X_1, X_2, X_3, X_4, X_5, X_6, X_8, X_9, X_{10}, X_{12}, X_{13}$)	0.4582	0.1876

Hasil ringkasan ketepatan klasifikasi pada Tabel 4.32 menunjukkan bahwa nilai tertinggi didapatkan pada saat menggunakan metode *hybrid* regresi logistik ordinal-algoritma genetika yang melibatkan variabel terpilih dari seleksi algoritma genetika. Nilai AUC yang didapatkan sebesar 0.7430 yang artinya 74.30% pasien telah diklasifikasikan dengan tepat sesuai dengan kelas risikonya. Nilai *geometric mean* yang didapatkan sebesar 0.7091 yang artinya terdapat 70.91% pasien telah diklasifikasikan dengan tepat sesuai dengan kelas risikonya. Selain itu tanda parameter

pada masing-masing variabel yang terpilih telah sesuai dengan tanda yang seharusnya sesuai dengan kondisi pada dunia medis.

4.10 Perbandingan Hasil Ketepatan Klasifikasi dengan Metode Sebelumnya

Pada penelitian sebelumnya analisis klasifikasi terhadap kelas risiko pasien pneumonia di RSUD Dr. Soetomo pernah dilakukan oleh Firdausanti (2017) menggunakan metode *Hybrid Analisis Diskriminan – Particle Swarm Optimization* (PSO). Adapun perbandingan hasil klasifikasi dari metode sebelumnya disajikan pada Tabel 4.35

Tabel 4.34 Perbandingan Ketepatan Klasifikasi dengan Metode Sebelumnya

Metode	Ketepatan Klasifikasi	
	AUC	<i>G_mean</i>
<i>Hybrid Analisis Diskriminan – PSO</i>	0.7163	0.6736
<i>Hybrid Regresi Logistik Ordinal- Algoritma Genetika</i>	0.7430	0.7091

Tabel 4.33 menunjukkan bahwa nilai ketepatan klasifikasi pada saat menggunakan *Hybrid Regresi Logistik Ordinal- Algoritma Genetika* lebih tinggi dibandingkan dengan metode sebelumnya sehingga dapat disimpulkan bahwa metode ini lebih baik dibandingkan dengan yang sebelumnya.

(Halaman ini sengaja dikosongkan)

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan analisis yang telah dilakukan maka didapatkan beberapa kesimpulan yaitu sebagai berikut.

1. Pasien pneumonia yang dirawat di RSUD Dr. Soetomo pada tahun 2015 paling banyak merupakan pasien pada kelas risiko II yaitu sebesar 44%. Kemudian jumlah pasien terbanyak kedua adalah pada kelas risiko III yaitu sebesar 31%. Proporsi pasien pneumonia pada kelas IV sebesar 21% dan yang paling sedikit adalah pasien pneumonia pada kelas risiko V yaitu hanya sebanyak 4%.
2. Metode *hybrid* Regresi Logistik Ordinal – Algoritma Genetika lebih baik dibandingkan metode Regresi Logistik Ordinal dan *Naïve Bayes Classification* dengan nilai AUC dan *Geometric mean* sebesar 74.30% dan 70.91%.
3. Nilai ketepatan klasifikasi yang didapatkan lebih baik dari penelitian sebelumnya yang dilakukan oleh Firdausanti (2017) dengan menggunakan metode *hybrid* Analisis Diskrimininan – *Particle Swarm Optimization* (PSO).

5.2 Saran

Berdasarkan penelitian yang telah dilakukan maka saran yang dapat diberikan oleh peneliti adalah menggunakan metode lain yang lebih sensitif terhadap data yang bersifat *unbalanced*. Saran yang dapat diberikan kepada tim medis di RSUD Dr. Soetomo adalah mempertimbangkan metode statistika seperti algoritma genetika dalam mengklasifikasikan pasien pneumonia agar pasien lebih cepat ditangani secara tepat.

