



TUGAS AKHIR - KS184822

**PREDIKSI *FINANCIAL DISTRESS* PADA  
PERUSAHAAN SEKTOR INDUSTRI DI INDONESIA  
*METODE SUPPORT VECTOR MACHINE,*  
*GENERALIZED EXTREME VALUE REGRESSION,*  
DAN *LOGISTIC REGRESSION* DENGAN SELEKSI  
VARIABEL**

RIZKI NANDA SAVERA  
NRP 062116 4000 0084

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

PROGRAM STUDI SARJANA  
DEPARTEMEN STATISTIKA  
FAKULTAS SAINS DAN ANALITIKA DATA  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2020





TUGAS AKHIR - KS184822

**PREDIKSI *FINANCIAL DISTRESS* PADA  
PERUSAHAAN SEKTOR INDUSTRI DI  
INDONESIA METODE *SUPPORT VECTOR  
MACHINE*, *GENERALIZED EXTREME VALUE  
REGRESSION*, DAN *LOGISTIC REGRESSION*  
DENGAN SELEKSI VARIABEL**

RIZKI NANDA SAVERA  
NRP 062116 4000 0084

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

PROGRAM STUDI SARJANA  
DEPARTEMEN STATISTIKA  
FAKULTAS SAINS DAN ANALITIKA DATA  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2020

*(Halaman ini sengaja dikosongkan)*



**FINAL PROJECT - KS184822**

**FINANCIAL DISTRESS PREDICTION IN INDUSTRIAL  
SECTOR COMPANIES IN INDONESIA USING  
SUPPORT VECTOR MACHINE, GENERALIZED  
EXTREME VALUE REGRESSION, AND LOGISTIC  
REGRESSION WITH FEATURE SELECTION  
METHODS**

**RIZKI NANDA SAVERA  
SN 062116 4000 0084**

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

**UNDERGRADUATE PROGRAMME  
DEPARTMENT OF STATISTICS  
FACULTY OF SCIENCE AND DATA ANALYTICS  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2020**

*(Halaman ini sengaja dikosongkan)*

## LEMBAR PENGESAHAN

# PREDIKSI *FINANCIAL DISTRESS* PADA PERUSAHAAN SEKTOR INDUSTRI DI INDONESIA *METODE SUPPORT VECTOR MACHINE,* *GENERALIZED EXTREME VALUE REGRESSION,* DAN *LOGISTIC REGRESSION* DENGAN SELEKSI VARIABEL

## TUGAS AKHIR

Diajukan untuk Memenuhi Salah Satu Syarat

Memperoleh Gelar Sarjana Statistika

pada

Program Studi Sarjana Departemen Statistika

Fakultas Sains dan Analitika Data

Institut Teknologi Sepuluh Nopember

Oleh:

**Rizki Nanda Savera**

NRP. 062116 4000 0084

Disetujui oleh Pembimbing:

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

NIP. 19831204 200812 1 002

Mengetahui,

Kepala Departemen Statistika



**Dr. Dra. Kartika Fithriasari, M.Si**  
NIP. 19691212 199303 2 002

SURABAYA, Agustus 2020

*(Halaman ini sengaja dikosongkan)*

**PREDIKSI FINANCIAL DISTRESS PADA  
PERUSAHAAN SEKTOR INDUSTRI DI INDONESIA  
METODE SUPPORT VECTOR MACHINE,  
GENERALIZED EXTREME VALUE REGRESSION,  
DAN LOGISTIC REGRESSION DENGAN SELEKSI  
VARIABEL**

**Nama Mahasiswa : Rizki Nanda Savera**

**NRP : 062116 4000 0084**

**Departemen : Statistika-FSAD-ITS**

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

**Abstrak**

*Sektor industri dalam perekonomian Indonesia memiliki andil sebagai penyumbang kontribusi terbesar untuk perekonomian di Indonesia. Eksistensi dari sektor industri menuntut kestabilan keuangan suatu perusahaan yang berkecimpung dibidang industri. Maka dari itu pada penelitian ini akan dianalisis perusahaan-perusahaan sektor industri yang dapat dikatakan mengalami financial distress berdasarkan rasio aktivitas, rasio profitabilitas, rasio solvabilitas, dan rasio likuiditas. Metode yang digunakan dalam penelitian antara lain yaitu Support Vector Machine, Generalized Extreme Value Regression, dan Logistic Regression. Selain itu pada masing-masing metode juga dilakukan seleksi variabel untuk meminimalisir adanya multikolinieritas. Analisis dilakukan menggunakan time window dengan pendekatan full memory time window. Setelah dilakukan permodelan secara serentak dan seleksi variabel diketahui bahwa untuk permodelan secara serentak dan seleksi variabel, metode Generalized Extreme Value Regression dan dengan metode yang sama menggunakan backward elimination memberikan hasil prediksi terbaik dan diberikan oleh data size 0 baik pada data train ataupun data test dengan ukuran AUC.*

**Kata Kunci:** *Financial Distress, Industri, Klasifikasi, Probabilitas, Seleksi Variabel.*

*(Halaman ini sengaja dikosongkan)*

# **FINANCIAL DISTRESS PREDICTION IN INDUSTRIAL SECTOR COMPANIES IN INDONESIA USING SUPPORT VECTOR MACHINE, GENERALIZED EXTREME VALUE REGRESSION, AND LOGISTIC REGRESSION WITH FEATURE SELECTION METHODS**

**Name : Rizki Nanda Savera  
Student Number : 062116 4000 0084  
Department : Statistics  
Supervisor : Dr.rer.pol. Dedy Dwi Prastyo, S.Si., M.Si.**

## ***Abstract***

*The industrial sector in the Indonesian economy has contributed as the largest contributor. The existence of the industrial sector demands the financial stability of a company working in the industry. Therefore this research will analyze industrial sector companies which can be said to experience financial distress based on their activity ratio, profitability ratio, solvency ratio, and liquidity ratio. The methods used in this research include Support Vector Machine, Generalized Extreme Value Regression, and Logistic Regression. In addition, each variable is also selected for variables to minimize the presence of multicollinearity. The analysis was performed using a time window with a full memory time window approach. After modeling simultaneously and with variable selection it is known that for modeling simultaneously, the Generalized Extreme Value Regression method provides the best prediction results and is given by data size 0 both on the data train or data test with AUC size. While based on the AUC obtained from modeling with variable selection, the best prediction results are given by data size 0 with the Backward Generalized Extreme Value Regression method both on the data train or the data test.*

***Keywords:*** *Clasification, Feature Selection, Financial Distress, Industry, Probability.*

*(Halaman ini sengaja dikosongkan)*

## KATA PENGANTAR

Puji Syukur penulis panjatkan atas berkat, rahmat, dan karunia yang telah diberikan Allah SWT sehingga penulis dapat menyelesaikan Laporan Tugas Akhir ini yang berjudul “**Prediksi Financial Distress pada Perusahaan Sektor Industri Di Indonesia Metode Support Vector Machine, Generalized Extreme Value Regression, dan Logistic Regression dengan Seleksi Variabel**” dengan tepat waktu.

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

1. Orang tua dan keluarga penulis yang mendukung dan memberikan doa selama penyusunan Tugas Akhir.
2. Dr. rer. pol. Dedy Dwi Prastyo, S.Si., M.Si. selaku dosen pembimbing yang telah memberikan bimbingan, saran, serta motivasi selama penyusunan Tugas Akhir berlangsung.
3. Santi Puteri Rahayu, S.Si., M.Si., Ph.D dan Dr. Ir. Setiawan, M.S selaku dosen penguji yang telah memberikan masukan dan bantuan dalam menyelesaikan Tugas Akhir.
4. Dr. Suhartono, S.Si., M.Sc selaku dosen wali yang telah banyak memberikan saran dan arahan dalam proses belajar di Departemen Statistika.
5. Dr. Dra. Kartika Fithriasari, M.Si selaku Kepala Departemen Statistika FSAD-ITS yang telah membantu memfasilitasi penulis dalam masa perkuliahan.
6. Al'awwa, Rivi, Kicky, Jemima, Widya, Adinda, Isye, Febri dan Dita yang selalu memberikan dukungan dan semangat dalam penyusunan Tugas Akhir.
7. Teman-teman seperjuangan TR16GER khususnya Thalia yang membantu penulis dalam running data dan sebagai teman diskusi serta teman-teman TR16GER lainnya yang selalu memberikan semangat kepada penulis dalam penyusunan Tugas Akhir.
8. Teman-teman UKM Sepakbola ITS khususnya divisi Futsal Putri dan tim serta staf pelatih Futsal Putri ITS yang selalu

memberi dukungan dalam masa perkuliahan dan dalam penyusunan Tugas Akhir.

9. Seluruh pihak yang turut membantu dalam penyelesaian laporan Tugas Akhir ini baik secara langsung maupun tidak langsung.

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

Surabaya, Agustus 2020

Penulis

## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	i
<b>LEMBAR PENGESAHAN .....</b>	Error! Bookmark not defined.
<b>KATA PENGANTAR.....</b>	xi
<b>DAFTAR ISI.....</b>	xiii
<b>DAFTAR TABEL .....</b>	xvii
<b>DAFTAR LAMPIRAN .....</b>	xxi
<b>BAB I PENDAHULUAN .....</b>	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	7
1.3 Tujuan .....	8
1.4 Manfaat Penelitian .....	9
1.5 Batasan Masalah .....	9
<b>BAB II TINJAUAN PUSTAKA .....</b>	11
2.1 Statistika Deskriptif.....	11
2.2 Imputasi <i>k</i> -Nearest Neighbor.....	12
2.3 <i>Support Vector Machine</i> .....	12
2.3.1 <i>Support Vector Machine Linear Classifier</i> .....	13
2.3.2 <i>Support Vector Machine Nonlinear Classifier</i> .....	17
2.4 <i>Regularized Support Vector Machine</i> .....	19
2.4.1 <i>Lasso Absolute Shrinkage and Selection Support Vector Machine</i> .....	19
2.4.2 <i>Elastic-Net Support Vector Machine</i> .....	20
2.5 <i>Isotonic Regression</i> .....	20
2.6 <i>Extreme Value Theory</i> .....	21
2.7 <i>Generalized Extreme Value Distribution</i> .....	22
2.8 <i>Generalized Extreme Value Regression</i> .....	23
2.9 Estimasi Parameter <i>Generalized Extreme Value Regression</i> .....	23
2.10 Regresi Logistik .....	25
2.11 Estimasi Parameter <i>Logistic Regression</i> .....	26
2.12 <i>Regularized Logistic Regression</i> .....	28
2.12.1 <i>Lasso Absolute Shrinkage and Selection Logistic Regression</i> .....	29
2.12.2 <i>Elastic-Net Logistic Regression</i> .....	29
2.13 <i>Stepwise Procedures</i> .....	30

2.14	Uji Signifikansi Parameter .....	30
2.15	Evaluasi Ketepatan Klasifikasi .....	31
2.16	Konsep Drift .....	33
2.17	Rasio Keuangan .....	34
2.18	<i>Financial Distress</i> .....	36
2.19	Indikator Ekonomi Makro .....	37
<b>BAB III METODOLOGI PENELITIAN</b>	.....	<b>39</b>
3.1	Sumber Data .....	39
3.4	Langkah Penelitian.....	45
<b>BAB IV ANALISIS DAN PEMBAHASAN</b>	.....	<b>51</b>
4.1	Karakteristik Data .....	51
4.2	Klasifikasi Menggunakan <i>Support Vector Machines</i> (SVM).....	74
4.2.2	<i>Lasso Support Vector Machine</i> .....	78
4.2.3	<i>Elastic-Net Support Vector Machine</i> .....	80
4.3	Klasifikasi Menggunakan <i>Generalized Extreme Value Regression</i> .....	82
4.3.1	<i>Generalized Extreme Value Regression</i> secara Serentak.....	83
4.3.2	<i>Generalized Extreme Value Regression</i> dengan <i>Backward Elimination</i> .....	86
4.4	Klasifikasi Menggunakan <i>Logistic Regression</i> .....	89
4.4.1	<i>Binary Logistic Regression</i> secara Serentak.....	89
4.4.2	<i>Binary Logistic Regression</i> secara <i>Stepwise</i> .....	93
4.4.3	<i>Lasso Logistic Regression</i> .....	96
4.4.4	<i>Elastic-Net Logistic Regression</i> .....	98
4.5	Perbaikan Kebaikan Model .....	100
<b>BAB V KESIMPULAN DAN SARAN</b>	.....	<b>105</b>
5.1	Kesimpulan.....	105
5.2	Saran .....	106
<b>DAFTAR PUSTAKA</b>	.....	<b>108</b>
<b>LAMPIRAN</b>	.....	<b>117</b>
<b>BIODATA PENULIS</b>	.....	<b>157</b>

## DAFTAR GAMBAR

<b>Gambar 2.1</b> Klasifikasi Linier SVM (a) <i>linearly separable</i> dan (b) <i>linearly nonseparable</i> .....	13
<b>Gambar 2.2</b> Hyperplane Klasifikasi Linier SVM .....	14
<b>Gambar 2.3</b> Pemetaan Ruang Dua Dimensi Data Menjadi Tiga Dimensi .....	18
<b>Gambar 2.4</b> Probability Density Function dari Distribusi Fréchet ( <i>dotted</i> ), Distribusi Gumbel ( <i>solid</i> ), dan Distribusi Weibull ( <i>dashed</i> ) .....	21
<b>Gambar 2.5</b> a) Skema <i>Block Maxima</i> dan b) Skema <i>Peak Over Threshold</i> .....	22
<b>Gambar 2.6</b> Kurva ROC .....	33
<b>Gambar 2.7</b> (a) <i>Full Memory Time Window</i> ; (b) <i>No Memory Time Window</i> .....	34
<b>Gambar 3.1</b> (a) Ilustrasi <i>full memory time window size 0</i> .....	44
<b>Gambar 3.1</b> (b) Ilustrasi <i>full memory time window size 1</i> .....	45
<b>Gambar 3.1</b> (c) Ilustrasi <i>full memory time window size 2</i> .....	45
<b>Gambar 3.1</b> (d) Ilustrasi <i>full memory time window size 3</i> .....	45
<b>Gambar 3.2</b> Diagram Alir Penelitian .....	49
<b>Gambar 4.1</b> Density Plot EBITA .....	52
<b>Gambar 4.2</b> Density Plot STA .....	53
<b>Gambar 4.3</b> Density Plot ITR .....	55
<b>Gambar 4.4</b> Density Plot DSI .....	56
<b>Gambar 4.5</b> Density Plot ROE .....	58
<b>Gambar 4.6</b> Density Plot NPM .....	59
<b>Gambar 4.7</b> Density Plot OPM .....	61
<b>Gambar 4.8</b> Density Plot DER .....	62
<b>Gambar 4.9</b> Density Plot DAR .....	63
<b>Gambar 4.10</b> Density Plot CR .....	65
<b>Gambar 4.11</b> Density Plot QR .....	66
<b>Gambar 4.12</b> Density Plot ETD .....	67
<b>Gambar 4.13</b> Density Plot WCA .....	69
<b>Gambar 4.14</b> Density Plot WCS .....	70
<b>Gambar 4.15</b> Density Plot RGDPG .....	72
<b>Gambar 4.16</b> Density Plot BI7DRR .....	72
<b>Gambar 4.17</b> Density Plot USD/IDR .....	73

<b>Gambar 4.18</b> Density Plot Inflasi .....	74
<b>Gambar 4.19a</b> SVM Linier <i>Size 0</i> .....	77
<b>Gambar 4.19b</b> SVM Linier <i>Size 1</i> .....	77
<b>Gambar 4.19c</b> SVM Linier <i>Size 2</i> .....	77
<b>Gambar 4.19d</b> SVM Linier <i>Size 3</i> .....	77
<b>Gambar 4.20a</b> Lasso SVM <i>Size 0</i> .....	79
<b>Gambar 4.20b</b> Lasso SVM <i>Size 1</i> .....	79
<b>Gambar 4.20c</b> Lasso SVM <i>Size 2</i> .....	80
<b>Gambar 4.20d</b> Lasso SVM <i>Size 3</i> .....	80
<b>Gambar 4.21a</b> Elastic-Net SVM <i>Size 0</i> .....	82
<b>Gambar 4.21b</b> Elastic-Net SVM <i>Size 1</i> .....	82
<b>Gambar 4.21c</b> Elastic-Net SVM <i>Size 2</i> .....	82
<b>Gambar 4.21d</b> Elastic-Net SVM <i>Size 3</i> .....	82
<b>Gambar 4.22a</b> AUC Model Serentak .....	101
<b>Gambar 4.22b</b> Accuracy Model Serentak .....	101
<b>Gambar 4.23a</b> AUC Model dengan Seleksi Variabel .....	104
<b>Gambar 4.23a</b> Accuracy Model dengan Seleksi Variabel .....	104

## DAFTAR TABEL

<b>Tabel 2.1</b> Confussion Matrix .....	32
<b>Tabel 2.2</b> Kategori Kebaikan Model Berdasarkan AUC .....	33
<b>Tabel 3.1</b> Variabel Penelitian .....	39
<b>Tabel 3.2</b> Struktur Data Penelitian .....	43
<b>Tabel 3.3</b> Time Window Size Data Penelitian .....	44
<b>Tabel 4.1</b> Statistika Deskriptif EBITA .....	51
<b>Tabel 4.2</b> Statistika Deskriptif STA .....	53
<b>Tabel 4.3</b> Statistika Deskriptif ITR .....	54
<b>Tabel 4.4</b> Statistika Deskriptif DSI .....	55
<b>Tabel 4.5</b> Statistika Deskriptif ROE .....	57
<b>Tabel 4.6</b> Statistika Deskriptif NPM .....	58
<b>Tabel 4.7</b> Statistika Deskriptif OPM .....	60
<b>Tabel 4.8</b> Statistika Deskriptif DER .....	61
<b>Tabel 4.9</b> Statistika Deskriptif DAR .....	63
<b>Tabel 4.10</b> Statistika Deskriptif CR .....	64
<b>Tabel 4.11</b> Statistika Deskriptif QR .....	65
<b>Tabel 4.12</b> Statistika Deskriptif ETD .....	67
<b>Tabel 4.13</b> Statistika Deskriptif WCA .....	68
<b>Tabel 4.14</b> Statistika Deskriptif WCS .....	69
<b>Tabel 4.15a</b> Confussion Matrix Model SVM Linier Size 0 .....	75
<b>Tabel 4.15b</b> Confussion Matrix Model SVM Linier Size 1 .....	75
<b>Tabel 4.15c</b> Confussion Matrix Model SVM Linier Size 2 .....	75
<b>Tabel 4.15d</b> Confussion Matrix Model SVM Linier Size 3 .....	75
<b>Tabel 4.16</b> Ketepatan Klasifikasi Model SVM Linier .....	76
<b>Tabel 4.17a</b> Confussion Matrix Model Lasso SVM Size 0 .....	78
<b>Tabel 4.17b</b> Confussion Matrix Model Lasso SVM Size 1 .....	78
<b>Tabel 4.17c</b> Confussion Matrix Model Lasso SVM Size 2 .....	78
<b>Tabel 4.17d</b> Confussion Matrix Model Lasso SVM Size 3 .....	79
<b>Tabel 4.18</b> Ketepatan Klasifikasi Model Lasso SVM .....	79
<b>Tabel 4.19a</b> Confussion Matrix Model Elastic-Net SVM Size 0 .....	80
<b>Tabel 4.19b</b> Confussion Matrix Model Elastic-Net SVM Size 1 .....	81
<b>Tabel 4.19c</b> Confussion Matrix Model Elastic-Net SVM Size 2 .....	81