(Halaman ini sengaja dikosongkan)

DAFTAR PUSTAKA

- Agresti, A. (2002). *Categorical Data Analysis Second Edition*. Florida: John Wiley & Sons, Inc.
- Ashari, A., Paryudi, I., & Tjoa, A. (2013). Performance Comparison Between Naïve Bayes, Decision Tree, and K-Nearest Neighbor in Searching Alternative Design in An Energy Simulation Tool. *Int. J. Adv. Comput. Sci. Appl*, 33-39.
- Bekkar, M., Djemaa, D. K., & Alitouche, D. A. (2013). Evaluation Measures for Model Assessment over Imbalanced Data Sets. *Journal of Information Engineering and Application*, 27-38.
- Depkes. (2005). Pharmaceutical Care untuk Penyakit Infeksi Saluran Pernapasan.
- Firdausanti, N. A., Irhamah, & Wulandari, L. (2017). *Klasifikasi Kelas Risiko Pneumonia di RSUD Dr. Soetomo menggunakan Metode Hybrid Analisis Diskriminan-Particle Swarm Optimization (ADL-PSO) dan Naïve Bayes*. Surabaya: ITS.
- Fridayanthie, E. W. (2015). Analisa Data Mining untuk Prediksi Penyakit Hepatitis dengan menggunakan Metode Naive Bayes dan Support Vector Machine . *Jurnal Katulistiwa Informatika*, 25-36.
- Gen, M., & Cheng, R. (1997). *Genetic Algorithms and Engineering Design*. New York: John Wiley & Sons, Inc.
- Gorunescu, F. (2011). *Data Mining Concepts, Models and Technique*. Jerman: Springer.
- Gradiana, M. D., & Irhamah. (2014). Klasifikasi Pasien Penderita Diabetes Melitus Tipe Dua Menggunakan Metode Analisis Diskriminan Hybrid Algoritma Genetika. *Jurnal Sains dan Seni POMITS*.
- Gunawan, Armanto, H., Santoso, J., Giovanni, D., Kurniawan, F., Yudianto, R., et al. (2012). Evolutionary Neural Network

- for Othello Game. *Procedia-Social and Behavioral Sciences*, 419-425.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Technique Third Edition*. United States of America: Elsevier Inc.
- Hosmer, D., & Lemeshow, S. (2000). *Applied Logistic Regression Second Edition*. Canada: John Wiley & Sons, Inc.
- 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.
- Johnson, P., Vandewater, L., Wilson, W., Maruff, P., & Savage, G. (2014). Genetic Algorithm with Logistic Regression for Prediction of Progression to Alzheimer's Disease. *BMC Bioinformatics*, 1-14.
- Kemenkes. (2014). Profil Kesehatan Indonesia 2014.
- Kusrini, E. (2009). *Algoritma Data Mining*. Yogyakarta: Andi.
- Nursalam. (2006). *Asuhan Keperawatan Pada Pasien dengan Gangguan Sistem*. Jakarta: Salemba Medika.
- PDPI. (2003). *Pneumonia Komuniti: Pedoman Diagnosis & Penatalaksanaan di Indonesia*.
- Rahimatun, S. M., Irhamah, & Wulandari, L. (2017). *Klasifikasi Penderita Pneumonia menggunakan Metode Analisis Diskriminan, Hybrid Analisis Diskriminan-Algoritma Genetika dan Nai ve Bayes*. Surabaya: ITS.
- Santoso, F. P., Wulandari, S. P., & Purhadi. (2012). Faktor-faktor Eksternal Pneumonia pada Balita di Jawa Timur dengan Pendekatan Geographically Weighted Regression. *Jurnal Sains dan Seni POMITS*.
- 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.

- Smeltzer, S. C., & Bare, B. G. (2001). *Buku Ajar Keperawatan Medikal*. Jakarta: EGC.
- Suyanto. (2005). *Algoritma Genetika dalam Matlab*. Yogyakarta: Andi Offset.
- Walpole, R., Myers, R., Myers, S., & Ye, K. (2011). *Probability & Statistics for Engineers and Scientists Ninth Edition*. United States: Prentice Hall.
- WHO. (2007). Pencegahan dan Pengendalian Infeksi Saluran Pernapasan Akut (ISPA) yang Cenderung menjadi Pandemi di Fasilitas Pelayanan Kesehatan. *Pedoman Interim WHO*.
- Wu, X., & Kumar, V. (2009). *The Top Ten Algorithms in Data Mining*. New York: CRC Press.
- 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.