<b>Tabel 4.19d</b> Confussion Matrix Model <i>Elastic-Net SVM Size</i>	81
3.....	81
<b>Tabel 4.20</b> Ketepatan Klasifikasi Model <i>Elastic-Net SVM</i> .....	81
<b>Tabel 4.21</b> Model Prediksi <i>Financial Distress</i> GEVR secara Serentak.....	83
<b>Tabel 4.22</b> Variabel Signifikan Model Serentak GEVR.....	84
<b>Tabel 4.23a</b> Confussion Matrix Model Serentak GEVR <i>Size</i>	84
0.....	84
<b>Tabel 4.23b</b> Confussion Matrix Model Serentak GEVR <i>Size</i>	85
1.....	85
<b>Tabel 4.23c</b> Confussion Matrix Model Serentak GEVR <i>Size</i>	85
2.....	85
<b>Tabel 4.23d</b> Confussion Matrix Model Serentak GEVR <i>Size</i>	85
3.....	85
<b>Tabel 4.24</b> Ketepatan Klasifikasi Model Serentak GEVR.....	85
<b>Tabel 4.25</b> Variabel Signifikan Model <i>Backward GEVR</i> .....	87
<b>Tabel 4.26</b> Model Prediksi <i>Financial Distress</i> GEVR dengan <i>Backward Elimination</i> .....	87
<b>Tabel 4.27a</b> Confussion Matrix Model <i>Backward GEVR Size</i>	88
0.....	88
<b>Tabel 4.27b</b> Confussion Matrix Model <i>Backward GEVR Size</i>	88
1.....	88
<b>Tabel 4.27c</b> Confussion Matrix Model <i>Backward GEVR Size</i>	88
2.....	88
<b>Tabel 4.27c</b> Confussion Matrix Model <i>Backward GEVR Size</i>	88
3.....	88
<b>Tabel 4.28</b> Ketepatan Klasifikasi Model <i>Backward GEVR</i> .....	89
<b>Tabel 4.29</b> Model Prediksi <i>Financial Distress</i> LOGIT secara Serentak.....	90
<b>Tabel 4.30</b> Variabel Signifikan Model Serentak LOGIT.....	90
<b>Tabel 4.31a</b> Confussion Matrix Model Serentak LOGIT <i>Size</i>	91
0.....	91
<b>Tabel 4.31b</b> Confussion Matrix Model Serentak LOGIT <i>Size</i>	91
1.....	91
<b>Tabel 4.31c</b> Confussion Matrix Model Serentak LOGIT <i>Size</i>	91
2.....	91

<b>Tabel 4.31d</b> Confussion Matrix Model Serentak LOGIT Size 3 .....	92
<b>Tabel 4.32</b> Ketepatan Klasifikasi Model Serentak LOGIT .....	92
<b>Tabel 4.33</b> Variabel Signifikan Model Stepwise LOGIT .....	93
<b>Tabel 4.34</b> Model Prediksi <i>Financial Distress</i> LOGIT secara Stepwise .....	93
<b>Tabel 4.35a</b> Confussion Matrix Model Stepwise LOGIT Size 0 .....	94
<b>Tabel 4.35b</b> Confussion Matrix Model Stepwise LOGIT Size 1 .....	94
<b>Tabel 4.35c</b> Confussion Matrix Model Stepwise LOGIT Size 2 .....	94
<b>Tabel 4.35d</b> Confussion Matrix Model Stepwise LOGIT Size 3 .....	95
<b>Tabel 4.36</b> Ketepatan Klasifikasi Model Stepwise LOGIT .....	95
<b>Tabel 4.37</b> Variabel Model Lasso LOGIT .....	96
<b>Tabel 4.38a</b> Confussion Matrix Model Lasso LOGIT Size 0 .....	96
<b>Tabel 4.38b</b> Confussion Matrix Model Lasso LOGIT Size 1 .....	96
<b>Tabel 4.38c</b> Confussion Matrix Model Lasso LOGIT Size 2 .....	97
<b>Tabel 4.38d</b> Confussion Matrix Model Lasso LOGIT Size 3 .....	97
<b>Tabel 4.39</b> Ketepatan Klasifikasi Model Lasso LOGIT .....	97
<b>Tabel 4.40</b> Variabel Model Elastic-Net LOGIT .....	98
<b>Tabel 4.41a</b> Confussion Matrix Model Elastic-Net LOGIT Size 0 .....	98
<b>Tabel 4.41b</b> Confussion Matrix Model Elastic-Net LOGIT Size 1 .....	98
<b>Tabel 4.41c</b> Confussion Matrix Model Elastic-Net LOGIT Size 2 .....	99
<b>Tabel 4.41d</b> Confussion Matrix Model Elastic-Net LOGIT Size 3 .....	99
<b>Tabel 4.42</b> Ketepatan Klasifikasi Model Elastic-Net LOGIT .....	99
<b>Tabel 4.43</b> Perbandingan Akurasi Model Serentak .....	100
<b>Tabel 4.44</b> Perbandingan Akurasi Model dengan Seleksi Variabel .....	102

*(Halaman ini sengaja dikosongkan)*

## DAFTAR LAMPIRAN

<b>Lampiran 1.</b> Data Penelitian.....	117
<b>Lampiran 2.</b> <i>Syntax Imputasi Data</i> .....	118
<b>Lampiran 3.</b> <i>Syntax Eksplorasi Data</i> .....	119
<b>Lampiran 4.</b> Syntax SVM Linier.....	120
<b>Lampiran 5.</b> <i>Syntax Lasso SVM</i> .....	121
<b>Lampiran 6.</b> <i>Syntax Elastic-Net SVM</i> .....	122
<b>Lampiran 7.</b> <i>Syntax GEVR</i> .....	124
<b>Lampiran 8.</b> <i>Syntax Regresi Logistik</i> .....	127
<b>Lampiran 9.</b> <i>Syntax Lasso Regresi Logistik</i> .....	129
<b>Lampiran 10.</b> <i>Syntax Elastic-Net Regresi Logistik</i> .....	130
<b>Lampiran 11.</b> ROC SVM.....	134
<b>Lampiran 12.</b> ROC GEVR.....	137
<b>Lampiran 13.</b> ROC LOGIT.....	139
<b>Lampiran 14.</b> Koefisien <i>Lasso SVM</i> .....	143
<b>Lampiran 15.</b> Koefisien <i>Elastic-Net SVM</i> .....	143
<b>Lampiran 16.</b> Nilai Log-Likelihood Parameter $\tau$ untuk GEVR.....	144
<b>Lampiran 17.</b> Estimasi Paramter Model dengan GEVR Serentak.....	144
<b>Lampiran 18.</b> Estimasi Paramter Model Akhir dengan <i>Back- ward GEVR</i> .....	147
<b>Lampiran 19.</b> Estimasi Paramter Model dengan LOGIT Se- rentak.....	149
<b>Lampiran 20.</b> Estimasi Paramter Model dengan <i>Stepwise LOGIT</i> .....	152
<b>Lampiran 21.</b> Estimasi Paramter Model dengan <i>Lasso LOGIT</i> .....	154
<b>Lampiran 22.</b> Estimasi Paramter Model dengan <i>Elastic-Net LOGIT</i> .....	154
<b>Lampiran 23.</b> Surat Keterangan Data.....	156

*(Halaman ini sengaja dikosongkan)*

## **BAB I**

### **PENDAHULUAN**

#### **1.1 Latar Belakang**

Manusia melakukan berbagai cara dalam upaya untuk memenuhi kebutuhannya menggunakan sumber daya yang ada baik sumber daya alam maupun sumber daya manusia itu sendiri. Sumber daya yang ada dapat dimanfaatkan langsung atau melalui proses pengolahan. Pengolahan yang dilakukan merupakan pengubahan bahan mentah yang didapat dari alam menjadi suatu barang atau produk yang dapat dimanfaatkan. Pengolahan dijalankan oleh suatu badan organisasi yang dapat dikatakan sebagai perusahaan. Beberapa tokoh memberikan definisi terhadap “perusahaan”, salah satunya adalah Molengraff yang dikutip oleh Abdulkadir (2000) mengemukakan bahwa perusahaan adalah semua perbuatan yang dilakukan dengan terus-menerus, bertindak keluar untuk mendapatkan penghasilan dengan cara memperniagakan barang-barang, menyerahkan barang-barang atau mengadakan perjanjian-perjanjian. Perusahaan secara garis besar digolongkan menjadi perusahaan jasa yang merupakan badan usaha yang dalam melakukan operasional kerja yaitu menjual pelayanan berupa jasa yang dimiliki oleh pribadi maupun kelompok, perusahaan dagang yaitu perusahaan yang melakukan penjualan barang yang telah dibeli dari suatu badan usaha yang lain dengan meningkatkan harga untuk mendapatkan laba, dan perusahaan manufaktur yaitu perusahaan yang sering disebut perusahaan industri karena kegiatan yang dilakukan adalah perusahaan mengolah bahan mentah menjadi barang atau produk yang siap dijual (Sujarweni, 2016).

Kuartal pertama tahun 2019 Badan Pusat Statistika menunjukkan bahwa sektor industri menjadi penyumbang paling besar terhadap perekonomian Indonesia yaitu sebesar 20,07% (Tobing, 2019). Pada Sindonews.com (2020) disebutkan bahwa pada 2019 industri merupakan sektor yang memiliki kontribusi terbesar dalam nilai ekspor nasional yaitu sepanjang Januari hingga Oktober 2019, ekspor produk hasil industri manufaktur mencapai 105,11 miliar USD. Agus Gumiwang selaku Menteri Perindustrian

menyatakan bahwa industri manufaktur juga sektor yang menyumbang kontribusi yang besar terhadap penerimaan pajak di mana hingga Triwulan III 2019 kontribusi industri manufaktur sebesar Rp. 245,60 triliun atau 29,23% dari penerimaan pajak neto nasional.

Krisis keuangan global pada tahun 2008-2009 memberi dampak yang signifikan terhadap perkembangan dunia bisnis. Kejadian serupa dapat terjadi lagi pada masa depan mengingat pasar modal yang menyadari bahwa minat masa depan mungkin akan lebih tinggi, sehingga dapat diamati ketidakpastian tersebut dan meningkatnya volatilitas. Situasi seperti itu menuntut perusahaan memperkuat fundamentalnya dalam menghadapi perubahan global yang terjadi. Perusahaan yang tidak mampu menghadapi dan mengikuti alur perubahan global lambat laun akan mengalami kesulitan keuangan (*financial distress*) yang berujung pada kebangkrutan.

Dalam Agusti (2013), Platt dan Platt mendefinisikan *financial distress* sebagai tahap penurunan kondisi keuangan secara terus menerus pada suatu perusahaan sebelum mengalami kebangkrutan. Perusahaan yang sukses diwajibkan memiliki manajemen keuangan yang efektif dan dapat meramalkan kemungkinan adanya *financial distress*. Pun dengan perusahaan yang bergerak pada sektor industri harus dapat memanajemen keuangan perusahaan agar tidak mengalami *financial distresses*. Sebagai penyumbang kontribusi terbesar terhadap perekonomian di Indonesia dan penyumbang jumlah ekspor terbesar, sektor industri tidak boleh melemah atau mengalami kebangkrutan. Apabila suatu perusahaan mengalami *financial distress*, maka dapat membuat para pemegang kepentingan perusahaan seperti kreditur dan pemegang saham kehilangan kepercayaan pada perusahaan tersebut. Dengan begitu para *stakeholder* tersebut akan mundur untuk bekerjasama dengan perusahaan dan apabila perusahaan gagal mencari jalan keluar maka perusahaan tersebut akan mengalami kebangkrutan. Seperti yang dilakukan oleh BEI (Bursa Efek Indonesia) pada kurun waktu 3 tahun terakhir telah melakukan *deleistung* pada 18 saham pada bank dan perusahaan yang tercatat. BEI adalah bursa efek atau yang juga dikenal dengan pasar modal yang telah berdiri dari 1912

di Batavia (sekarang Jakarta). Menjalani pasang surut selama beroperasi bursa efek ini sempat ditutup lagi dan kembali dibuka lagi pada 1977 oleh Presiden Soeharto. Hingga saat ini terdapat lebih dari 675 perusahaan yang tergabung (IDX, 2020).

Berdasarkan fenomena di atas maka perlu adanya sistem yang dapat digunakan untuk menganalisis sistem keuangan untuk mengetahui apakah terdapat kemungkinan suatu perusahaan mengalami *financial distress* yang berakibat pada kebangkrutan. Pada penelitian yang dilakukan oleh Kurniasari dan Ghozali (2013), untuk mengetahui kinerja keuangan dilakukan analisis terhadap rasio-rasio keuangan yaitu *Capital, Assets Quality, Management, Earnings, dan Liquidity* atau sering disebut rasio CAMEL. Dalam penelitian tersebut, Kurniasari dan Ghozali mengkategorikan tingkat kesehatan suatu bank menjadi sehat dan tidak sehat. Selain itu pada penelitian yang dilakukan oleh Almilia dan Kristijadi (2003) menyebutkan bahwa untuk menguji apakah suatu perusahaan mengalami *financial distress* dapat ditentukan melalui berbagai cara. Lau (1987) dan Hill, dkk. (1996) menggunakan adanya pemberhentian tenaga kerja atau menghilangkan pembayaran deviden untuk menentukan adanya *financial distress*. Sedangkan Asquith, dkk. (1994) menggunakan ICR (*Interest Coverage Ratio*) sebagai penentu. Pengertian ICR sendiri adalah rasio hutang dan profitabilitas yang digunakan untuk menentukan seberapa mudah suatu perusahaan dapat membayar bunga pinjamannya. Berbeda lagi dengan Whitaker (1999) yang mengukur *financial distress* dengan melihat adanya kas yang lebih kecil dari utang jangka panjang saat ini. Alat yang paling sering digunakan oleh peneliti dalam menentukan *financial distress* bagi perusahaan adalah ICR yang digunakan oleh Classens, dkk. (1999), Whardani (2006), Hapsari (2015), dan Astuti & Pamudji (2015). Keempat penelitian menyebutkan bahwa perusahaan yang memiliki ICR kurang dari satu adalah perusahaan yang mengalami *financial distress* dan perusahaan yang sehat secara keuangan atau tidak mengalami *financial distress* memiliki ICR lebih dari sama dengan satu.

Altman (1968) melakukan penelitian prediksi kebangkrutan perusahaan menggunakan rasio keuangan dan metode analisis

diskriminan. Altman membagi kategori perusahaan menjadi dua yaitu perusahaan yang mengalami kebangkrutan dan perusahaan yang tidak mengalami kebangkrutan. Penelitian mengenai *financial distress* juga pernah dilakukan oleh Widyarani (2018) yang melakukan prediksi bank umum di Indonesia mengalami *financial distress* atau tidak. Pada penelitian tersebut kondisi kesehatan keuangan bank umum dikategorikan menjadi dua yaitu bank yang mengalami *financial distress* dan bank sehat. Metode yang dilakukan dalam penelitian oleh Widyarani (2018) antara lain yaitu metode analisis diskriminan kernel, regresi logistik, dan *generalized extreme value regression*. Untuk mendapatkan prediksi dilakukan permodelan secara *univariate*, serentak, dan menggunakan variabel terpilih saja. Variabel yang digunakan Widyarani untuk memprediksi *financial distress* antara lain yaitu *Loan to Deposit Ratio*, *Return to Asset*, *Return on Equity*, *Net Interest Margin*, *Net Interest Margin*, Beban Operasi terhadap Pendapatan Operasional, Aktiva Produktif Bermasalah, *Non-Performing Loan*, Penyisihan Penghapusan Aktiva Produktif (PPAP) terhadap Aktiva Produktif, Pemenuhan PPAP, *Capital Adequacy Ratio*, Aktiva Tetap terhadap Modal, Inflasi, dan Pertumbuhan Ekonomi. Hasil penelitian menunjukkan bahwa metode regresi logistik lebih baik untuk memodelkan prediksi secara *univariate*. Sedangkan permodelan secara serentak menunjukkan bahwa metode GEVR lebih baik. Metode GEVR menunjukkan hasil yang lebih baik karena dapat menangkap adanya *unbalanced* data yaitu hanya terdapat 2 bank yang mengalami *financial distress* diantara 110 bank yang digunakan. Untuk menentukan model prediksi yang terbaik digunakan rasio akurasi, ROC (*Receiver Operating Characteristic*), dan AUC (*Area Under Curve*) atau area dibawah ROC. ROC dan AUC baik digunakan untuk mengatasi masalah dimana data yang digunakan adalah data *unbalanced* antara kategori yang ada.

Penelitian terkait *financial distress* juga dilakukan oleh Theodorus dan Artini (2018) yang dilakukan pada perusahaan perbankan yang tercatat di BEI. Menggunakan populasi yaitu perusahaan perbankan yang *go-public* yang beroperasi dan terdaftar di BEI dari periode 2010 hingga 2015. Tujuan dari

penelitian yang dilakukan oleh Theodorus dan Artini adalah mengetahui implikasi rasio CAMEL terhadap *financial distress* dan kemampuan rasio-rasio tersebut dalam memprediksi *financial distress*. Indikator yang digunakan untuk mengetahui apakah perusahaan perbankan mengalami *financial distress* adalah perusahaan yang mengalami kerugian minimal dalam waktu 2 tahun. Sedangkan variabel-variabel independen yang digunakan terdapat 6 variabel antara lain *Capital*, *Assets*, *Management*, *Earnings*, *Liquidity*, dan *Size*. Menggunakan metode regresi logistik dalam melakukan prediksi, diperoleh ketepatan klasifikasi *financial distress* yaitu sebesar 98,8%. Sedangkan variabel yang memberi pengaruh yang signifikan terhadap *financial distress* adalah variabel *Management* dan *Earnings*.

Pada 2005, Härdle, dkk. melakukan penelitian tentang prediksi kebangkrutan menggunakan metode *Support Vector Machines*. Tujuan dari penelitian ini adalah mengenalkan salah satu teknik statistika yang telah dikembangkan dan menjanjikan yaitu metode *Support Vector Machine* untuk melakukan analisis dalam menentukan kebangkrutan menggunakan rasio keuangan sebagai prediktornya. Melalui penelitian dapat dilihat bahwa metode SVM baik digunakan untuk mengekstraksi informasi dari data keuangan, meskipun data yang besar dibutuhkan untuk memanfaatkan seluruh kekuatan klasifikasinya. Metode SVM dapat dikatakan telah sukses diaplikasikan pada pengenalan karakter secara optik, diagnosis medis, dan klasifikasi teks. Selain itu dalam melakukan klasifikasi, metode SVM memberikan hasil yang lebih baik dibandingkan metode parametrik dan beberapa metode non-parametrik yang populer dan luas seperti *neural networks* yang dianggap sebagai salah satu metode dengan akurasi tinggi. Akurasi yang dihasilkan dari prediksi menggunakan SVM adalah sebesar 62% yang lebih besar dari metode analisis diskriminan yaitu sebesar 60%. Prastyo (2015) melakukan penelitian tentang penggunaan metode *Regularized Support Vector Machine* untuk melakukan prediksi kegagalan suatu perusahaan. Dengan menggunakan dua model yaitu *Lasso Absolute Shrinkage and Selection* (Lasso) dan *Smoothly Clipped Absolutely Deviation* (SCAD) didapat hasil prediksi dengan akurasi yang sama antar

kedua model. Namun model Lasso memberikan hasil interpretasi yang lebih masuk akan berdasarkan prediktor yang signifikan terhadap modelnya.

Salah satu metode yang juga digunakan oleh Widyarani (2018) untuk melakukan klasifikasi adalah analisis regresi logistik. Metode serupa yang telah dikembangkan yaitu  *$l_1$ -Regularized Logistic Regression* juga digunakan oleh Escobar & Menendez (2019) dalam melakukan klasifikasi biner pada proses monitor kualitas. Escobar & Menendez juga menyebutkan bahwa  *$l_1$ -Regularized Logistic Regression* dapat dengan baik diaplikasikan pada sistem manufaktur untuk deteksi kejadian terkait kualitas yang langka.  *$l_1$ -Regularized Logistic Regression* memberikan hasil yang baik pada data yang *unbalance*. Evaluasi dari hasil  *$l_1$ -Regularized Logistic Regression* diukur menggunakan AIC (*Akaike Infomation Criterion*), BIC (*Bayesian Information Criterion*), CEE (*Cross-Entropy Error*), dan McFAR2 (*McFadden's Adjusted R-Squared*). Pendekatan McFAR2 memberikan putusan yang sangat lemah untuk model yang kompleks.

Terdapat banyak alat atau uji yang dapat digunakan untuk mengukur akurasi prediksi klasifikasi. Widyarani (2018) menggunakan rasio akurasi, ROC, dan AUC untuk mengukur ketepatan klasifikasi prediksi *financial distress* pada bank umum di Indonesia. Penelitian yang dilakukan oleh SirElKhatim dan Salim (2017) tentang pengoptimalan metode *naïve bayes* dan regresi logistik dalam melakukan klasifikasi. Ketepatan klasifikasi diukur dengan beberapa ukuran yaitu melalui rasio akurasi, APER (*Apparent Error Rate*), *specificity*, dan *sensitivity*. Dewi (2016) menggunakan ROC dan AUC untuk mengukur ketepatan klasifikasi *data mining* pada prediksi keberhasilan pemasaran produk layanan perbankan. Huang & Ling (2005) menyatakan bahwa akhir-akhir ini AUC diusulkan sebagai alat ukur ketepatan klasifikasi hasil prediksi pertama dibandingkan ukuran lainnya. AUC memberikan hasil yang konsisten secara statistika dan lebih dikriminatif daripada ukuran akurasi lainnya pada data biner atau multinomial baik yang *balanced* maupun *unbalanced*. Selain itu, AUC juga *robust* terhadap data yang *unbalanced*. Hasil prediksi

yang baik memiliki nilai AUC yang lebih tinggi dari pada model prediksi lainnya.

Menggunakan latar belakang yang telah diuraikan, maka penelitian tentang prediksi *financial distress* terhadap perusahaan sektor industri menggunakan metode-metode yang dianggap akan memberikan hasil prediksi terbaik. Metode-metode yang dimaksudkan antara lain *Support Vector Machine*, *Generalized Extreme Value Regression*, dan *Logistic Regression*. Prediktor yang digunakan dalam penelitian ini merupakan rasio-rasio keuangan yang didapat melalui perhitungan dari laporan keuangan yang ada. Rasio-rasio keuangan diidentifikasi sering terjadi adanya hubungan linier antar rasio keuangan atau terjadi multikolinieritas. Antar rasio keuangan yang memiliki hubungan linier dapat menyebabkan adanya rasio keuangan yang tidak berpengaruh signifikan sehingga dalam kasus klasifikasi dapat menyebabkan *overfitting*, maka dari itu untuk dilakukan seleksi variabel juga pada penelitian ini. Dalam melakukan seleksi variabel digunakan teknik regularisasi pada metode *Support Vector Machine* dan *Logistic Regression*. Regularisasi dilakukan untuk mendapatkan variabel yang terseleksi guna menghindari *overfitting*, terutama ketika hanya terdapat sedikit data *training* atau ketika terdapat parameter dengan jumlah yang besar. Seleksi variabel juga dilakukan pada metode *Generalized Extreme Value Regression* dan *Logistic Regression* dengan pendekatan *stepwise*. Pada metode *Generalized Extreme Value Regression* dan *Logistic Regression* akan menghasilkan *output* dalam bentuk probabilitas, maka agar dapat dibandingkan melalui probabilitas pada metode *Support Vector Machine* dilakukan *Isotonic Regression* pada skor yang dihasilkan dengan *Pool Adjacent Violators Algorithm*. Dalam penelitian ini juga akan dilakukan *windowing* terhadap *time* atau waktu menggunakan pendekatan *full memory time window*. Model-model yang didapatkan akan dibandingkan untuk mendapat model terbaik menggunakan AUC untuk mendapatkan variabel-variabel yang berpengaruh secara signifikan.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah disebutkan pada sub bab sebelumnya, sektor industri merupakan sektor yang menyumbang

kontribusi terbesar bagi perekonomian di Indonesia. Hal tersebut tentu dapat terjadi lagi pada periode-periode berikutnya mengingat sumber daya yang akan diolah masih melimpah. Perusahaan industri memiliki kewajiban untuk menjaga stabilitas keuangan agar tidak mengalami masalah pada sistem keuangannya yang apabila terjadi gangguan pada stabilitas keuangannya akan mengakibatkan peristiwa yaitu *financial distress* dan bila tidak teratas akan mengakibatkan kebangkrutan. Kebangkrutan perusahaan sektor industri secara besar-besaran akan mengakibatkan gangguan perekonomian di Indonesia. Secara umum permasalahan yang dibahas dalam penelitian ini adalah bagaimana prediksi *financial distress* terhadap perusahaan sektor industri di Indonesia. Prediksi tersebut dilakukan menggunakan rasio-rasio keuangan yang ada pada laporan keuangan yang didapat. Namun melalui seluruh rasio keuangan yang ada bias terdapat kejadian multikolinieritas antar rasio keuangan dalam memprediksi *financial distress* sehingga prediksi dilakukan dengan serentak dan dengan seleksi variabel. Secara spesifik permasalahan yang akan dibahas dalam penelitian ini adalah sebagai berikut:

1. Bagaimana karakteristik rasio keuangan dari perusahaan sektor industri yang sehat dan perusahaan sektor industri yang mengalami *financial distress* di Indonesia?
2. Bagaimana prediksi *financial distress* perusahaan sektor industri di Indonesia menggunakan metode *Support Vector Machine*, *Generalized Extreme Value Regression*, dan *Logistic Regression* dengan seleksi variabel?
3. Variabel apa saja yang memiliki pengaruh signifikan terhadap kondisi keuangan perusahaan sektor industri di Indonesia?

### **1.3 Tujuan**

Menjawab rumusan penelitian di atas, maka tujuan dari penelitian ini adalah prediksi *financial distress* pada perusahaan sektor industri di Indonesia. Tujuan penelitian ini secara spesifik dijelaskan pada poin-poin di bawah ini:

1. Mendapatkan karakteristik rasio keuangan dari perusahaan sektor industri yang sehat dan perusahaan sektor industri yang bermasalah di Indonesia.
2. Mendapatkan prediksi *financial distress* perusahaan sektor industri di Indonesia menggunakan metode *Support Vector Machine*, *Generalized Extreme Value Regression*, dan *Logistic Regression* dengan seleksi variabel.
3. Mendapatkan variabel-variabel yang memiliki pengaruh signifikan terhadap kondisi keuangan perusahaan sektor industri di Indonesia.

#### **1.4 Manfaat Penelitian**

Penelitian ini dharapkan dapat memberi manfaat secara umum kepada seluruh pihak. Secara khusus manfaat yang dapat diambil antara lain:

1. Mejadi *early warning system* kepada pihak-pihak terkait seperti pemerintah, kreditur, dan pihak-pihak yang akan mendapat dampak dari kebangkrutan perusahaan sektor industri terkait dan perusahaan itu sendiri agar melakukan antisipasi bila terjadi penurunan nilai ICR.
2. Memberi pemahaman mengenai metode-metode yang digunakan dalam analisis diantaranya yaitu metode *Support Vector Machine*, *Generalized Extreme Value Regression*, dan *Logistic Regression* dengan seleksi variabel.

#### **1.5 Batasan Masalah**

Dalam melakukan penelitian ini digunakan beberapa batasan agar penelitian dapat terfokus pada topik dan tujuan. Batasan-batasan yang digunakan antara lain:

1. Data yang digunakan merupakan data laporan keuangan kuartalan sejak tahun 2005 hingga 2018 dari perusahaan sektor industri yang tercatat oleh Bank Indonesia.
2. Data yang digunakan merupakan data rasio keuangan kuartal empat tiap tahun kecuali pada tahap *imputing* data *missing value* menggunakan data seluruh kuartal.

*(Halaman ini sengaja dikosongkan)*

## **BAB II**

### **TINJAUAN PUSTAKA**

Dalam tinjauan pustaka akan dibahas metode-metode yang akan digunakan dalam penelitian ini. Metode yang dimaksud antara lain yaitu *Support Vector Machine*, *Generalized Extreme Value Regression*, dan *Logistic Regression*. Selain itu teori-teori terkait maupun asumsi-asumsi yang mendukung metode-metode yang telah disebutkan juga dibahas dalam bab ini.

#### **2.1 Statistika Deskriptif**

Statistika deskriptif yaitu merupakan metode-metode yang berkaitan dengan pengumpulan dan penyajian suatu gugus data sehingga dapat memberikan informasi yang berguna. Dalam menyajikan data, statistika deskriptif menyajikan dalam bentuk tabel, diagram grafik dan dengan beberapa ukuran lainnya (Irianto, 2012).

Salah satu cara visualisasi data yang popular adalah *density plot*. Merupakan variasi dari histogram yang menggunakan *kernel smoothing*. Puncak dari *density plot* menunjukkan nilai yang memiliki frekuensi tertinggi dalam interval. Histogram adalah grafik dari distribusi frekuensi suatu variabel. Penyajian data dalam histogram yaitu sumbu absis adalah sumbu X dan Y sebagai ordinat. Sumbu menunjukkan frekuensi dan sumbu x merupakan kelas-kelas interval variabel yang divisualisaikan.

Ukuran untuk menggambarkan data secara umum dibagi menjadi ukuran pemusatan dan penyebaran data. Ukuran pemusatan data adalah sembarang ukuran yang menunjukkan pusat segugus data. Beberapa contoh ukuran pemusatan data antara lain yaitu *mean*, *median*, dan *modus*. *Mean* atau rata-rata adalah nilai yang diperoleh dari jumlah sekelompok data dibagi dengan banyaknya data. Median adalah nilai data yang terletak di tengah setelah data diurutkan. Sedangkan modus adalah nilai yang memiliki frekuensi paling besar dalam suatu gugus data. Ukuran lainnya yaitu ukuran penyebaran data, suatu ukuran yang menyatakan seberapa besar nilai-nilai data berbeda atau bervariasi

dengan nilai ukuran pusatnya atau seberapa besar penyimpangan nilai-nilai data dengan nilai pusatnya. Salah satu nilai penyebaran data yang sering digunakan yaitu varians adalah ukuran seberapa jauh sebuah kumpulan bilangan tersebut.

## 2.2 Imputasi *k-Nearest Neighbor*

Imputasi *missing value* menggunakan metode *k-Nearest Neighbor* (kNN) didasarkan pada observasi donor. Agregat nilai  $k$  yang dianggap nilai yang berdekatan merupakan nilai yang digunakan mengisi *missing value* pada suatu gugus data. Perhitungan jarak yang dilakukan untuk menentukan nilai terdekat menggunakan perpanjangan dari jarak Gower yang dapat mengatasi jarak pada variabel-variabel dengan tipe kategori maupun kontinu. Jarak antar dua observasi adalah kontribusi dari rata-rata masing-masing variabel yang terboboti, dimana bobot yang digunakan mewakili prioritas tiap variabel (Kowarik & Templ, 2016). Berikut merupakan rumus yang dapat digunakan untuk menentukan jarak antar observasi ke- $i$  dan ke- $j$ :

$$d_{i,j} = \frac{\sum_{k=1}^p w_k \delta_{i,j,k}}{\sum_{k=1}^p w_k}, \quad (2.1)$$

dengan  $w_k$  adalah bobot dan  $\delta_{i,j,k}$  adalah kontribusi variabel ke- $k$ .

Nilai dari  $\delta_{i,j,k}$  untuk variabel kontinu dapat dihitung menggunakan persamaan berikut:

$$\delta_{i,j,k} = |x_{i,k} - x_{j,k}|r_k, \quad (2.2)$$

dengan  $x_{i,k}$  adalah nilai dari observasi ke- $i$  pada variabel ke- $k$  dan  $x_{j,k}$  merupakan nilai dari observasi ke- $j$  pada variabel ke- $k$ . Sedangkan  $r_k$  menotasikan *range* dari variabel ke- $k$ . Untuk variabel nominal yang biner (0/1), jarak didapat menggunakan rumus berikut:

$$\delta_{i,j,k} = \begin{cases} 0 & \text{jika } x_{i,k} = x_{j,k}, \\ 1 & \text{jika } x_{i,k} \neq x_{j,k}. \end{cases} \quad (2.3)$$

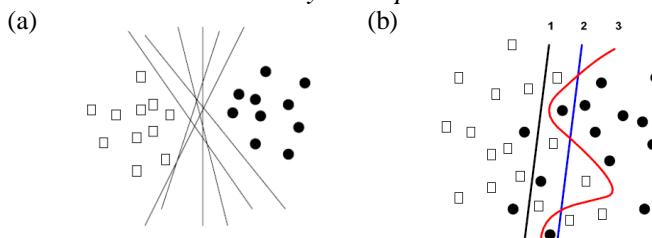
## 2.3 Support Vector Machine

SVM adalah salah satu metode dalam *machine learning* yang dapat dikatakan paling bagus untuk klasifikasi. *Support Vector*

*Machine* ditemukan dan diperkenalkan pertama kali oleh Vapnik pada tahun 1992. Metode SVM merupakan salah satu metode yang paling menjanjikan diantara metode dalam *machine learning* karena memiliki generalisasi yang tinggi dan mampu diterapkan secara luas dengan baik untuk masalah klasifikasi, peramalan, dan estimasi (Wu & Wang, 2013). Cara kerja metode SVM adalah dengan memaksimumkan marjin atau jarak antar kelas data menggunakan *optimum hyperplane*. Berdasarkan jarak antar kelas tersebut, pada dasarnya SVM adalah *linear classifier* yang telah dikembangkan agar dapat bekerja pada masalah *nonlinear classifier*.

### 2.3.1 Support Vector Machine Linear Classifier

Dalam kasus *linear classifier*, SVM merupakan alat klasifikasi dengan batas keputusan secara linier. Klasifikasi linier SVM dibagi lagi menjadi dua berdasarkan pemisahnya yaitu SVM *in linearly separable* dan SVM *in linearly nonseparable*.



**Gambar 2.1** Klasifikasi Linier SVM (a) *linearly separable* dan (b) *linearly nonseparable*. Sumber: (Härdle, dkk., 2014)

#### 1. Support Vector Machine in Linearly Separable

Setiap observasi memiliki pasangan prediktor sejumlah  $p$ ,  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip}) \in \mathbb{R}^p$  dengan  $i = 1, 2, \dots, n$  dimana  $n$  adalah banyaknya observasi. Sementara  $y_i \in \mathbf{y} \in \{-1, 1\}$  adalah label atau kelas dari data respon. Apabila  $\mathbf{x}_i$  merupakan anggota dari kelas 1, maka  $\mathbf{x}_i$  memiliki label  $y_i = 1$ , begitu pula sebaliknya. Data yang diberikan merupakan himpunan data *training* dari kedua kelas berupa data pasangan yang akan diklasifikasi. Himpunan tersebut dapat dituliskan dalam persamaan berikut (Härdle, dkk., 2014).

$$D_n = \{(x_1, y_1), \dots, (x_n, y_n)\} \in \mathbb{R}^p \times \{-1, 1\}. \quad (2.4)$$

$D_n$  berisi informasi untuk memprediksi  $y$  pada observasi baru. Label  $y_i$  pada data *training* disebut *trainor* atau supervisor. Konsep utama yang digunakan untuk menetapkan pemisah pada SVM *in linearly separable* adalah *dot product*. Keluarga  $F$  dari fungsi klasifikasi yang terdapat pada ruang data adalah sebagai berikut.

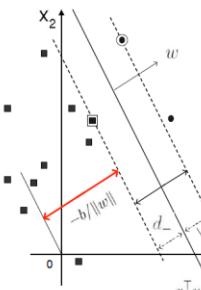
$$F = \{\mathbf{x}_i' \mathbf{w} + b, \mathbf{w} \in \mathbb{R}^p, b \in \mathbb{R}\}, \quad (2.5)$$

dengan  $\mathbf{w}$  adalah vektor bobot yang tegak lurus terhadap *hyperplane* dan  $b$  merupakan posisi-posisi bidang relatif terhadap pusat koordinat yang ditunjukkan pada Gambar 2.2. Persamaan *hyperplane* pemisah dituliskan pada persamaan berikut.

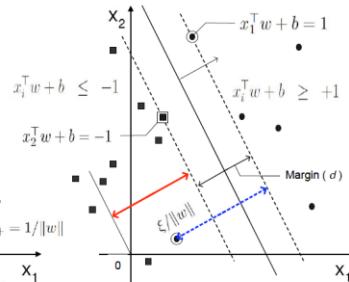
$$f(\mathbf{x}_i') = \mathbf{x}_i' \mathbf{w} + b = 0, \quad (2.6)$$

dimana  $\mathbf{x}_i' \mathbf{w} + b \geq 1$  untuk  $y_i = 1$  merupakan bidang pembatas kelas 1 dan  $\mathbf{x}_i' \mathbf{w} + b \leq -1$  untuk  $y_i = -1$  merupakan bidang pembatas kelas 2.

(a)



(b)



Gambar 2.2 Hyperplane Klasifikasi Linier SVM. Sumber: (Härdle, dkk., 2014)

Berdasarkan Gambar 2.2 dapat ditunjukkan *hyperplane* memisahkan ruang menjadi dua bagian. Panjang vektor  $\mathbf{w}$  adalah  $\|\mathbf{w}\| = \sqrt{\mathbf{w}'\mathbf{w}}$ . Bidang pembatas pertama dan kedua mempunyai bobot  $\mathbf{w}$  dimana jarak tegak lurus dari titik asal adalah sebesar  $\frac{|1-b|}{\|\mathbf{w}\|}$  dan  $\frac{|-1-b|}{\|\mathbf{w}\|}$ . Sedangkan jarak antara *margin* dengan *hyperplane* (bidang pemisah) adalah  $d_+ = d_- = \frac{1}{\|\mathbf{w}\|}$ . Sehingga didapatkan nilai maksimum *margin* (berdasarkan rumus jarak garis ke titik pusat) adalah  $\frac{1-b-(1-b)}{\|\mathbf{w}\|} - \frac{2}{\|\mathbf{w}\|}$  atau ekuivalen dengan  $\min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|^2$  (Härdle, dkk., 2014).