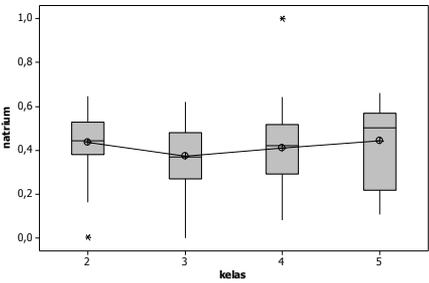
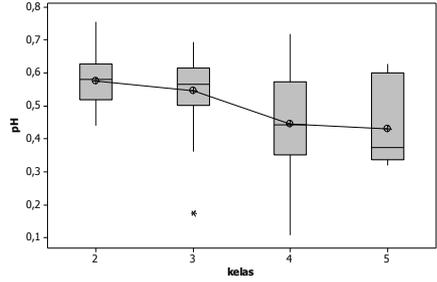
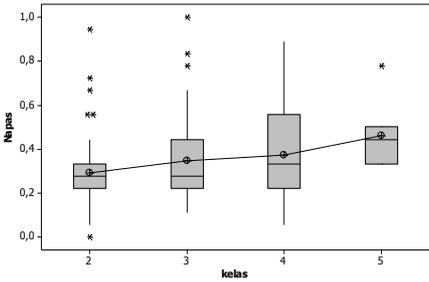
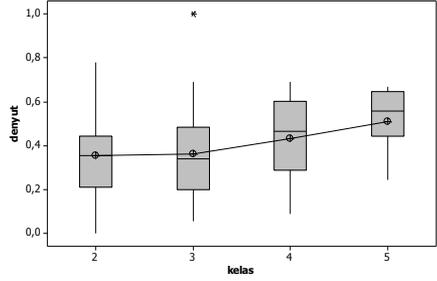
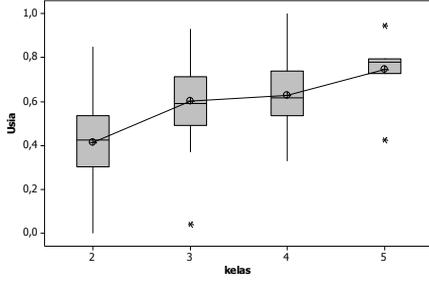
(Halaman ini sengaja dikosongkan)

LAMPIRAN

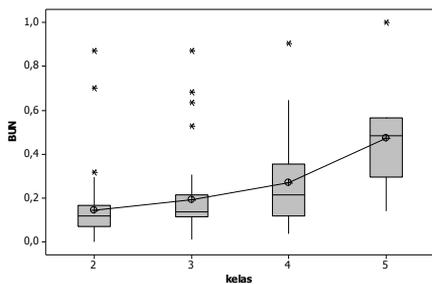
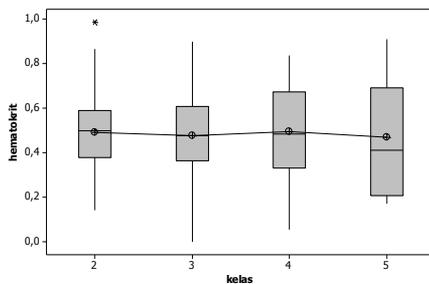
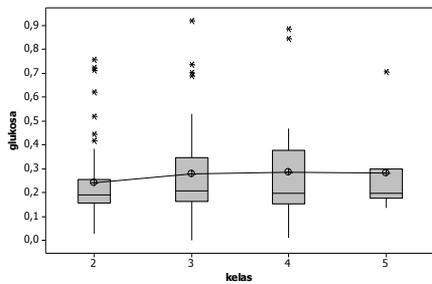
Lampiran 1. Data Pasien Pneumonia

Kelas	Usia	Sistolik	Diastolik	...	Hematokrit	PO₂	BUN
2	2	54	100	...	60	41.4	112
2	1	24	100	...	70	30.2	147
2	2	47	100	...	60	41.8	100
2	1	58	125	...	86	44.3	54.8
2	2	50	100	...	70	40.4	56.4
∴	∴	∴	∴	...	∴	∴	∴
2	1	65	185	...	100	39.7	83
3	2	43	130	...	98	42.7	71
3	1	35	80	...	40	32.3	86.8
3	2	67	130	...	100	36.7	69.3
3	2	46	110	...	70	41.7	43.3
3	2	71	110	...	70	46.2	80
∴	∴	∴	∴	...	∴	∴	∴
3	2	43	130	...	80	37.8	36
4	2	78	110	...	60	37.4	79
4	2	70	100	...	70	48.7	68
4	1	72	120	...	60	53.1	86
4	1	55	130	...	90	44.7	54
4	2	64	115	...	80	39.7	133
∴	∴	∴	∴	...	∴	∴	∴
4	1	75	80	...	50	45.5	66
5	2	74	100	...	60	28.7	105.7
5	2	73	150	...	80	44.7	59
5	2	73	115	...	70	51.9	126.7
5	2	73	140	...	80	27.4	106.6
5	2	85	90	...	50	35.4	88
∴	∴	∴	∴	...	∴	∴	∴
5	2	69	120	...	80	33.7	101