Menentukan *hyperplane* terbaik dapat menggunakan *Quadratic Programming* (QP) *problem* yaitu mencari titik minimal persamaan (2.7) dengan memperhatikan *constraint* pada persamaan (2.8).

$$\min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|^2, \quad (2.7)$$

$$y_i(\mathbf{x}'_i \mathbf{w} + b) - 1 \geq 0. \quad (2.8)$$

Lagrangian untuk masalah mendasar pada kasus ini adalah sebagai berikut.

$$\min_{\mathbf{w}, b} L_p(\mathbf{w}, b) = \min_{\mathbf{w}, b} \left[ \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i (y_i(\mathbf{x}'_i \mathbf{w} + b) - 1) \right]. \quad (2.9)$$

Meminimalkan  $L$  terhadap  $\mathbf{w}$  dan  $b$  dapat diberikan sebagai berikut.

$$\begin{aligned} \frac{\partial L_p(\mathbf{w}, b)}{\partial \mathbf{w}} &= 0, \\ \mathbf{w} - \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i &= 0, \end{aligned} \quad (2.10)$$

$$\begin{aligned} \mathbf{w} &= \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i, \\ \frac{\partial L_p(\mathbf{w}, b)}{\partial b} &= 0, \end{aligned} \quad (2.11)$$

$$\sum_{i=1}^n \alpha_i y_i = 0.$$

Lagrangian untuk permasalahan ganda dilakukan dengan mensubstitusi persamaan (2.10) dan (2.11) adalah sebagai berikut.

$$\max_{\alpha} L_D(\alpha) = \max_{\alpha} \left[ \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}'_i \mathbf{x}_j \right]. \quad (2.12)$$

Setelah menyelesaikan permasalahan berganda, pengklasifikasian suatu objek dapat dilakukan dengan aturan berikut.

$$\hat{f}(\mathbf{x}'_i) = \text{sign}(\mathbf{x}' \hat{\mathbf{w}} + b), \quad (2.13)$$

dengan  $\hat{\mathbf{w}} = \sum_{i=1}^n \hat{\alpha}_i y_i \mathbf{x}_i$  dengan  $\mathbf{x}_i$  adalah *support vector*. Nilai  $b = \frac{1}{n_{sv}} \left( \sum_{i=1}^{n_{sv}} \frac{1}{y_i} - (\mathbf{x}'_i \hat{\mathbf{w}}) \right)$  dengan  $\mathbf{x}'_i$  merupakan data yang diklasifikasikan dan  $n_{sv}$  adalah jumlah *support vector*.

## 2. Support Vector Machine in Linearly Nonseparable

Kasus *linearly nonseparable*, pemisalah antar kelas tidak bersifat linier sempurna sehingga diperlukan variabel *slack*  $\xi$ . Penggunaan variabel tersebut dapat menyebabkan suatu titik berada di dalam *error margin*  $0 \leq \xi_i \leq 1$  atau disebut

misklasifikasi,  $\xi > 1$  sehingga klasifikasi  $\mathbf{x}_i$  adalah sebagai berikut (Härdle, dkk., 2014).

$$\begin{aligned} \mathbf{x}'_i \mathbf{w} + b &\geq 1 - \xi_i, \text{ untuk } y_i = 1, \\ \mathbf{x}'_i \mathbf{w} + b &\geq -(1 - \xi_i), \text{ untuk } y_i = -1 \\ \xi_i &\geq 0, \end{aligned}$$

dimana persamaan di atas dapat digabung menjadi

$$y_i = (\mathbf{x}'_i \mathbf{w} + b) \geq 1 - \xi_i, \quad (2.14a)$$

$$\xi_i \geq 0. \quad (2.14b)$$

Penalti untuk kesalahan klasifikasi terkait dengan jarak titik kesalahan klasifikasi  $\mathbf{x}_i$  dari *hyperplane* yang membatasi kelasnya. Jika  $\xi_i > 0$ , maka kesalahan dalam memisahkan dua set terjadi. Fungsi objektif dilakukan dengan memaksimalkan *penalized margin* kemudian diformulasikan ke dalam rumus berikut.

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

dengan nilai  $\xi_i > 0$  dan parameter pinalti  $C > 0$  dimana  $C$  adalah parameter yang menentukan besar kecilnya bobot akibat *misclassification* yang nilainya ditentukan.

Fungsi *lagrange* dalam kasus ini adalah sebagai berikut:

$$L_p(\mathbf{w}, b, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i (y_i (\mathbf{x}'_i \mathbf{w} + b) - 1 + \xi_i) - \sum_{i=1}^n \mu_i \xi_i, \quad (2.16)$$

dengan  $\alpha_i \geq 0$  dan  $\mu_i \geq 0$  merupakan *Lagrange Multiplier*. Nilai optimal dari persamaan (2.16) di atas dapat dihitung dengan meminimalkan  $L$  terhadap  $\mathbf{w}, b$  dan  $\xi$  serta memaksimalkan  $L$  terhadap  $\alpha$  sehingga diperoleh persamaan di bawah ini.

$$\max_{\alpha} L_p(\alpha) = \max_{\alpha} \left( \min_{\mathbf{w}, b, \xi} L(\mathbf{w}, b, \xi) \right). \quad (2.17)$$

Sedangkan untuk meminimalkan  $L$  terhadap  $\mathbf{w}, b$  dan  $\xi$  diberikan pada persamaan berikut.

$$\begin{aligned} \frac{\partial L_p(\mathbf{w}, b, \xi)}{\partial \mathbf{w}} &= 0, \\ \mathbf{w} - \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i &= 0, \end{aligned} \quad (2.18)$$

$$\begin{aligned} \mathbf{w} &= \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i, \\ \frac{\partial L_p(\mathbf{w}, b, \xi)}{\partial b} &= \sum_{i=1}^n \alpha_i y_i = 0, \end{aligned} \quad (2.19)$$

$$\begin{aligned}\frac{\partial L_p(\mathbf{w}, b, \xi)}{\partial \xi} &= 0, \\ \frac{\partial L_p(\mathbf{w}, b, \xi)}{\partial \xi} &= C - \alpha_i - \mu_i = 0, \quad (2.20) \\ \alpha_i &= C - \mu_i.\end{aligned}$$

Dari persamaan-persamaan di atas dapat dilakukan substitusi ke dalam persamaan 2.15).

$$\max_{\alpha} L_D(\alpha) = \max_{\alpha} \left[ \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i' \mathbf{x}_j \right]. \quad (2.21)$$

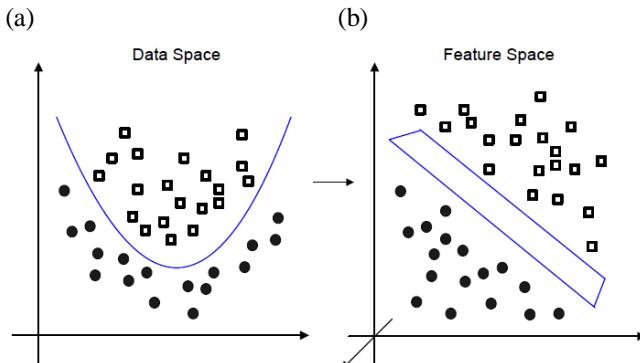
*Constraint* yang digunakan untuk memaksimalkan  $\alpha_i$  diberikan pada persamaan di bawah ini.

$$\sum_{i=1}^n \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C. \quad (2.22)$$

Sampel  $\mathbf{x}_i$  untuk  $\alpha_i > 0$  (*support vector*) merupakan titik yang berada di atas marjin atau di dalam marjin ketika *soft margin* mungkin digunakan. *Support vector* sering menyebar dan level penyebarannya berada pada batas atas untuk *misclassification rate* (Schölkopf & Smola, 2002).

### 2.3.2 Support Vector Machine Nonlinear Classifier

Kasus yang sering ditemui dalam dunia nyata lebih banyak bersifat nonlinier daripada linier. Hal tersebut menyebabkan SVM dikembangkan dengan memasukkan fungsi kernel nonlinier agar SVM dapat digunakan dengan baik pada kasus nonlinier. Data nonlinier dipetakan dengan menggunakan fungsi pemetaan  $\psi: \mathcal{R}^p \rightarrow \mathbf{H}$  ke dalam ruang yang berdimensi tinggi  $\mathbf{H}$ , dimana aturan klasifikasi bersifat linier yang ilustrasinya dapat dilihat pada Gambar 2.3 di bawah ini. Menggunakan Gambar 2.3 (a) dapat dinyatakan bahwa data tidak dapat dipisahkan secara linier menggunakan pemisah dua dimensi sehingga data dipetakan dengan fungsi  $\psi$  menjadi ruang tiga dimensi (b) dan data dapat dipisahkan secara linier.



**Gambar 2.3** Pemetaan Ruang Dua Dimensi Data Menjadi Tiga Dimensi.  
Sumber: (Härdle, dkk., 2014)

Fungsi transformasi pada SVM menggunakan *kernel trick* yang digunakan untuk menghitung *scalar product* melalui fungsi *kernel* (Schölkopf & Smola, 2002). Proyeksi fungsi  $\psi: \mathcal{R}^p \rightarrow \mathcal{H}$  memastikan bahwa *inner product*  $\psi(\mathbf{x}_i)' \psi(\mathbf{x}_j)$  ditunjukkan oleh fungsi *kernel* berikut:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \psi(\mathbf{x}_i)' \psi(\mathbf{x}_j). \quad (2.23)$$

Optimasi pada persamaan (2.21) diubah menjadi persamaan di bawah ini:

$$\max_{\alpha} L_D(\alpha) = \max_{\alpha} \left[ \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \right]. \quad (2.24)$$

*Constraint* yang digunakan untuk memaksimalkan  $\alpha_i$  di atas dituliskan dalam persamaan berikut.

$$\sum_{i=1}^n \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, n. \quad (2.25)$$

Salah satu model *kernel* yang sering digunakan adalah model *Radial Basis Function* (RBF) *kernel* yang dirumuskan pada persamaan di bawah:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp \left( -\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2} \right). \quad (2.26)$$

Berdasarkan adanya tambahan fungsi *kernel* maka didapatkan fungsi keputusan sebagai berikut.

$$\hat{f}(\mathbf{x}_i) = \text{sign} \left( \sum_{i=1}^n \hat{\alpha}_i y_i K(\mathbf{x}_i, \mathbf{x}_j) + \hat{b} \right). \quad (2.27)$$

## 2.4 Regularized Support Vector Machine

Pemisah nonlinier didapatkan melalui penambahan *kernel trick* pada *dual problem* yang akan ditunjukkan pada persamaan di bawah.

$$\begin{aligned} & \max_{\alpha} L_D(\alpha) \\ & s.t. \sum_{i=1}^n \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, \end{aligned} \quad (2.28)$$

dengan  $L_D(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$ , dan  $K(\mathbf{x}_i, \mathbf{x}_j)$  merupakan fungsi *kernel* dan  $\alpha$  merupakan *Lagrangian Multiplier*. Konstanta  $C$  merupakan *cost parameter* yang menunjukkan besarnya misklasifikasi.

Standar  $L_2$ -norm SVM dapat dituliskan dalam bentuk *regularization* seperti berikut (Prastyo, 2015).

$$\min_{w,b} \sum_{i=1}^n \{1 - y_i f(\mathbf{x}_i)\}_+ + \lambda \|w\|_2^2, \quad (2.29)$$

dengan  $\lambda > 0$  adalah *tunning parameter* dan *Hinge-loss function* adalah sebagai berikut.

$$\begin{aligned} L\{y_i, f(\mathbf{x}_i)\} &= \{1 - y_i f(\mathbf{x}_i)\}_+ \\ &= \max\{0, 1 - y_i f(\mathbf{x}_i)\}. \end{aligned} \quad (2.30)$$

*Penalty function* pada persamaan (2.29) adalah  $P(w) = \|w\|_2^2$  menggunakan seluruh input untuk mendapatkan pemisah sehingga tidak dapat memilih prediktor yang relevan. *Regularization SVM* menggunakan istilah untuk penalti lain yang dapat memilih prediktor yang relevan bersama dengan langkah-langkah estimasi parameter.

### 2.4.1 Lasso Absolute Shrinkage and Selection Support Vector Machine

Salah satu teknik yang digunakan dalam metode *Regularized* adalah *Lasso Absolute Shrinkage and Selection*. Metode ini lebih akrab dikenal dengan istilah *Lasso SVM* dimana istilah yang digunakan sebelum *Lasso* adalah  $L_1$ -norm yang digunakan dalam penelitian Bradley & Magasarian (1998) dan Zhu, dkk., (2004). Teknik *Lasso* sendiri pertama kali diperkenalkan oleh Tibshirani (1996). *Lasso* merupakan teknik yang membangun model secara bersamaan dan dapat menghasilkan model yang akurat dan bersifat *parsimonious*.

$$\min_{\mathbf{w}, \mathbf{b}} \sum_{i=1}^n \{1 - y_i f(\mathbf{x}_i)\}_+ + \lambda \|\mathbf{w}\|_1, \quad (2.31)$$

dimana  $\lambda > 0$  merupakan *tunning parameter*.

#### 2.4.2 Elastic-Net Support Vector Machine

*Elastic-net* merupakan salah satu metode regularisasi yang digunakan untuk mengatasi kelemahan dari metode *Lasso*. *Elastic-net* pertama kali dipopulerkan oleh Zou & Hastie (2005), dimana *elastic-net* merupakan campuran dari penalti *L<sub>1</sub>-norm* dan penalti *L<sub>2</sub>-norm*. Kelebihan dari *elastic-net* untuk mengatasi kekurangan *Lasso* adalah pemilihan prediktor tidak dibatasi ukuran sampel dan kelompok prediktor yang berkorelasi dapat dipilih bersama-sama (pemilihan kelompok). Penalti *L<sub>1</sub>-norm* berperan dalam pemilihan prediktor, sedangkan penalti *L<sub>2</sub>-norm* berperan dalam pemilihan kelompok. Fungsi persamaan *elastic-net* adalah sebagai berikut.

$$\min_{\mathbf{w}, \mathbf{b}} \sum_{i=1}^n \{1 - y_i f(\mathbf{x}_i)\}_+ + \lambda_1 \|\mathbf{w}\|_1 + \lambda_2 \|\mathbf{w}\|_2^2, \quad (2.32)$$

dengan  $\lambda_1$  dan  $\lambda_2$  adalah *tunning parameter*.

#### 2.5 Isotonic Regression

SVM merupakan salah satu metode klasifikasi dengan hasil akhir atau *output* berupa skor. Baru-baru ini terdapat upaya untuk memetakan skor SVM tersebut ke dalam bentuk probabilitas. Salah satu pendekatan yang umum adalah menggunakan *isotonic regression* untuk memetakan skor SVM ke dalam bentuk probabilitas. Fungsi *isotonic* memiliki tren monoton naik. Jika skor SVM diperengkatkan dengan benar, maka probabilitas  $P(Y_j = 1 | \mathbf{x}_i)$  dimana  $Y_j \in \{-1, 1\}$  diasumsikan akan naik seiring kenaikan skor SVM. Sehingga *isotonic regression* dapat digunakan untuk memetakan skor SVM dalam bentuk probabilitas. Algoritma yang umum digunakan pada *isotonic regression* adalah *Pool-Adjacent-Violators Algorithm* (PAVA) (Leeuw, dkk., 2009).

Pada algoritma ini  $g(x_{(i)})$  adalah probabilitas masing-masing skor  $x_i$  terperengkat. Tahap pertama yaitu memetakan  $g(x_{(i)}) = 1$  untuk contoh positif dan  $g(x_{(i)}) = 0$  pada pengamatan lainnya.

Jika  $g(x_{(i)})$  sudah *isotonic* (monoton naik) untuk semua  $x_{(i)}$ , maka persyaratan telah terpenuhi. Namun jika  $g(x_{(i-1)}) > g(x_{(i)})$  maka nilai  $x_{(i-1)}$  dan  $x_{(i)}$  sebagai *pair-adjacent* melanggar asumsi (*violator*). Nilai dari  $g(x_{(i-1)})$  dan  $g(x_{(i)})$  digantikan dengan masing-masing rata-rata, hingga nilai-nilai tersebut tidak melanggar asumsi *isotonic*. Tahap ini diulang terus-menerus hingga didapat nilai yang *isotonic*. Pada akhirnya, akan didapatkan daftar probabilitas bersama dengan skor SVM yang berdekatan dengan pengamatan data *training*. Sehingga pada data baru, dapat dihitung probabilitas yang dipetakan berdasarkan skor yang diperoleh  $x_i$  dari aturan keputusan SVM. Biasanya, akan ada interval skor dengan probabilitas yang sama. Karena tidak ada batasan yang dikenakan pada skor SVM, interval terendah dimulai dari  $-\infty$  dan interval tertinggi berakhir pada  $+\infty$  (Lambrou, dkk., 2012).

## 2.6 Extreme Value Theory

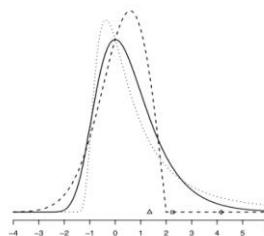
Pada umumnya *extreme value theory* (EVT) digunakan untuk menganalisis kejadian yang dianggap sangat jarang terjadi atau kejadian ekstrim yang memiliki dampak yang besar (Dharmawan, 2012). Terdapat tiga distribusi nilai ekstrim yang mempresentasikan distribusi  $G$  adalah sebagai berikut:

Fréchet:  $G_{1,\tau}(x) = e^{(-x^{-\tau})}$ ,  $x \geq 0$ , untuk  $\tau > 0$

Gumbel:  $G_0(x) = e^{(-x^{-\tau})}$ ,  $x \in R$

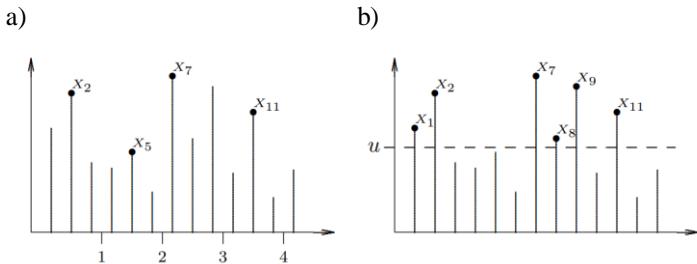
Weibull:  $G_{2,\tau}(x) = e^{(-|x|^{-\tau})}$ ,  $x \leq 0$ , untuk  $\tau < 0$

dimana  $\tau$  adalah kuantil yang digunakan dalam perhitungan tingkat resiko.



**Gambar 2.4** Probability Density Function dari Distrbusi Fréchet (dotted), Distribusi Gumbel (solid), dan Distribusi Weibull (dashed). Sumber: (Wang & Dey, 2010)

Dalam EVT terdapat dua pendekatan yang ada untuk saat ini yaitu pendekatan *Block Maxima* (BM) dan *Peak Over Threshold* (POT).



**Gambar 2.5** a) Skema *Block Maxima* dan b) Skema *Peak Over Threshold*.

Sumber: (Gilli & Kellezi, 2006).

Pendekatan *block maxima* mempertimbangkan variabel maksimum yang dibutuhkan dalam periode berturut-turut. Observasi yang terpilih merupakan suatu kejadian ekstrim yang disebut *block maxima*. Sedangkan pendekatan *peak over threshold* berfokus pada realisasi yang melampaui ambang batas (tinggi) yang diberikan (Gilli & Kellezi, 2006).

## 2.7 Generalized Extreme Value Distribution

*Generalized Extreme Value* (GEV) distribution pertama kali diperkenalkan oleh Jenkinson. Distribusi GEV merupakan salah satu dari distribusi probabilitas kontinu yang berkaitan erat dengan EVT. Distribusi GEV memiliki tiga parameter dengan berbagai metode yang dapat digunakan untuk mengestimasi parameternya, yaitu  $\tau$  merupakan parameter bentuk,  $\mu$  adalah parameter lokasi, dan  $\sigma$  merupakan parameter skala. Variabel random  $\mathbf{X}$  dikatakan beristribusi GEV jika memiliki *probability density function* (pdf) sebagai berikut:

$$f(x) = \frac{1}{\sigma} \exp[-(1 - \tau)Y - \exp(-Y)], \quad (2.33)$$

$$\text{dengan, } Y = \begin{cases} -\frac{1}{\tau} \log \left( 1 - \frac{\tau(x-\mu)}{\sigma} \right), & \tau \neq 0 \\ \frac{(x-\mu)}{\sigma}, & \tau = 0. \end{cases}$$

Berikut adalah *cummulative density function*:

$$F_X(x) = \exp \left\{ - \left[ 1 + \tau \left( \frac{x-\mu}{\sigma} \right) \right]^{-1/\tau} \right\}, \quad (2.34)$$

dengan  $-\infty < \tau < +\infty$ ,  $-\infty < \mu < +\infty$ , dan  $\alpha > 0$  (Calabrese & Osmetti, 2013).

## 2.8 Generalized Extreme Value Regression

*Generalized Extreme Value Regression* (GEVR) merupakan salah satu pemodelan regresi *Generalized Linear Model* (GLM) dengan variabel dependen bersifat biner. GEVR merupakan fungsi kuantil dari distribusi GEV sebagai *link function*. GEVR digunakan untuk mengatasi kekurangan regresi logistik untuk mengestimasi kejadian langka atau *extreme value* pada suatu gugus data. Probabilitas bank ke- $i$  dikategorikan mengalami *financial distress* adalah sebagai berikut (Calabrese & Osmetti, 2013):

$$\pi(\mathbf{x}_i) = \exp \left\{ - [1 + \tau(\boldsymbol{\beta}' \mathbf{x}_i)]^{-1/\tau} \right\}. \quad (2.35)$$

*Link function* untuk GEVR adalah sebagaimana berikut:

$$\frac{-\ln[\pi(\mathbf{x}_i)]^{-\tau}-1}{\tau} = \boldsymbol{\beta}' \mathbf{x}_i, \quad (2.36)$$

dengan  $\tau$  merupakan parameter bentuk dan  $\pi(\mathbf{x}_i)$  adalah peluang  $x$  diklasifikasikan dalam kelompok bank yang mengalami *financial distress*. *Link function* pada GEVR dinamakan dengan *gevit* yang diberikan pada persamaan di bawah ini (Calabrese & Giudici, 2015):

$$\text{gevit}(\pi_i) = \frac{-\ln(\pi_i)^{-\tau}-1}{\tau} = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} = \eta_i. \quad (2.37)$$

## 2.9 Estimasi Parameter Generalized Extreme Value Regression

Estimasi parameter yang dilakukan untuk metode *Generalized Extreme Value Regression* dilakukan dengan memaksimumkan fungsi *likelihood*. Fungsi probabilitas yang digunakan untuk mendapatkan fungsi *likelihood* adalah:

$$f(y_i) = \pi(\mathbf{x}_i)^{y_i} (1 - \pi(\mathbf{x}_i))^{1-y_i}, \quad (2.38)$$

dengan

$$\pi(\mathbf{x}_i) = \exp \left\{ - [1 + \tau(\boldsymbol{\beta}' \mathbf{x}_i)]^{-\frac{1}{\tau}} \right\}, \quad (2.39)$$

Sehingga fungsi *likelihood* untuk metode GEVR dapat dituliskan melalui persamaan berikut:

$$\begin{aligned}
l(\boldsymbol{\beta}, \tau) &= \prod_{i=1}^n f(y_i) = \prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i} \\
&= \prod_{i=1}^n \exp \left\{ -[1 + \tau(\boldsymbol{\beta}' \mathbf{x}_i)]^{-\frac{1}{\tau}} \right\}^{y_i} \left( 1 - \exp \left\{ -[1 + \right. \right. \\
&\quad \left. \left. \tau(\boldsymbol{\beta}' \mathbf{x}_i)]^{-\frac{1}{\tau}} \right\} \right)^{1-y_i}. \tag{2.40}
\end{aligned}$$

Berdasarkan fungsi *likelihood* yang telah didapatkan, maka akan diperoleh fungsi *log-likelihood* berikut:

$$\begin{aligned}
\ln(l(\boldsymbol{\beta}, \tau)) &= \sum_{i=1}^n \left\{ -y_i [1 + \tau(\boldsymbol{\beta}' \mathbf{x}_i)]^{-\frac{1}{\tau}} + (1 - y_i) \ln \left[ 1 - \right. \right. \\
&\quad \left. \left. \exp \left[ -[1 + \tau(\boldsymbol{\beta}' \mathbf{x}_i)]^{-\frac{1}{\tau}} \right] \right] \right\}. \tag{2.41}
\end{aligned}$$

Invers dari persamaan tersebut merupakan CDF yang hanya berlaku pada nilai  $\{\mathbf{x}_i : 1 + \tau \mathbf{x}_i > 0\}$ , persamaan tersebut hanya berlaku pada  $\{\mathbf{x}_i : 1 + \tau \mathbf{x}_i > 0\}$ . *Score function* didapatkan dengan melakukan *differencing* pada fungsi *log-likelihood* dengan parameter  $\boldsymbol{\beta}$  dan  $\tau$ .

$$\frac{\partial l(\boldsymbol{\beta}, \tau)}{\partial \beta_j} = -\sum_{i=1}^n x_{ij} \frac{\ln[\pi(x_i)] y_i - \pi(x_i)}{1 + \tau \boldsymbol{\beta}' \mathbf{x}_i (1 - \pi(x_i))}, \tag{2.42}$$

dengan  $j = 0, 1, \dots, p$  serta

$$\begin{aligned}
\frac{\partial l(\boldsymbol{\beta}, \tau)}{\partial \tau} &= \sum_{i=1}^n \left[ \frac{1}{\tau^2} \ln(1 + \tau \boldsymbol{\beta}' \mathbf{x}_i) - \right. \\
&\quad \left. \frac{\boldsymbol{\beta}' \mathbf{x}_i}{\tau(1 + \tau \boldsymbol{\beta}' \mathbf{x}_i)} \right] \frac{y_i - \pi(x_i)}{1 - \pi(x_i)} \ln[\pi(x_i)]. \tag{2.43}
\end{aligned}$$

Kadangkala, metode maksimum *likelihood* tidak memberikan hasil akhir yang *closed form* sehingga perl iterasi numerik dibutuhkan dalam melakukan estimasi parameter. *Initial value* yang disarankan menurut Calabrese dan Osmetti (2013) untuk parameter dilakukan perhitungan dengan persamaan *log-likelihood* dari distribusi Gumbel. Peluang perusahaan diklasifikasikan kedalam kategori yang mengalami *financial distress* berubah menjadi berikut:

$$\pi(x_i) = \exp(-\exp(\boldsymbol{\beta}' \mathbf{x}_i)), \tag{2.44}$$

Sehingga *log-likelihood* dari distribusi Gumbel adalah sebagai berikut:

$$l(\boldsymbol{\beta}) = \sum_{i=1}^n \{y_i[-\exp(\boldsymbol{\beta}' \mathbf{x}_i)] + (1 - y_i)\ln[1 - \exp[-\exp(\boldsymbol{\beta}' \mathbf{x}_i)]]\}. \quad (2.45)$$

Estimasi parameter kemudian dilanjutkan dengan iterasi numerik.

## 2.10 Regresi Logistik

Regresi logistik adalah metode pemodelan dengan respon berskala nominal atau ordinal. Pada penelitian ini tipe data respon adalah biner dimana hanya terdapat dua nilai (dikotomi) yaitu sukses yang dinotasikan sebagai 1 dan gagal dinotasikan dengan 0 dan prediktor bersifat polikotomus (Hosmer & Lemshow, 2000). Oleh sebab itu respon ini disebut memiliki distribusi Bernoulli. Fungsi probabilitas untuk setiap observasi diberikan pada persamaan berikut (Agresti, 2002).

$$f(y) = \pi^y(1 - \pi)^{1-y}; y = 0, 1, \quad (2.46)$$

dimana jika  $y = 0$  maka  $P(y = 0) = (1 - \pi)$  dan jika  $y = 1$  maka  $P(y = 1) = \pi$ . Model regresi logistik adalah *Generalized Linear Model* dengan komponen random binomial dan *logit link function*. Kelebihan dari model regresi logistik adalah banyaknya informasi yang dapat diambil. Salah satu informasi penting yang sering digunakan untuk interpretasi model adalah *odds ratio*. Berikut adalah persamaan yang digunakan untuk menghitung nilai *odds ratio* (Johnson & Wichern, 2007):

$$\text{odds ratio} = \frac{\pi(x)}{1-\pi(x)}, \quad (2.39)$$

dengan  $\pi(x)$  adalah probabilitas suatu perusahaan sektor industri dinyatakan mengalami *financial distress*. Probabilitas perusahaan sektor industri mengalami *financial distress* untuk perusahaan sektor industri ke- $i$  yaitu didapat jika probabilitas yang dihasilkan lebih besar dari 0.5 dirumuskan sebagai berikut (Härdle, dkk., 2014).

$$\pi(\mathbf{x}_i) = P(y_i = 1|\mathbf{x}_i) = \frac{\exp(\alpha + \mathbf{x}_i' \boldsymbol{\beta})}{1 + \exp(\alpha + \mathbf{x}_i' \boldsymbol{\beta})}, \quad (2.40)$$

dan probabilitas perusahaan sektor industri tidak mengalami *financial distress* adalah sebagai berikut:

$$P(y_i = 0|\mathbf{x}_i) = 1 - P(y_i = 1|\mathbf{x}_i) = \frac{1}{1 + \exp(\mathbf{x}_i' \boldsymbol{\beta})}. \quad (2.41)$$

*Link Function* yang digunakan untuk menemukan model regresi logistik adalah sebagai berikut:

$$\frac{\pi(x)}{1-\pi(x)} = \exp(\alpha + \mathbf{x}'_i \boldsymbol{\beta}), \quad (2.42)$$

maka

$$g(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \mathbf{x}'_i \boldsymbol{\beta}, \quad (2.43)$$

dimana *log* dari persamaan tersebut memiliki hubungan yang linier  $\log \frac{\pi(x)}{1-\pi(x)} = \alpha + \mathbf{x}'_i \boldsymbol{\beta}$ . Oleh karena itu, *link function* yang digunakan adalah transformasi dari *log* persamaan yang disebut dengan logit (Agresti, 2002).

## 2.11 Estimasi Parameter *Logistic Regression*

Metode yang digunakan dalam menaksir parameter *Logistic Regression* adalah *Maximum Likelihood Ratio* (MLE). Fungsi *likelihood* diberikan pada persamaan berikut:

$$l(\boldsymbol{\beta}) = \prod_{i=1}^n f(y_i) = \prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i}, \quad (2.55)$$

dengan  $n$  merupakan jumlah dari observasi,  $x_i$  adalah nilai prediktor ke- $i$ , dan , dan  $y_i$  merupakan respon untuk observasi ke- $i$  (Agresti, 2002). Langkah selanjutnya adalah merubah fungsi *likelihood* ke dalam bentuk *log-likelihood* yaitu:

$$\begin{aligned} \ln(l(\boldsymbol{\beta})) &= \ln\left[\prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i}\right] \\ &= \sum_{i=0}^n (y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]). \\ &= \sum_{j=0}^p (\sum_{i=1}^n y_i x_{ij}) \beta_j - \sum_{i=1}^n \ln\left[1 + \exp\left(\sum_{j=1}^p \beta_j x_{ij}\right)\right]. \end{aligned} \quad (2.56)$$

Nilai estimasi untuk parameter  $\hat{\boldsymbol{\beta}}$ , didapatkan melalui turunan pertama fungsi *likelihood* yang kemudian disamadengankan nol. Berikut merupakan persamaan turun pertama fungsi *likelihood*:

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^n y_i x_{ij} - \sum_{i=1}^n x_{ij} \left( \frac{e^{\sum_{j=0}^p \beta_j x_{ij}}}{1 + e^{\sum_{j=0}^p \beta_j x_{ij}}} \right) = 0. \quad (2.57)$$

Sehingga,

$$\sum_{i=1}^n y_i x_{ij} - \sum_{i=1}^n x_{ij} \hat{\pi}(x_i) = 0. \quad (2.58)$$

Estimasi varians kovarians dikembangkan melalui teori MLE dari koefisien parameternya yang didapatkan dari turunan kedua  $l(\boldsymbol{\beta})$ .

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_j \partial \beta'_j} = \sum_{i=1}^n \sum_{j=1}^p x_{ij} x'_{ij} \pi(x_i)(1 - \pi(x_i)) \quad (2.59)$$

Hasil persamaan yang tidak *closed form* seringkali didapatkan ketika estimasi dilakukan menggunakan metode MLE. Salah satu alat untuk mengatasi masalah tersebut adalah iterasi numerik. Dari beberapa metode iterasi numerik, metode *Newton Raphson* merupakan metode yang sering digunakan. Estimasi parameter dengan metode *Newton Raphson* melibatkan matriks Hessian dalam perhitungannya. Berikut merupakan persamaan yang digunakan untuk menghitung matriks Hessian:

$$\mathbf{u}' = \left( \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_0}, \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_1}, \dots, \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_p} \right), \quad (2.60)$$

dengan

$$\mathbf{u}_j^{(t)} = \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^n (y_i - u_j^{(t)}) x_{ij}. \quad (2.61)$$

Matriks Hessian dapat dihitung melalui turunan kedua dari persamaan *log-likelihood* berikut:

$$\mathbf{H} = \begin{bmatrix} h_{00} & \cdots & h_{0p} \\ \vdots & \ddots & \vdots \\ h_{p0} & \cdots & h_{pp} \end{bmatrix}, h_{ij} = \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_i}. \quad (2.62)$$

Matriks Hessian dengan iterasi sebanyak  $n$  dapat ditulis menjadi:

$$H(\boldsymbol{\beta}^{(t)}) = -\{\mathbf{x}' \text{diag}[\pi(\mathbf{x}_1)^{(t)}(1 - (\mathbf{x}_1)^{(t)}), \dots, \pi(\mathbf{x}_n)^{(t)}(1 - (\mathbf{x}_n)^{(t)})]\mathbf{x}\}^{-1} \quad (2.63)$$

Sehingga didapatkan estimasi sebagai berikut:

$$\hat{\boldsymbol{\beta}}^{(t+1)} = \hat{\boldsymbol{\beta}}^{(t)} + \{\mathbf{x}' \text{diag}[\pi(\mathbf{x}_1)^t(1 - (\mathbf{x}_1)^{(t)}), \dots, \pi(\mathbf{x}_n)^{(t)}(1 - (\mathbf{x}_n)^{(t)})]\mathbf{x}\}^{-1} \mathbf{x}' (\mathbf{y} - \pi(\mathbf{x}_n)^{(t)}) \quad (2.64)$$

Langkah-langkah yang digunakan dalam iterasi numerik menggunakan *Newton Raphson* adalah sebagai berikut:

1. Melakukan estimasi parameter  $\boldsymbol{\beta}$  dengan menggunakan metode *Ordinary Least Square* (OLS) yang menghasilkan nilai parameter awal  $\boldsymbol{\beta}^0$  sehingga diperoleh  $\pi(x_i)^0$ .
2. Nilai  $\pi(x_i)^0$  yang telah diperoleh digunakan untuk mendapatkan matriks Hessian  $\mathbf{H}^0$  dan  $\mathbf{u}^0$ .
3. Melakukan iterasi hingga didapatkan nilai  $\boldsymbol{\beta}$  yang konvergen, dengan nilai  $\boldsymbol{\beta}$  pada iterasi ke- $(t + 1)$  adalah  $\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^t - (\mathbf{H}^{(t)})^{-1} \mathbf{u}^{(t)}$ .

## 2.12 Regularized Logistic Regression

Dalam penelitian yang dilakukan oleh Härdle dan Prastyo (2004) menyatakan bahwa probabilitas pelanggan ke- $i$  mengalami gagal bayar oleh prediktor gagal bayar diberikan oleh persamaan berikut.

$$\begin{aligned} P(y_i = 1 | \mathbf{x}_i) &= \frac{e^{\beta_0 + \mathbf{x}'_i \beta}}{1 + e^{\beta_0 + \mathbf{x}'_i \beta}} \\ &= \frac{1}{1 + e^{-(\beta_0 + \mathbf{x}'_i \beta)}}, \end{aligned} \quad (2.44)$$

dan

$$\begin{aligned} P(y_i = 0 | \mathbf{x}_i) &= 1 - P(y_i = 1 | \mathbf{x}_i) \\ &= \frac{1}{1 + e^{(\beta_0 + \mathbf{x}'_i \beta)}}. \end{aligned} \quad (2.45)$$

Log odds ratio model regresi linier dimana,

$$\log \left\{ \frac{P(y_1 = 1 | \mathbf{x}_i)}{P(y_1 = 0 | \mathbf{x}_i)} \right\} = \beta_0 + \mathbf{x}'_i \beta. \quad (2.46)$$

Dilanjutkan dengan memaksimalkan fungsi *regularized log-likelihood* dengan persamaan dibawah ini yaitu:

$$\max_{\beta, \beta_0} \{\ell(\beta, \beta_0) - \lambda R_\gamma(\beta)\}, \quad (2.47)$$

dengan  $R_\gamma(\beta)$  adalah fungsi regualisasi dari fungsi *likelihood* yang diberikan pada persamaan berikut.

$$\begin{aligned} \ell(\beta, \beta_0) &= n^{-1} \sum_{i=1}^n [y_i \log P(y_i = 1 | \mathbf{x}_i) + (1 - y_i) \log \{1 - P(y_i = 1 | \mathbf{x}_i)\}], \\ &= n^{-1} \sum_{i=1}^n \{I(y_i = 1) \log P(y_i = 1 | \mathbf{x}_i) + I(y_i = 0) \log P(y_i = 0 | \mathbf{x}_i)\}, \\ &= n^{-1} \sum_{i=0}^n I(y_i = 1) \left[ (\beta_0 + \mathbf{x}'_i \beta) - \log \left\{ 1 + e^{\beta_0 + \mathbf{x}'_i \beta} \right\} \right]. \end{aligned} \quad (2.48)$$

Persamaan (2.44) dimaksimalkan menggunakan algoritma penurunan koordinat titik. Langkah-langkah dalam algoritma penurunan koordinat siklik yaitu lingkaran luar, tengah, dan dalam. Nilai  $\lambda$  diatur dan menghasilkan  $(\tilde{\beta}_0, \tilde{\beta})$  pada loop luar. Estimasi kuadratik (ekspansi Taylor) dari fungsi *log-likelihood* diperbarui pada loop tengah sehingga didapatkan persamaan berikut:

$$\ell_Q(\beta_0, \beta) = -\frac{1}{2n} \sum_{i=1}^n \omega_i (z_i - \beta_0 - \mathbf{x}'_i \beta)^2 + C(\tilde{\beta}_0, \tilde{\beta}), \quad (2.49)$$

dengan respon

$$z_i = \tilde{\beta}_0 + \mathbf{x}'_i \tilde{\beta} + \frac{y_i - \tilde{P}(y_i=1|x_i)}{\tilde{P}(y_i=1|x_i)\tilde{P}(y_i=0|x_i)}, \quad (2.50)$$

dan bobot sebagai berikut:

$$\omega_i = \tilde{P}(y_i = 1|\mathbf{x}_i)\tilde{P}(y_i = 0|\mathbf{x}_i), \quad (2.51)$$

dimana  $\tilde{P}(y_i = 1|\mathbf{x}_i)$  dan  $\tilde{P}(y_i = 0|\mathbf{x}_i)$  dievaluasi pada estimasi yang sedang dilakukan. Sedangkan  $C(\tilde{\beta}_0, \tilde{\beta})$  adalah konstan.  $\ell_Q(\beta_0, \beta)$  pada awalnya diartikan sebagai kuadrat terkecil yang terboboti. Sementara itu algoritma penurunan koordinat digunakan untuk memecahkan masalah kuadrat terkecil tertimbang (PWLS) pada loop dalam adalah sebagai berikut:

$$\min_{\beta_0, \beta} \{-\ell_Q(\beta_0, \beta) + \lambda R_\gamma(\beta)\}. \quad (2.52)$$

Setiap penurunan loop koordinat dalam tetap berlanjut sampai perubahan maksimum pada persamaan (2.48) kurang dari *threshold* yang sangat kecil simana nilai *threshold* yang digunakan dalam penelitian ini adalah 1E-7. Langkah selanjutnya adalah mengurangi nilai  $\lambda$  dan mengulangi ketiga loop sampai mendapatkan persamaan yang konvergen.