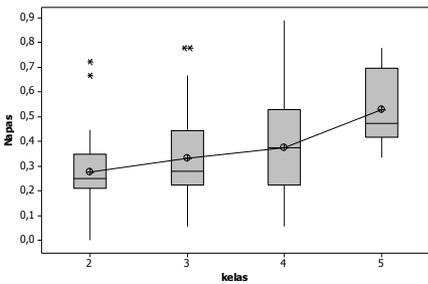
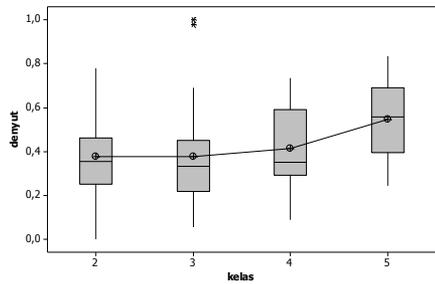
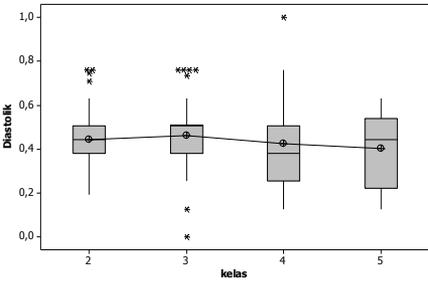
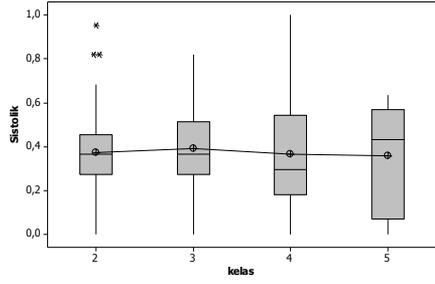
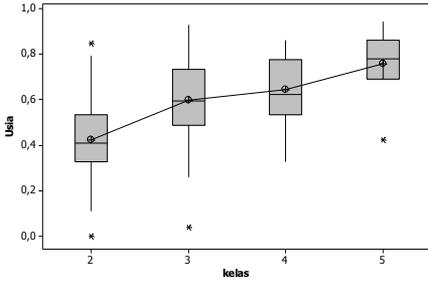
Lampiran 2. *Boxplot* Data Training pada Seleksi *Backward*, *Forward*, dan *Stepwise*



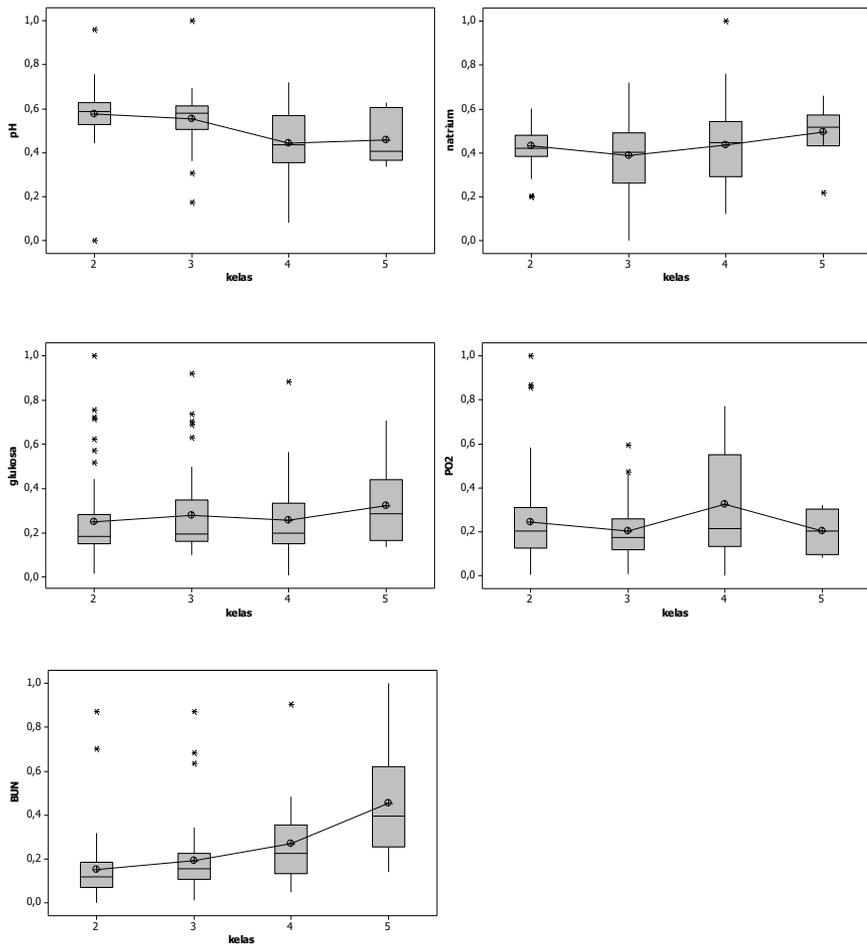
Lampiran 2. *Boxplot* Data Training pada Seleksi *Backward*, *Forward*, dan *Stepwise* (Lanjutan)



Lampiran 3. *Boxplot* Data Training pada Seleksi Algoritma Genetika



Lampiran 3. *Boxplot* Data Training pada Seleksi Algoritma Genetika (Lanjutan)



Lampiran 4. Tahapan Seleksi Variabel Metode *Backward*

	<i>df</i>	AIC
<i>Start: AIC=227.89</i>		
$Y = f(X_1, X_2, \dots, X_{13})$		
$-X_7$	1	225.92
$-X_{12}$	1	226.20
\vdots	\vdots	\vdots
$-X_2$	1	284.49
<i>Step 1: AIC=225.92</i>		
$Y = f(X_1, X_2, \dots, X_{13})$		
$-X_{12}$	1	224.24
$-X_4$	1	224.36
\vdots	\vdots	\vdots
$-X_2$	1	282.49
\vdots	\vdots	\vdots
<i>Step 4: AIC=221.03</i>		
$Y = f(X_1, X_2, \dots, X_{13})$		
<i>none</i>		221.03
$-X_{11}$	1	223.38
$-X_{10}$	1	224.18
\vdots	\vdots	\vdots
$-X_2$	1	279.51

Lampiran 5 Tahapan Seleksi Variabel Metode *Forward*

	<i>df</i>	AIC
<i>Start: AIC=362.12</i>		
$Y = 1$		
+ X_2	1	313.40
+ X_8	1	335.15
⋮	⋮	⋮
+ X_{11}	1	364.11
<i>Step 1: AIC=313.4</i>		
$Y = f(X_2)$		
+ X_8	1	294.88
+ X_6	1	296.37
⋮	⋮	⋮
+ X_4	1	314.85
⋮	⋮	⋮
<i>Step 9: AIC=221.03</i>		
$Y = f(X_1, X_2, \dots, X_{13})$		
<i>none</i>		221.03
+ X_{12}	1	222.67
+ X_3	1	222.72
+ X_7	1	222.86
+ X_4	1	223.03

Lampiran 6 Tahapan Seleksi Variabel Metode *Stepwise*

	<i>df</i>	AIC
<i>Start: AIC=362.12</i>		
$Y = 1$		
+ X_2	1	313.40
+ X_8	1	335.15
⋮	⋮	⋮
+ X_{11}	1	364.11
<i>Step 1: AIC=313.4</i>		
$Y = f(X_2)$		
+ X_8	1	294.88
+ X_6	1	296.37
⋮	⋮	⋮
- X_2	1	362.12
⋮	⋮	⋮
<i>Step 10: AIC=221.03</i>		
$Y = f(X_1, X_2, \dots, X_{13})$		
<i>none</i>		221.03
+ X_{12}	1	222.67
+ X_3	1	222.72
⋮	⋮	⋮
- X_2	1	279.51

Lampiran 7. *Syntax Cross Validation dan Regresi Logistik Ordinal*

```

library(MXM)
library(foreign)
library(ggplot2)
library(MASS)
library(Hmisc)
library(reshape2)
library(caret)
library(rminer)
data=read.csv("e:/datapneumoniastd.csv",header=TRUE,sep=";")
data[,1]=as.factor(data[,1])
data[,2]=as.factor(data[,2])
a=generatefolds(data$kelas,nfolds=5,stratified=TRUE,seed=12345)
AUC=rep(0,5)
g=rep(0,5)
for(i in 1:5)
{
  train=data[-a[[i]],]
  test=data[a[[i]],]
  model = polr(kelas~., data=train , Hess = TRUE)
  pred=predict(model,test)
  tabel1=table(test$kelas,pred)
  AUC[i]=0.25*(((tabel1[1,1]/sum(tabel1[1,]))+(tabel1[2,2]/sum(tabel1[2,]))
)+tabel1[3,3]/sum(tabel1[3,])+tabel1[4,4]/sum(tabel1[4,]))
  g[i]=(((tabel1[1,1]/sum(tabel1[1,]))*(tabel1[2,2]/sum(tabel1[2,]))*tabel1[3,3]
)/sum(tabel1[3,])*tabel1[4,4]/sum(tabel1[4,]))^(0.25)
}
summary(model)
mean(AUC)
mean(g)