### **2.12.1 Lasso Absolute Shrinkage and Selection Logistic Regression**

Konsep *Lasso Regression* adalah apabila terdapat koefisien dari beberapa variabel kurang berkontribusi maka akan dipaksa tepat nol (James, dkk., 2013). Pada model akhir metode ini hanya menggunakan variabel yang signifikan. *Regularization* untuk regresi logistik diberikan pada persamaan berikut.

$$\min_{\beta_0, \beta} \sum_{i=1}^n L\{y_i, f(x_i) + \lambda(\gamma\|\beta\|_1)\}, \quad (2.53)$$

dengan  $\gamma = 1$  dan  $\lambda > 0$  adalah *tunning parameter*.

### **2.12.2 Elastic-Net Logistic Regression**

*Elastic-Net* pada model regresi logistik merupakan kombinasi antara *Ridge* dan *Lasso Regression* (Friedman, dkk. 2010). Konsep metode ini adalah menyusutkan beberapa koefisien ke nol (*Ridge Regression*) dan mengatur beberapa koefisien ke nol (*Lasso Regression*).

$$\min_{\beta_0, \beta} \sum_{i=1}^n L\left\{y_i, f(x_i) + \lambda \left[ (1-\gamma) \frac{1}{2} \|\beta\|_2^2 + \gamma \|\beta\|_1 \right] \right\}, \quad (2.54)$$

dengan bobot  $0 < \gamma < 1$  yang harus dioptimalkan bersamaan dengan  $\lambda > 0$  yang merupakan *tuning parameter*.

### **2.13 Stepwise Procedures**

Dalam studi eksplorasi, sebuah metode *algorithmic* dapat digunakan untuk mencari diantara model-model yang informatif. Goodman dalam Agresti (2000) mengusulkan metode *forward selection* dan *backward elimination* untuk mendapatkan model dari respon dari beberapa prediktor. *Forward selection* menambahkan variabel peubah secara berurutan hingga penambahan variabel tidak memberikan peningkatan terhadap hasil ketepatan estimasi respon. Urutan penambahan peubah dengan menambahkan peubah yang memberikan peningkatan estimasi respon terbesar. *Backward elimination* diawali dengan membuat model kompleks lalu menghilangkan variabel peubah yang tidak berpengaruh signifikan hingga didapat model terbaik dimana variabel peubah secara keseluruhan signifikan.

*Backward elimination* lebih dianjurkan daripada *forward selection*. *Backward elimination* dianggap lebih aman dan lebih efektif untuk menghapus variabel peubah dari model kompleks daripada melakukan penambahan variabel peubah secara berurutan. Selain itu pada *forward selection* variabel peubah yang telah masuk tidak dapat dikeluarkan lagi dari model, dimana variabel peubah tersebut dapat tidak memberi pengaruh yang signifikan terhadap model.

### **2.14 Uji Signifikansi Parameter**

Prediktor yang memiliki pengaruh signifikan terhadap respon ditunjukkan dengan uji signifikansi parameter. Uji signifikansi parameter dilakukan secara serentak dan secara parsial. Berikut pengujian parameter yang dilakukan setelah mendapatkan model.

#### **2.14.1 Uji Serentak**

Uji serentak dapat dihitung melalui rasio *likelihood* (Hosmer & Lemeshow, 2000). Uji ini merupakan uji *chi-square* yang menggunakan nilai *maximum likelihood*. Pengujian ini bertujuan untuk memeriksa apakah prediktor berpengaruh signifikan terhadap respon. Berikut adalah hipotesis yang digunakan pada pengujian serentak.

$$H_0: \beta_0 = \beta_1 = \dots = \beta_p = 0$$

$$H_1: \text{minimal ada satu } \beta_j \neq 0, j = 0, 1, \dots, p.$$

Statistik uji:

$$G^2 = -2\ln \left[ \frac{\left( \frac{n_1}{n} \right)^{n_1} \left( \frac{n_0}{n} \right)^{n_0}}{\prod_{i=1}^n \hat{\pi}_i^{y_i} (1-\hat{\pi}_i)^{1-y_i}} \right] \quad (2.65)$$

dimana

$n_1$ : jumlah perusahaan yang mengalami *financial distress*

$n_0$ : jumlah perusahaan lainnya (sehat).

Berdasarkan hasil perhitungan secara statistik dapat diperoleh kesimpulan dari daerah penolakan hipotesis yaitu tolak  $H_0$  jika  $G^2 > X_{\alpha,p}^2$  atau  $p\text{-value} < \alpha$  (5%), yang berarti bahwa minimal terdapat satu prediktor berpengaruh signifikan terhadap respon.

## 2.14.2 Uji Parsial

Pengujian secara parsial dilakukan untuk mengatahui signifikansi masing-masing prediktor terhadap respon. Pengujian parameter secara parameter menggunakan uji *Wald* dengan hipotesis sebagai berikut:

$$H_0: \beta_j = 0$$

$$H_0: \beta_j \neq 0.$$

Statistik uji sebagai berikut.

$$W^2 = \frac{\hat{\beta}_j^2}{\text{var}(\hat{\beta}_j)} \quad (2.66)$$

Berdasarkan hasil perhitungan secara statistik dapat diperoleh kesimpulan dari daerah penolakan hipotesis yaitu tolak  $H_0$  jika  $W^2 > X_{\alpha,p}^2$  atau  $p\text{-value} < \alpha$  (5%). Sehingga didapat kesimpulan bahwa prediktor ke- $j$  berpengaruh signifikan terhadap respon.

## 2.15 Evaluasi Ketepatan Klasifikasi

Model terbaik merupakan model yang memiliki ketepatan klasifikasi paling tinggi diantara model-model lainnya. Perhitungan terhadap ketepatan klasifikasi dilakukan untuk mengevaluasi model agar mendapatkan model dengan hasil prediksi terakurat. Evaluasi ketepatan klasifikasi dapat bermacam-macam dan yang umum digunakan adalah *accuracy*, *sensitivity*, *specifity*, dan *precision*. Diantara data *training* maupun *testing* keduanya dilakukan perhitungan untuk mendapatkan seberapa kuat

hasil klasifikasinya (Han, dkk., 2011). Sebelum mengukur ketepatan klasifikasi menggunakan ukuran-ukuran yang telah disebutkan, terlebih dahulu dibuat *confussion matrix* sebagai berikut (Härdle, dkk., 2014).

**Tabel 2.1 Confussion Matrix**

<i>Predicted</i> $\hat{Y}$	<i>Sample</i> ( $\hat{Y}$ )	
	<i>Default</i> (1)	<i>Non-Default</i> (0)
<i>Default</i> (1)	<i>True Positive</i>	<i>False Negative</i>
<i>Non-Default</i> (0)	<i>False Positive</i>	<i>True Negative</i>
<i>Total</i>	<i>Positive</i>	<i>Negative</i>

Berdasarkan *confussion matrix* pada Tabel 2.1, maka dapat dihitung ukuran-ukuran ketepatan klasifikasi menggunakan rumus-rumus dibawah ini:

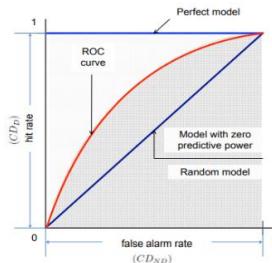
$$\text{Accuracy} = P(\hat{Y} = Y) = \frac{\text{True Positive} + \text{True Negative}}{\text{Positive} + \text{Negative}},$$

$$\text{Sensitivity} = P(\hat{Y} = 1|Y = 1) = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}},$$

$$\text{Specificity} = P(\hat{Y} = 0|Y = 0) = \frac{\text{True Negative}}{\text{True Negarive} + \text{False Positive}},$$

$$\begin{aligned} \text{Precision} &= \frac{(\hat{Y} = 1|Y = 1)}{(\hat{Y} = 1|Y = 1) + (\hat{Y} = 1|Y = 0)} \\ &= \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}. \end{aligned}$$

Keempat nilai yang dapat menunjukkan ketepatan klasifikasi di atas memiliki kekurangan yaitu kurang baik digunakan pada data yang *unbalanced*. Solusi yang dapat digunakan untuk mengatasi ketepatan klasifikasi pada data *unbalanced* adalah menggunakan kurva ROC (*Receiver Operating Characteristic*). Kurva ROC pada model tertentu mnunjukkan *trade-off* antara *true positive rate* (TPR) dengan *false positive rate* (FPR). Kurva ROC menggambarkan *trade-off* antara tingkat dimana model dapat mengidentifikasi secara tepat kasus yang positif dan secara keliru mengidenifikasi kasus negatif. Area di bawah ROC (AUC) adalah ukuran secara nilai yang digunakan untuk menunjukkan keakuratan hasil klasifikasi (Han, dkk., 2011). Semakin besar nilai dari AUC maka keakuratan prediksi dinilai semakin besar. Sedangkan kurva ROC digambarkan pada Gambar 2.6.



**Gambar 2.6** Kurva ROC (Sumber: Härdle, dkk., 2014)

AUC adalah area terarsir dibawah kurva ROC. Ukuran AUC digunakan untuk merangkum kurva ROC ke dalam suatu nilai dimana semakin besar AUC maka dapat dikatakan model semakin baik (Witten, dkk., 2011). Persamaan untuk mendapatkan nilai AUC ditunjukkan pada rumus dibawah ini:

$$AUC = \frac{Sensitivity + Specificity}{2}.$$

AUC diperoleh berdasarkan nilai *sensitivity* dan *specificity*, yang merupakan probabilitas suatu kelas diklasifikasikan pada kelas yang tepat. Sehingga AUC lebih *robust* daripada *Accuracy* secara total. Evaluasi ketepatan klasifikasi berdasarkan nilai AUC diberikan pada Tabel 2.2 berikut (Gorunescu, 2011).

**Tabel 2.2** Kategori Kebaikan Model Berdasarkan AUC

Nilai AUC	Kategori
0,9-1	<i>Excellent classification</i>
0,8-0,9	<i>Good classification</i>
0,7-0,8	<i>Fair classification</i>
0,6-0,7	<i>Poor classification</i>
<0,6	<i>Failure</i>

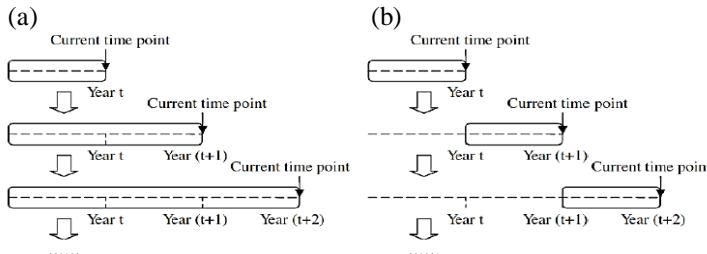
## 2.16 Konsep Drift

Konsep *drift* sering diartikan sebagai perbaian konsep target yang disebabkan oleh perubahan *hidden context*. Perubahan tersebut dapat terjadi secara langsung maupun tidak langsung. Problem ini dapat diatasi menggunakan *window size* pada data *training*. Secara khusus, konsep ini memperhatikan pengaruh kejadian pada waktu  $t$  terhadap kejadian yang terjadi pada waktu selanjutnya. Misal  $z = (x, y)$  terdiri atas variabel prediktor  $x \in \mathbb{R}^P$  dan variabel  $y = \{0,1\}$ . Data untuk seiap periode

dinamakan dengan *batch*, dimana pada masing-masing *batch* terdiri dari  $m$  sampel (Klinkenberg, 2013):

$$z_{(1,1)}, \dots, z_{(1,m)}, z_{(2,1)}, \dots, z_{(2,m)}, \dots, z_{(t,1)}, \dots, z_{(t,m)}, z_{(t+1,1)}, \dots, \\ z_{(t+1,m)}$$

dimana  $z_{i,j}$  merupakan sampel ke- $j$  pada *batch* ke- $i$ . Masing-masing *batch* bersifat independen, identik dan mengikuti distribusi  $Pri(x, y)$ . Jenis *time window* yang digunakan dalam konsep *drift* diantanya yaitu *full memory time window* dan *no memory time window*. Pada Gambar 2.7 di bawah ini akan digambarkan konsep dari *time window* menggunakan *full memory time window* dan *no memory time window*.



**Gambar 2.7** (a) *Full Memory Time Window*; (b) *No Memory Time Window*  
(Sumber: Klinkenberg, 2013)

*Full memory window size* mempertimbangkan data pada waktu sebelumnya untuk memprediksi kejadian pada waktu tersebut. Maksud dari pernyataan tersebut adalah *fill memory window size* melibatkan data baru dan data sebelumnya. Namun model ini memiliki kelemahan yaitu kurangnya kemampuan model beradaptasi dengan kondisi baru karena model masih mengandung konsep kondisi pada waktu sebelumnya. Di sisi lain, model *no memory window size* hanya mengasumsikan bahwa kondisi sebelumnya tidak memiliki hubungan dengan kondisi saat ini. Kelemahan dari model ini adalah kurangnya generalisasi pada jumlah data yang kecil untuk data *training* ketika model harus stabil.

## 2.17 Rasio Keuangan

Rasio keuangan adalah alat yang dijadikan perusahaan untuk menganalisis dan memantau kondisi keuangan oleh suatu perusahaan. Rasio keuangan sendiri dapat dilihat melalui laporan

keuangan yang biasanya diterbitkan tahunan (Firdaus, 2012). Rasio keuangan akan dijadikan calon investor dan debitor untuk mengambil keputusan atau pertimbangan mengenai prospek suatu perusahaan dimasa yang akan datang. Selain itu fungsi lain dari rasio keuangan adalah untuk menilai peluang dan resiko yang akan dihadapi perusahaan. Berikut merupakan beberapa jenis rasio keuangan yang digunakan dalam penelitian ini.

### 1. Rasio Aktivitas

Rasio aktivitas merupakan alat yang digunakan untuk mengukur seberapa efektif perusahaan memanfaatkan aset-aset yang dimiliki. Rasio aktivitas melibatkan perbandingan antara tingkat penjualan dengan investasi pada berbagai jenis aktiva diantaranya yaitu: persediaan, piutang, aktiva tetap, dan aktiva lain (Sawir, 2005). Secara umum rasio aktivitas dikelompokkan menjadi aktivitas jangka pendek (*short-term activity*) dan aktivitas jangka panjang (*long-term activity*). Variabel yang digunakan dalam penelitian ini yang termasuk dalam jenis rasio aktivitas adalah sebagai berikut:

- a. *Earning Before Interest and Tax to Total Asset*
- b. *Sales to Total Asset*
- c. *Inventory Turnover Ratio*
- d. *Day's Sales in Inventory Ratio*

### 2. Rasio Profitabilitas

Rasio profitabilitas digunakan untuk melihat gambaran efektivitas pengelolaan perusahaan karena profitabilitas merupakan hasil akhir bersih dari berbagai kebijakan yang diputuskan oleh perusahaan (Sawir, 2005). Beberapa ukuran rasio profitabilitas yang digunakan pada penelitian sebagai variabel penelitian adalah:

- a. *Return on Equity*
- b. *Gross Profit Margin*
- c. *Net Profit Margin*
- d. *Operating Profit Margin*

### 3. Rasio Solvabilitas

Rasio solvabilitas adalah rasio yang dapat menggambarkan kemampuan perusahaan dalam memenuhi kewajibaban jangka panjangnya. Berdasarkan pandangan pemegang saham dengan

meingkatnya rasio solvabilitas akan mengakibatkan kenaikan pembayaran bunga yang tinggi dan pada akhirnya akan mengurangi pembayaran deviden (Prihadi, 2010). Rasio solvabilitas yang digunakan menjadi variabel penelitian adalah sebagai berikut:

- a. *Debt to Equity Ratio*
- b. *Debt to Asset Ratio*
- 4. Rasio Likuiditas

Jika rasio solvabilitas adalah alat untuk mengukur kemampuan perusahaan untuk memenuhi kewajiban jangka panjang, maka rasio likuiditas adalah alat yang digunakan untuk mengukur kemampuan perusahaan dalam memenuhi kewajiban jangka pendek. Rasio likuiditas juga dapat dikatakan sebagai kemampuan perusahaan untuk memenuhi kewajiban dalam hal keuangan pada saat jatuh (Prihadi, 2010). Variabel penelitian yang termasuk dalam rasio likuiditas adalah sebagai berikut:

- a. *Current Ratio*
- b. *Quick Ratio*
- c. *Earning to Debt*
- d. *Working Capital to Total Asset*
- e. *Working Capital to Sales*

## **2.18 Financial Distress**

Platt mendefinisikan *financial distress* sebagai tahap penurunan kondisi keuangan yang terjadi sebelum terjadinya kebangkrutan ataupun likuidasi. Selain itu *financial distress* dapat membawa suatu perusahaan megalami kegagalan pembayaran (*default*), tidak sesuai dengan kontrak yang telah disepakati (Argo, 2010). Definisi lain dari *financial distress* disebutkan oleh Hosfer yang mengumpamakan kondisi *financial distress* sebagai suatu kondisi dari perusahaan yang megalami laba bersih negatif selama beberapa tahun.