```

Lampiran 8. *Syntax Seleksi Backward*

```

library(MXM)
library(foreign)
library(ggplot2)
library(MASS)
library(Hmisc)
library(reshape2)
library(caret)
library(rminer)
library(stepwise)
AUC=rep(0,5)
g=rep(0,5)
a=generatefolds(data$kelas,nfolds=5,stratified=TRUE,seed=12345)
for(i in 1:5)
{
  train=data[-a[[i]],]
  test=data[a[[i]],]
  fullmodel = polr(kelas~., data=train , Hess = TRUE)
  model=step(fullmodel)
  pred=predict(model,test)
  tabel1=table(test$kelas,pred)

  AUC[i]=0.25*((tabel1[1,1]/sum(tabel1[1,]))+(tabel1[2,2]/sum(tabel1[2,])
)+tabel1[3,3]/sum(tabel1[3,])+tabel1[4,4]/sum(tabel1[4,]))

  g[i]=((tabel1[1,1]/sum(tabel1[1,]))*(tabel1[2,2]/sum(tabel1[2,]))*tabel1[3
,3]/sum(tabel1[3,])*tabel1[4,4]/sum(tabel1[4,]))^(0.25)
}
summary(model)
mean(AUC)
mean(g)

```

Lampiran 9. *Syntax Seleksi Forward*

```

library(MXM)
library(foreign)
library(ggplot2)
library(MASS)
library(Hmisc)
library(reshape2)
library(caret)
library(rminer)
library(stepwise)
AUC=rep(0,5)
g=rep(0,5)
a=generatefolds(data$kelas,nfolds=5,stratified=TRUE,seed=12345)
for(i in 1:5)
{
  train=data[-a[[i]],]
  test=data[a[[i]],]
  fullmodel = polr(kelas~., data=train , Hess = TRUE)
  nothing = polr(kelas~ 1,data=train , Hess = TRUE)
  model = step(nothing,
scope=list(lower=formula(nothing),upper=formula(fullmodel)),direction=
"forward")
  pred=predict(model,test)
  tabel1=table(test$kelas,pred)
  AUC[i]=0.25*((tabel1[1,1]/sum(tabel1[1,]))+(tabel1[2,2]/sum(tabel1[2,])
)+tabel1[3,3]/sum(tabel1[3,])+tabel1[4,4]/sum(tabel1[4,]))
  g[i]=((tabel1[1,1]/sum(tabel1[1,]))*(tabel1[2,2]/sum(tabel1[2,]))*tabel1[3
,3]/sum(tabel1[3,])*tabel1[4,4]/sum(tabel1[4,]))^(0.25)
}summary(model)
mean(AUC)
mean(g)

```

Lampiran 10. *Syntax Seleksi Stepwise*

```

library(MXM)
library(foreign)
library(ggplot2)
library(MASS)
library(Hmisc)
library(reshape2)
library(caret)
library(rminer)
library(stepwise)
AUC=rep(0,5)
g=rep(0,5)
a=generatefolds(data$kelas,nfolds=5,stratified=TRUE,seed=12345)
for(i in 1:5)
{
  train=data[-a[[i]],]
  test=data[a[[i]],]
  fullmodel = polr(kelas~., data=train , Hess = TRUE)
  nothing = polr(kelas~ 1,data=train , Hess = TRUE)
  model = step(nothing,
               scope=list(lower=formula(nothing),upper=formula(fullmodel)),
               direction="both")
  pred=predict(model,test)
  tabel1=table(test$kelas,pred)
  AUC[i]=0.25*((tabel1[1,1]/sum(tabel1[1,]))+(tabel1[2,2]/sum(tabel1[2,])
  )+tabel1[3,3]/sum(tabel1[3,])+tabel1[4,4]/sum(tabel1[4,]))
  g[i]=((tabel1[1,1]/sum(tabel1[1,]))*(tabel1[2,2]/sum(tabel1[2,]))*tabel1[3,3]/sum(tabel1[3,])*tabel1[4,4]/sum(tabel1[4,]))^(0.25)
}summary(model)
mean(AUC)
mean(g)