*Financial distress* dapat membawa suatu perusahaan mengalami kegagalan pembayaran (*default*), tidak sesuai dengan kontrak yang telah disepakati. Kegagalan pembayaran tersebut mendorong debitur untuk mencari penyelesaian dengan pihak kreditur, yang pada akhirnya dapat dilakukan restrukturisasi keuangan antara perusahaan, kreditor, dan investor (Ross, dkk.,

2002). Perusahaan yang mengalami *financial distress* akan menghadapi kondisi: a) tidak mampu memenuhi jadwal atau kegagalan pembayaran kembali hutang yang sudah jatuh tempo kepada kreditor; b) perusahaan dalam kondisi tidak *solvable (insolvency)*.

Asquith, dkk. (1994) mendefinisikan pengukuran perusahaan mengalami *financial distress* atau tidak menggunakan *interest coverage ratio* (ICR). ICR adalah rasio keuangan yang menjelaskan seberapa baik perusahaan dapat membayar bunga pada hutang yang belum dibayarkan. ICR sering digunakan oleh pemodal, investor, dan kreditor untuk melihat resiko bila melakukan investasi pada perusahaan tersebut. ICR didapatkan melalui persamaan berikut:

$$\text{ICR} = \frac{\text{EBIT}}{\text{Interest Expense}},$$

dengan EBIT adalah *Earnings Before Interest and Tax* yaitu pendapatan sebelum adanya bunga dan pajak dan *Interest Expense* adalah beban bunga yang dijatuhkan kepada perusahaan. Dalam penelitian ini perusahaan yang mengalami *financial distress* diindikasikan menggunakan ICR yaitu kurang dari 1 (satu). Selain menggunakan ICR, penelitian ini juga mendefinisikan bahwa perusahaan yang memiliki *Return On Asset* (ROA) negatif merupakan perusahaan yang mengalami *financial distress*. ROA digunakan untuk mengukur kemampuan perusahaan dalam mendayagunakan aset untuk memperoleh laba dan mengukur hasil total untuk seluruh penyedia sumber dana. Nilai ROA yang baik yaitu konstan di atas 20%, semakin besar nilai ROA maka kemampuan perusahaan dalam mendayagunakan asetnya semakin baik sehingga peluang perusahaan mengalami *financial distress* semakin kecil (Sawir, 2005). Rumus yang digunakan untuk mendapatkan ROA adalah sebagai berikut:

$$\text{ROA} = \frac{\text{Net Income}}{\text{Total Asset}}.$$

## 2.19 Indikator Ekonomi Makro

Ekonomi makro merupakan gambaran perubahan ekonomi yang dapat mempengaruhi pola perubahan dalam masyarakat. Hasil kajian dalam ruang lingkup ekonomi makro dapat digunakan untuk menganalisis target-target kebijakan ekonomi yang

dijalankan oleh pemerintah, salah satunya untuk membantu pengembangan dan evaluasi kebijakan ekonomi dan strategi bisnis. Kondisi perikonomian negara dapat mempengaruhi kinerja perusahaan dalam beroperasi, khususnya perusahaan sektor industri yang memberikan kontribusi terbesar untuk nilai ekspor Indonesia sehingga dalam penelitian ini melibatkan indikator ekonomi makro (IDX, 2016). Indikator ekonomi makro yang digunakan adalah *Real GDP Growth*, *BI 7-Day Repo Rate*, *USD/IDR*, dan Inflasi.

*Real GDP Growth* adalah alat ukur pertumbuhan ekonomi menggunakan Produk Domestik Bruto (PDB) dari suatu tahun ke tahun lainnya yang disesuaikan dengan inflasi dan dinyatakan secara riil yang dinyatakan dalam persentase. *BI 7-Day Repo Rate* adalah kebijakan baru yang dilakukan oleh Bank Indonesia untuk menggantikan *BI-Rate* sejak 19 Agustus 2016. *USD/IDR* merupakan nilai tukar rupiah terhadap Dolar AS. Sedangkan inflasi adalah suatu keadaan perekonomian suatu wilayah atau negara dimana terjadi kecenderungan kenaikan harga-harga barang secara terus menerus yang biasanya diukur dengan menghitung perubahan tingkat persentas perubahan sebuah indeks harga.

### **BAB III**

### **METODOLOGI PENELITIAN**

#### **3.1 Sumber Data**

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari laporan keuangan oleh Bank Indonesia dari tahun 2005 hingga 2018. Data penelitian ini merupakan hasil pengolahan dari suatu nilai keuangan perusahaan menjadi rasio-rasio keuangan perusahaan. Data yang akan digunakan terdiri dari 80% data *training* dan 20% data *testing*.

Variabel yang digunakan dalam penelitian ini meliputi respon dan prediktor. Variabel yang mewakili respon adalah variabel ICR dan ROA. Respon dikelompokkan menjadi 2 yaitu kelompok perusahaan sehat dan kelompok perusahaan mengalami *financial distress*. Jumlah prediktor yang digunakan terdiri dari 14 variabel rasio keuangan perusahaan, 4 variabel indikator makro ekonomi. Tabel 3.1 merupakan tabel yang berisi variabel-variabel yang digunakan dalam penelitian ini.

**Tabel 3.1** Variabel Penelitian

Notasi	Varibel	Keterangan
Y	ICR atau ROA	1: Perusahaan <i>financial distress</i> 0: Perusahaan sehat
X <sub>1</sub>	EBITA	<i>Earning Before Income Tax to Asset</i>
X <sub>2</sub>	STA	<i>Sales to Total Asset</i>
X <sub>3</sub>	ITR	<i>Inventory Turnover Ratio</i>
X <sub>4</sub>	DSIR	<i>Day's Sales in Inventory Ratio</i>
X <sub>5</sub>	ROE	<i>Return on Equity</i>
X <sub>6</sub>	NPM	<i>Net Profit Margin</i>
X <sub>7</sub>	OPM	<i>Operating Profit Margin</i>
X <sub>8</sub>	DER	<i>Debt to Equity</i>
X <sub>9</sub>	DAR	<i>Debt to Asset</i>
X <sub>10</sub>	CR	<i>Current Ratio</i>
X <sub>11</sub>	QR	<i>Quick Ratio</i>
X <sub>12</sub>	ETD	<i>Earning to Debt</i>
X <sub>13</sub>	WCA	<i>Working Capital to Total Asset</i>
X <sub>14</sub>	WCS	<i>Working Capital to Sales Ratio</i>
X <sub>15</sub>	rgdpq	<i>Real GDP Growth (%yoy)</i>
X <sub>16</sub>	bi7dr	<i>BI 7-Day Repo Rate</i>
X <sub>17</sub>	USD/IDR	USD/IDR
X <sub>18</sub>	Inflasi	Inflasi

Adapun deskripsi dari masing-masing variabel yang telah disebutkan dalam uraian-uraian di bawah ini.

### 1. *Earning Before Interest and Tax to Total Asset*

EBITA digunakan untuk mengukur kemampuan perusahaan dalam mengelola sumber dayanya secara efektif dengan melihat dari hasil penjualan dan investasinya. Semakin besar penambahan tiap satuan nilai EBITA maka perubahan peluang perusahaan mengalami *financial distress* semakin kecil (Sawir, 2005).

$$\text{EBITA} = \frac{\text{Earning Before Interest and Tax}}{\text{Total Asset}}$$

### 2. *Sales to Total Asset*

STA merupakan rasio keuangan standar yang menggambarkan kemampuan aset perusahaan dalam menghasilkan penjualan (Altman, 1968). Semakin besar penambahan tiap satuan nilai STA maka perubahan peluang perusahaan mengalami *financial distress* semakin kecil. Nilai STA dapat dihitung menggunakan fungsi berikut.

$$\text{STA} = \frac{\text{Sales}}{\text{Total Asset}}$$

### 3. *Inventory Turnover Ratio*

*Inventory turnover ratio* merupakan rasio keuangan yang digunakan untuk menggambarkan likuiditas perusahaan. Rasio ini juga disebut sebagai rasio perputaran perusahaan. Semakin tinggi rasio perputaran persediaan maka semakin baik pengelolaan persediaan perusahaan (Sjam & Dharma, 2010). Untuk mendapatkan nilai dari rasio ini dapat digunakan rumus berikut.

$$\text{ITR} = \frac{\text{COGS}}{\text{Inventory}}$$

### 4. *Day's Sales in Inventory*

*Day's sales in inventory* (DSI) merupakan rasio yang mengukur waktu rata-rata *inventory* disimpan di dalam perusahaan, yaitu lamanya waktu *inventory* berada di gudang sebelum dikonversikan menjadi penjualan dan pada akhirnya menghasilkan kas. Rasio ini menunjukkan periode menahan persediaan rata-rata atau periode rata-rata persediaan barang berada di gudang (Sjam & Dharma, 2010).

$$DSI = \frac{\text{Inventory}}{\text{COGS}/365}$$

#### 5. *Return to Equity*

ROE digunakan untuk melihat sejauh mana perusahaan perusahaan mengolah modal sendiri (*net worth*) secara efektif dan mengukur tingkat keuntungan dari investasi yang telah dilakukan pemilik modal sendiri atau pemegang saham perusahaan. Semakin besar nilai ROE maka semakin baik pula kemampuan perusahaan mengolah modal sendiri sehingga peluang perusahaan mengalami *financial distress* semakin kecil (Sawir, 2005).

$$ROE = \frac{\text{Net Income}}{\text{Total Equity}}$$

#### 6. *Net Profit Margin*

NPM merupakan rasio yang menunjukkan besar persentase pendapatan bersih yang diperoleh dari setiap penjualan. Semakin besar rasio ini maka kondisi perusahaan semakin baik karena perusahaan memiliki kemampuan cukup besar dalam menghasilkan laba. NPM dapat dihitung menggunakan fungsi berikut.

$$NPM = \frac{\text{Net Profit}}{\text{Sales}}$$

#### 7. *Operating Profit Margin*

OPM digunakan sebagai indikator perusahaan dalam mencapai laba bisnis utama. Laba usaha tanpa dipengaruhi struktur modal, keputusan investasi di surat berharga (*marketable securities*) atau laba dari afiliasi (*income from affiliate*) dan tingkat pajak. Semakin besar nilai OPM maka semakin baik pencapaian laba bisnis utama perusahaan dan peluang perusahaan mengalami *financial distress* semakin kecil (Sawir, 2000).

$$OPM = \frac{\text{Earnings Before Interest and Tax}}{\text{Sales}}$$

#### 8. *Debt to Equity Ratio*

DER adalah variabel yang digunakan untuk mengukur seberapa besar modal dapat menjamin hutang sehingga dapat menggambarkan struktur modal perusahaan atau persentase dari hutang dan modal yang digunakan perusahaan. Semakin kecil nilai

DER maka semakin besar modal perusahaan yang menjamin hutang, DER merupakan bagian dari analisis kredit (Prihadi, 2010).

$$\text{DER} = \frac{\text{Total Liabilities}}{\text{Total Equity}}$$

#### 9. *Debt to Asset Ratio*

DAR merupakan variabel penelitian yang digunakan untuk mengukur kemampuan aset perusahaan untuk membayar kewajiban jangka panjang perusahaan (*solvability*). Semakin besar penambahan tiap satuan nilai DAR maka perubahan peluang perusahaan mengalami *financial distress* akan semakin kecil. DAR merupakan rasio yang termasuk dalam analisis kredit (Prihadi, 2020).

$$\text{DAR} = \frac{\text{Total Liabilities}}{\text{Total Asset}}$$

#### 10. *Current Ratio*

*Current ratio* atau CR digunakan untuk mengukur kemampuan aset lancar perusahaan untuk melunasi kewajiban jangka pendeknya. Semakin besar penambahan tiap satuan nilai CR maka perubahan peluang perusahaan mengalami *financial distress* akan semakin kecil. Rasio lancar merupakan rasio yang termasuk dalam analisis kredit (Prihadi, 2010).

$$\text{CR} = \frac{\text{Current Asset}}{\text{Current Liabilities}}$$

#### 11. *Quick Ratio*

Keterbatasan modal dan rasio lancar adalah tidak mempertimbangkan jenis aset lancar yang dimiliki persahaan dan tingkat kemudahan dalam mengubahnya menjadi kas. Oleh karena itu dimungkinkan dua perusahaan yang memiliki modal kerja dan rasio lancar yang sama tetapi memiliki kemampuan yang berbeda untuk membayar liabilitas lancarnya (Warren, dkk., 2017).

$$\text{QR} = \frac{\text{Current Asset-Inventory}}{\text{Current Liabilities}}$$

#### 12. *Earning to Debt*

ETD digunakan untuk mengukur kemampuan perusahaan dalam menjamin hutang (Altman, 1968).

$$ETD = \frac{\text{Net Income}}{\text{Total Liabilities}}$$

### 13. Working Capital to Total Asset

Rasio ini menunjukkan kemampuan perusahaan untuk menghasilkan modal kerja bersih dari seluruh total aset yang dimilikinya. Semakin tinggi nilai rasio ini maka kemampuan perusahaan untuk terus beroperasi akan semakin baik (Fitriyah & Hariyati, 2013).

$$WCA = \frac{\text{Working Capital}}{\text{Total Asset}}$$

### 14. Working Capital to Sales Ratio

Baik *current ratio* maupun *quick ratio* merupakan ukuran likuiditas jangka pendek yang paling luas digunakan, namun keduanya menghadapi masalah yang statis. Untuk menghadapi kelemahan ini, ada pendapat bahwa arus kas jangka pendek di masa depan akan menjadi indikator kemampuan membayar yang lebih baik (Sjam & Dharma, 2010). *Working capital to sales ratio* memenuhi tujuan ini pada tingkat tertentu.

$$WCS = \frac{\text{Current Assets-Current Liabilities}}{\text{Sales}}$$

Struktur data penelitian akan ditunjukkan pada Tabel 3.2. Data yang akan digunakan merupakan data kuartal IV tiap tahunnya. Berikut struktur data yang akan digunakan dalam penelitian.

Tabel 3.2 Struktur Data Penelitian

Perusahaan	Periode	t	$Y_t$	$X_{it,1}$	$X_{it,2}$	...	$X_{it,18}$
1	2005Q4	1	$Y_{11}$	$X_{11,1}$	$X_{11,2}$		$X_{11,18}$
	2006Q4	2	$Y_{12}$	$X_{12,1}$	$X_{12,2}$	..	$X_{12,18}$
	:	:	:	:	:		:
	2018Q4	t	$Y_{1t}$	$X_{1t,1}$	$X_{1t,2}$		$X_{1t,18}$
2	2005Q4	1	$Y_{21}$	$X_{21,1}$	$X_{21,2}$		$X_{21,18}$
	2006Q4	2	$Y_{22}$	$X_{22,1}$	$X_{22,2}$	..	$X_{22,18}$
	:	:	:	:	:		:
	2018Q4	t	$Y_{2t}$	$X_{2t,1}$	$X_{2t,2}$		$X_{2t,18}$

**Tabel 3.2** Struktur Data Penelitian (Lanjutan)

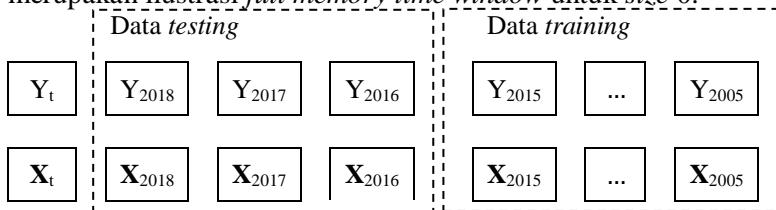
⋮		⋮	⋮	⋮	⋮	⋮	⋮
i	2005Q <sub>4</sub>	1	Y <sub>i1</sub>	X <sub>i1,1</sub>	X <sub>i1,2</sub>		X <sub>i1,18</sub>
	2006Q <sub>4</sub>	2	Y <sub>i2</sub>	X <sub>i2,1</sub>	X <sub>i2,2</sub>	⋮	X <sub>i2,18</sub>
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	2018Q <sub>4</sub>	t	Y <sub>it</sub>	X <sub>it,1</sub>	X <sub>it,2</sub>		X <sub>it,18</sub>
⋮		⋮	⋮	⋮	⋮	⋮	⋮
n	2005Q <sub>4</sub>	1	X <sub>n1,17</sub>	X <sub>n1,1</sub>	X <sub>n1,2</sub>		X <sub>n1,18</sub>
	2006Q <sub>4</sub>	2	X <sub>n2,17</sub>	X <sub>n2,1</sub>	X <sub>n2,2</sub>	⋮	X <sub>n2,18</sub>
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	2018Q <sub>4</sub>	t	X <sub>nt,17</sub>	X <sub>nt,1</sub>	X <sub>nt,2</sub>		X <sub>nt,18</sub>

Penelitian ini dilakukan menggunakan konsep *full memory window size* yaitu untuk menentukan suatu perusahaan mengalami *financial distress* atau tidak pada waktu ke-*t* diprediksi dengan rasio keuangan pada waktu ke-*t* dan waktu-waktu sebelumnya. Berikut pada tabel 3.3 akan diberikan *time window size* yang digunakan dalam penelitian.

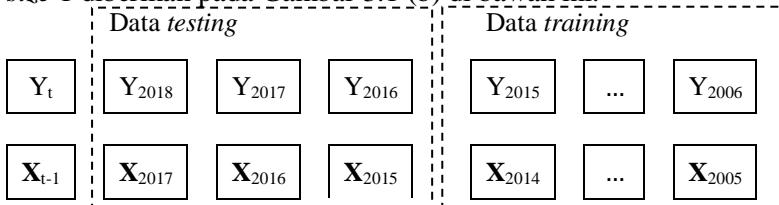
**Tabel 3.3** Time Window Size Data Penelitian

Size	Respon	Prediktor
0		$\mathbf{X}_t$
1	$Y_t$	$\mathbf{X}_{t-1}$
2		$\mathbf{X}_{t-2}$
3		$\mathbf{X}_{t-3}$

Ilustrasi penggunaan *time window* pada masing-masing *size* diberikan pada gambar-gambar di bawah ini. Gambar 3.1 (a) merupakan ilustrasi *full memory time window* untuk *size 0*.

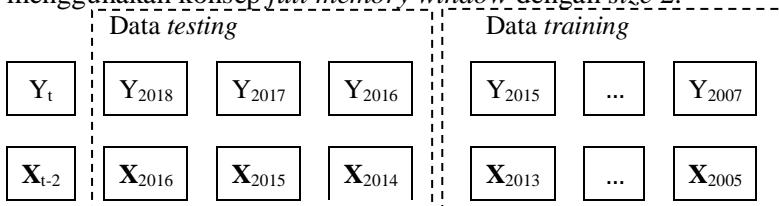
**Gambar 3.1** (a) Ilustrasi *full memory time window size 0*

Skema yang digunakan untuk *full memory time window* untuk size 1 diberikan pada Gambar 3.1 (b) di bawah ini.



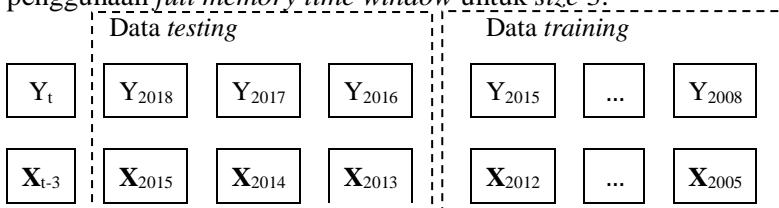
Gambar 3.1 (b) Ilustrasi *full memory time window size 1*

Gambar 3.1 (c) adalah ilustrasi yang digunakan untuk analisis menggunakan konsep *full memory window* dengan size 2.