```

Lampiran 11. *Syntax* Seleksi Algoritma Genetika

```

library(MXM)
library(foreign)
library(ggplot2)
library(MASS)
library(Hmisc)
library(reshape2)
library(caret)
library(rminer)
library(car)
library(GA)
library(xlsx)
a=generatefolds(data$kelas,nfolds=5,stratified=TRUE,seed=12345)
train=data[-a[[4]],]
test=data[a[[4]],]
train[,2]=as.factor(train[,2])
test[,2]=as.factor(test[,2])
xtrain=train[,2:14]
ytrain=train[,1]
fitness <- function(string)
{
  inc <- which(string == 1)
  xtrain <- xtrain[,inc]
  b=cbind(ytrain,xtrain)
  model = polr(as.factor(ytrain)~., data=b, Hess = TRUE)
  pred=predict(model,xtrain)
  tabel1=table(ytrain,pred)
  AUC=0.25*((tabel1[1,1]/sum(tabel1[1,]))+(tabel1[2,2]/sum(tabel1[2,]))+
  tabel1[3,3]/sum(tabel1[3,])+tabel1[4,4]/sum(tabel1[4,])) return(AUCw)
}
gaControl("binary"=list(selection="ga_rwSelection"))
GA <- ga("binary", fitness = fitness, nBits = ncol(xtrain),
  names = colnames(xtrain),
  selection=gaControl("binary")$selection,popSize=100, pcrossover=0.8,
  pmutation=0.1, elitism=base::max(1,round(100*0.2)), monitor =
  plot,maxiter = 60)

```

Lampiran 6. *Syntax* Seleksi Algoritma Genetika (Lanjutan)

```

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

train=data[-a[[4]],]
test=data[a[[4]],]
xtest=test[,2:14]
xt1=xtrain[,f2]
xt2=xtest[,f2]
xt1[,1]=as.factor(xt1[,1])
xt2[,1]=as.factor(xt2[,1])
write.xlsx(xt1,'e:/datatrainerpilihG4auc.xlsx',sheetName = 'Sheet1')
write.xlsx(xt2,'e:/datatestterpilihGA4auc.xlsx',sheetName = 'Sheet1')
b=cbind(ytrain,xt1)
model = polr(as.factor(ytrain)~., data=b , Hess = TRUE)
pred=predict(model,xt2)
tabel1=table(test$kelas, pred)
AUC=0.25*((tabel1[1,1]/sum(tabel1[1,]))+(tabel1[2,2]/sum(tabel1[2,]))+t
abel1[3,3]/sum(tabel1[3,])+tabel1[4,4]/sum(tabel1[4,]))
model
AUC

```

Lampiran 12. *Syntax* Optimasi Algoritma Genetika

```

function[fitness]=optimasi(B)
%DATA
persistent dblArray;

if isempty(dblArray)
    disp 'Reading array from spreadsheet ...'
    dblArray=xlsread('e:/ts3.xlsx');
end
x1=dblArray(:,1);
x2=dblArray(:,2);
x3=dblArray(:,3);
x4=dblArray(:,4);
x5=dblArray(:,5);
x6=dblArray(:,6);
x7=dblArray(:,7);
x8=dblArray(:,8);
x9=dblArray(:,9);
x10=dblArray(:,10);
x11=dblArray(:,11);
x12=dblArray(:,12);
x13=dblArray(:,13);
y=dblArray(:,14);
x=[x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12 x13];
%- Membuat model
gy=B(4)*x(:,1)+B(5)*x(:,2)+B(6)*x(:,3)+B(7)*x(:,4)+B(8)*x(:,5)+B(9)*
x(:,6)+B(10)*x(:,7)+B(11)*x(:,8)+B(12)*x(:,9)+B(13)*x(:,10)+B(14)*x(
,11)+B(15)*x(:,12)+B(16)*x(:,13);
p2=1./(1-exp(gy-B(1)));
p3=(1./(1-exp(gy-B(2))))-p2;
p4=(1./(1-exp(gy-B(3))))-p3;
p5=1-p2-p3-p4;
%- Mencari y topi
pgabung=[p2 p3 p4 p5];
n=length(y);

```

Lampiran 7. *Syntax* Optimasi Algoritma Genetika

```
for i=1:n
    pmax=max(pgabung(i,:));
    if pmax==pgabung(i,1)
        ytopi(i)=2;
    else if pmax==pgabung(i,2)
        ytopi(i)=3;
    else if pmax==pgabung(i,3)
        ytopi(i)=4;
    else
        ytopi(i)=5;
    end;
end;
end;
end;
ytopi=ytopi;
tabel=confusionmat(y,ytopi);
auc=0.25*(tabel(1,1)/sum(tabel(1,:))+tabel(2,2)/sum(tabel(2,:))+tabel(3,3)
/sum(tabel(3,:))+tabel(4,4)/sum(tabel(4,:)));
fitness=1-auc;
end
```

Lampiran 13. *Syntax Memanggil Function Optimasi Algoritma Genetika*

```
n=length(B);
option = gaoptimset ('PopulationSize', 100, 'InitialPopulation',
B, 'MutationFcn', { @mutationuniform,0.1 }, 'CrossoverFraction', [0.8],
'CrossoverFcn', @crossoveringlepoint, 'EliteCount', 20, 'SelectionFcn',
@selectionroulette)
[a,fval,exitflag,output,population,scores]=ga(@optimasi,n,[],[],[],[],[],[]
,option)

xlswrite('fitnessfullmodAUC1.xlsx',[population scores])
```

Lampiran 14. *Syntax Naïve Bayes Classification*

```

library(MXM)
library(e1071)
library(caret)
library(rminer)
a=generatefolds(data$kelas,nfolds=5,stratified=TRUE,seed=12345)
AUC=rep(0,5)
g=rep(0,5)
for(i in 1:5)
{
  train=data[-a[[i]],]
  test=data[a[[i]],]
  model = naiveBayes(kelas~., data=train)
  pred=predict(model,test)
  tabel1=table(test$kelas,pred)
  AUC[i]=0.25*((tabel1[1,1]/sum(tabel1[1,]))+(tabel1[2,2]/sum(tabel1[2,])
)+tabel1[3,3]/sum(tabel1[3,])+tabel1[4,4]/sum(tabel1[4,]))
  g[i]=((tabel1[1,1]/sum(tabel1[1,]))*(tabel1[2,2]/sum(tabel1[2,]))*tabel1[3
,3]/sum(tabel1[3,])*tabel1[4,4]/sum(tabel1[4,]))^(0.25)
}
mean(AUC)
mean(g)

```

Lampiran 15. Surat Pernyataan Pengambilan Data

SURAT PERNYATAAN

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

Nama : Ratna Juwita

NRP : 062114 4000 0091

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



(Ratna Juwita)
NRP. 062114 4000 0091

*(coret yang tidak perlu)

(Halaman ini sengaja dikosongkan)

BIODATA PENULIS



Ratna Juwita atau yang akrab disapa Ratna atau Raju merupakan anak sulung dari dua bersaudara yang lahir di Tulungagung, 23 Mei 1996. Putri dari pasangan Bapak Sumali dan Ibu Komasri ini berdomisili di Tulungagung dan telah menempuh pendidikan normal di SDN II Suruhan Kidul (2002-2008), SMPN 1 Bandung (2008-2011), dan SMAN 1 Kauman (2011-2014). Penulis memilih untuk melanjutkan studi guna menempuh gelar sarjana di Institut Teknologi Sepuluh Nopember (ITS) Surabaya. Pada tahun 2014, penulis dinyatakan lolos SBMPTN sebagai mahasiswa Departemen Statistika, Fakultas Matematika, Komputasi, dan Sains Data. Semasa kuliah penulis aktif di kepanitiaan dan organisasi kemahasiswaan ITS tingkat jurusan yakni Himpunan Mahasiswa Statistika ITS (HIMASTA-ITS) pada periode 2015-2016 sebagai *staff* Departemen Sosial Masyarakat dan pada periode 2016-2017 sebagai ketua biro Sosial Masyarakat. Selain itu penulis turut berpartisipasi dalam kepanitiaan seperti Pekan Raya Statistika, GERIGI dan ITS Megajar for Indonesia. Penulis juga aktif di organisasi luar kampus sebagai *volunteer* bidang HRD di Gerakan Melukis Harapan (GMH) yaitu sebuah yayasan yang melakukan pemberdayaan masyarakat di eks-lokalisasi Dolly. Penulis juga pernah berpartisipasi sebagai mentor bersama dengan dosen Statistika ITS dalam Pelatihan Pembelajaran Eksak Berbasis TIK di SMA Muhammadiyah 1 Gresik. Pada bulan Juli-Agustus 2017 penulis berkesempatan untuk melakukan kerja praktik di BMKG Jakarta. Segala kritik dan saran serta diskusi lebih lanjut mengenai Tugas Akhir ini dapat dikirimkan melalui surat elektronik (*e-mail*) ke juwitaratna767@gmail.com.

(Halaman ini sengaja dikosongkan)