Gambar 3.1 (c) Ilustrasi *full memory time window size 2*

Selanjutnya Gambar 3.1 (d) adalah skema atau ilustrasi dari penggunaan *full memory time window* untuk size 3.



Gambar 3.1 (d) Ilustrasi *full memory time window size 3*

### 3.4 Langkah Penelitian

Tujuan penelitian yang telah dirumuskan akan dicapai melalui langkah-langkah analisis sebagai berikut:

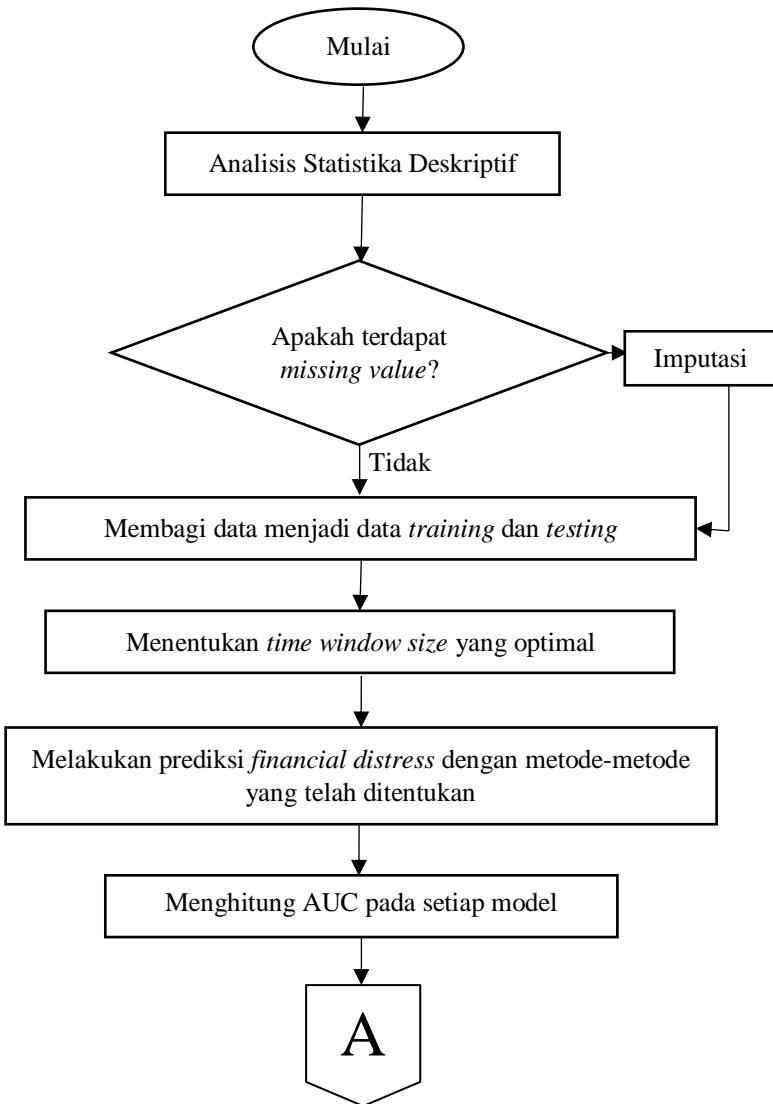
1. Melakukan pemeriksaan terhadap *missing value* pada data rasio keuangan perusahaan sektor industri di Indonesia serta melakukan imputasi *missing value* dengan metode *k-Nearest Neighbor* (kNN).
2. Melakukan analisis karakteristik rasio keuangan perusahaan yang mengalami *financial distress* dan tidak.

- a. Membuat *density plot* seluruh rasio keuangan dan variabel makroekonomi.
  - b. Melakukan analisis statistika deskriptif pada data rasio keuangan yang meliputi nilai *mean*, median, varians, dan modus.
3. Melakukan prediksi *financial distress* perusahaan industry secara serentak dan seleksi variabel.
    - a. *Support Vector Machine*
      - i. Melakukan *fitting* model SVM dengan data *train*.
      - ii. Melakukan prediksi terhadap data *train* dan *test* berdasarkan model yang didapatkan.
      - iii. Menghitung performansi model dengan nilai AUC dan *Accuracy*.
    - b. *Lasso Support Vector Machine*
      - i. Melakukan *tunning* parameter dengan menentukan nilai-nilai  $\lambda$  yang digunakan.
      - ii. Melakukan *fitting* model-model SVM dengan  $l_1$  *penalty* menggunakan data *train*.
      - iii. Melakukan prediksi terhadap data *train* berdasarkan model-model yang didapat.
      - iv. Menghitung performansi model-model dengan nilai AUC.
      - v. Mendapatkan nilai  $\lambda$  optimal dan subset variabel yang relevan berdasarkan nilai performansi metode pada data *train* paling tinggi.
      - vi. Melakukan prediksi terhadap data *test* menggunakan nilai  $\lambda$  optimum.
      - vii. Menghitung performansi model dengan nilai AUC dan *Accuracy*.
    - c. *Elastic-Net Support Vector Machine*
      - i. Melakukan *tunning* parameter dengan menentukan nilai-nilai  $\lambda_1$  dan  $\lambda_2$  yang digunakan.
      - ii. Melakukan *fitting* model-model SVM dengan *elastic-net* *penalty* menggunakan data *train*.
      - iii. Melakukan prediksi terhadap data *train* berdasarkan model-model yang didapat.

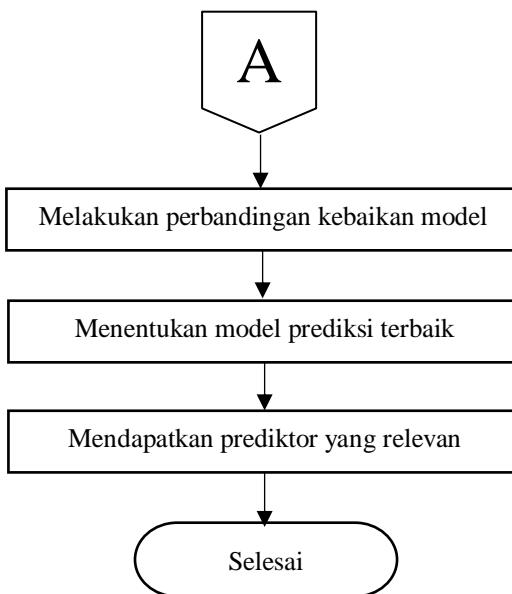
- iv. Menghitung performansi model-model dengan nilai AUC.
  - v. Mendapatkan kombinasi nilai  $\lambda_1$  dan  $\lambda_2$  optimal dan subset variabel yang relevan berdasarkan nilai performansi metode pada data *train* paling tinggi.
  - vi. Melakukan prediksi terhadap data *test* menggunakan kombinasi nilai  $\lambda_1$  dan  $\lambda_2$  optimum.
  - vii. Menghitung performansi model dengan nilai AUC dan Accuracy.
- d. *Generalized Extreme Value Regression*
    - i. Melakukan *fitting* model GEVR dengan data *train*.
    - ii. Melakukan uji signifikansi parameter.
    - iii. Melakukan prediksi terhadap data *train* dan *test* menggunakan model awal.
    - iv. Menghitung performansi model dengan nilai AUC dan Accuracy.
  - e. *Generalized Extreme Value Regression* dengan *Backward Elimination*
    - i. Melakukan eliminasi terhadap parameter/prediktor paling tidak signifikan satu per satu berdasarkan hasil uji tahap d(ii) hingga didapat model dengan seluruh parameter signifikan.
    - ii. Melakukan prediksi terhadap data *train* dan *test* berdasarkan model akhir yang didapat dari tahap sebelumnya.
    - iii. Menghitung performansi model dengan nilai AUC dan Accuracy.
  - f. *Logistic Regression*
    - i. Melakukan *fitting* model Logit dengan data *train*.
    - ii. Melakukan uji signifikansi parameter.
    - iii. Melakukan prediksi terhadap data *train* dan *test* menggunakan model awal.
    - iv. Menghitung performansi model dengan nilai AUC dan Accuracy.
  - g. *Stepwise Logistic Regression*
    - i. Melakukan *fitting* model Logit terhadap data *train* dengan prosedur *stepwise*.

- ii. Melakukan prediksi terhadap data *train* dan *test* berdasarkan model yang didapat.
  - iii. Menghitung performansi model dengan nilai AUC dan *Accuracy*.
- h. *Lasso Logistic Regression*
- i. Menetapkan nilai  $\gamma = 1$ .
  - ii. Melakukan *tunning* parameter dengan menentukan nilai-nilai  $\lambda$  yang akan digunakan.
  - iii. Melakukan *fitting* model dengan menggunakan  $l_1$  *penalty*.
  - iv. Melakukan prediksi terhadap data *train* dan *test* berdasarkan model yang didapatkan.
  - v. Menghitung performansi metode pada data *train* dan *test* berdasarkan nilai AUC dan *Accuracy*.
  - vi. Mendapatkan nilai  $\lambda$  optimal dan subset variabel yang relevan berdasarkan nilai performansi metode yang paling tinggi.
- i. *Elastic-Net Logistic Regression*
- i. Melakukan *tunning* parameter dengan menentukan nilai-nilai  $\lambda$  dan  $\gamma$  yang akan digunakan.
  - ii. Melakukan *fitting* model dengan menggunakan *elastic-net penalty*.
  - iii. Melakukan prediksi terhadap data *train* dan *test* berdasarkan model yang didapatkan.
  - iv. Menghitung performansi metode pada data *train* dan *test* berdasarkan nilai AUC dan *Accuracy*.
  - v. Mendapatkan nilai  $\lambda$  dan  $\gamma$  optimal dan subset variabel yang relevan berdasarkan nilai performansi metode yang paling tinggi.

Langkah analisis yang telah diuraikan di atas akan lebih mudah dipahami dengan diagram alir analisis. Maka Gambar 3.2 adalah diagram alir analisis pada penelitian ini.



Gambar 3.2 Diagram Alir Penelitian



**Gambar 3.2** Diagram Alir Penelitian (Lanjutan)

## BAB IV

### ANALISIS DAN PEMBAHASAN

Analisis dan pembahasan menyajikan hasil dari proses yang dilakukan. Proses yang dilakukan meliputi tujuan dilakukannya penelitian, yaitu karakteristik rasio keuangan perusahaan yang mengalami *financial distress* dan perusahaan yang tidak mengalami *financial distress*, melakukan prediksi *financial distress* yang akan dilakukan menggunakan metode *Support Vector Machine*, *Generalized Extreme Value Regression*, dan *Logistic Regression* menggunakan seleksi variabel untuk mendapatkan rasio-rasio keuangan yang saling berhubungan dan berpengaruh terhadap perusahaan yang mengalami *financial distress*.

#### 4.1 Karakteristik Data

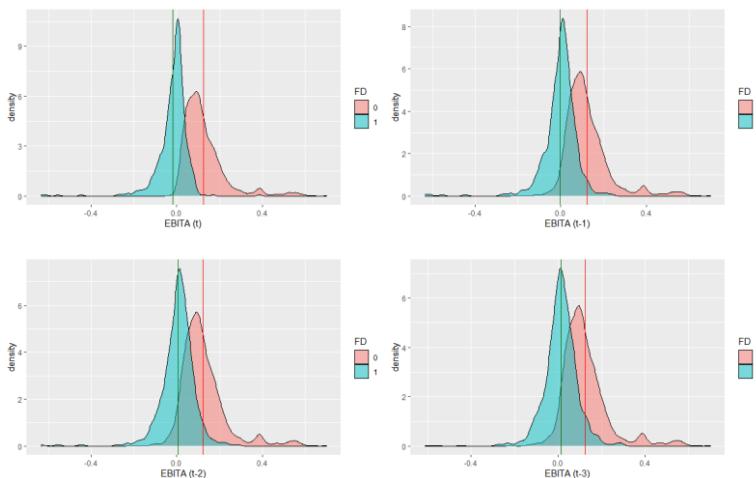
Karakteristik data yang diulas untuk menggambarkan kondisi perusahaan yang mengalami dan tidak mengalami *financial distress* merupakan statistika deskriptif yaitu *mean*, median, modus, dan varians serta *density plot*. Analisis karakteristik data dilakukan untuk mendapatkan gambaran umum pemerataan dan distribusi data. Selain itu karakteristik data juga digunakan untuk melihat perbedaan rasio keuangan antara perusahaan yang mengalami dan tidak mengalami *financial distress*. Pada Tabel 4.1 di bawah akan diberikan statistika deskriptif rasio keuangan EBITA dari tiap-tiap *size* penelitian.

**Tabel 4.1** Statistika Deskriptif EBITA

Statistik	EBITA							
	Size 0		Size 1		Size 2		Size 3	
	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0
Mean	-0,02	0,13	-0,003	0,12	0,005	0,12	0,01	0,12
Med	-0,003	0,1	0,005	0,1	0,01	0,1	0,01	0,1
Mod	0,07	0,19	0,07	0,12	0,07	0,12	0,07	0,16
Var	0,005	0,01	0,007	0,01	0,007	0,01	0,001	0,01

Ukuran rata-rata, median, modus, dan varians rasio keuangan EBITA pada perusahaan yang mengalami *financial distress* lebih kecil daripada perusahaan yang sehat baik untuk EBITA pada tahun yang sama atau tahun-tahun sebelumnya. Median rasio

keuangan EBITA pada perusahaan yang sehat di tahun yang sama dan tahun-tahun sebelumnya adalah sama yaitu 0,1. Selain itu persebaran rasio keuangan EBITA pada perusahaan yang mengalami *financial distress* dan perusahaan yang sehat digambarkan dalam Gambar 4.1 di bawah ini.



**Gambar 4.1 Density Plot EBITA**

Rata-rata rasio keuangan EBITA perusahaan yang mengalami *financial distress* pada tahun yang sama dan pada tahun sebelumnya negatif karena memiliki frekuensi EBITA negatif yang sangat tinggi walaupun modusnya mendekati nol dan positif. Sedangkan rata-rata dan modus rasio keuangan EBITA perusahaan yang mengalami *financial distress* pada dua dan tiga tahun sebelumnya positif. Dari seluruh skenario yang dibuat rata-rata dan modus rasio keuangan EBITA perusahaan yang sehat secara finansial bernilai lebih dari nol. Secara keseluruhan skenario ditinjau dari bentuk *density plot* rasio keuangan EBITA diduga berpengaruh terhadap kondisi perusahaan yang mengalami *financial distress* dan tidak mengalami *financial distress*.

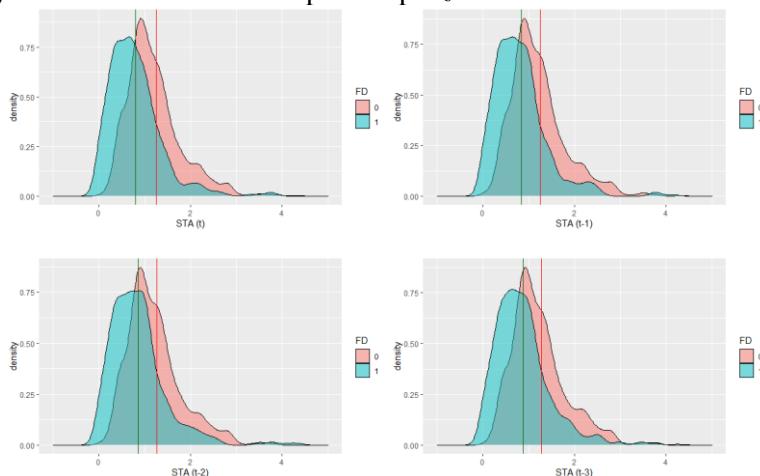
Rasio keuangan lain yang diduga berpengaruh terhadap kondisi perusahaan mengalami *financial distress* adalah STA. Berikut Tabel 4.2 adalah tabel statistika deskriptif rasio keuangan STA

pada perusahaan yang mengalami *financial distress* dan perusahaan yang sehat.

**Tabel 4.2** Statistika Deskriptif STA

Statistik	STA							
	Size 0		Size 1		Size 2		Size 3	
	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0
Mean	0,81	1,25	0,84	1,25	0,86	1,26	0,89	1,28
Med	0,71	1,09	0,73	1,1	0,75	1,11	0,77	1,11
Mod	0,29	0,5	0,29	0,92	0,29	0,92	0,29	1,04
Var	0,4	0,79	0,49	0,78	0,41	0,81	0,5	0,81

Rata-rata rasio keuangan STA pada perusahaan yang tidak mengalami *financial distress* lebih tinggi pada tiap *size* daripada perusahaan yang mengalami *financial distress* dengan perbedaan rata-rata antar *size* tidak signifikan begitupula dengan rata-rata STA antar *size* perusahaan yang mengalami *financial distress*. Hal ini juga berlaku pada median, modus dan varians, namun modus STA perusahaan yang tidak mengalami *financial distress* pada tiap *size* mengalami perubahan yang besar kecuali pada *size 1* dan *size 2*. Sedangkan modus STA perusahaan yang tidak mengalami *financial distress* konstan pada tiap *size*.



**Gambar 4.2** Density Plot STA

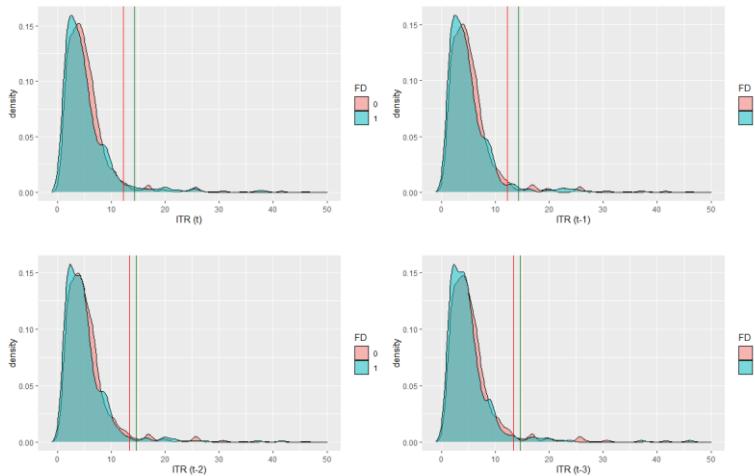
Pada Gambar 4.2 diberikan *density plot* rasio keuangan STA masing-masing *size*. Berdasarkan *density plot*, rata-rata STA

perusahaan yang mengalami *financial distress* lebih kecil daripada perusahaan yang tidak mengalami *financial distress* baik pada periode yang sama, dan periode-periode sebelumnya. Seluruh rata-rata STA perusahaan baik yang mengalami *financial distress* masing-masing *size* lebih besar daripada nol. Secara keseluruhan meninjau dari bentuk *density plot* diduga STA mempengaruhi kondisi keuangan perusahaan sehingga mengalami *financial distress*.

**Tabel 4.3** Statistika Deskriptif ITR

Statistik	ITR							
	Size 0		Size 1		Size 2		Size 3	
	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0
Mean	5,63	6,06	5,57	6,13	5,68	6,19	5,64	6,29
Med	3,99	4,53	4,12	4,56	4,2	4,6	4,18	4,62
Mod	1,29	2,53	2,78	2,53	2,78	2,53	2,78	2,77
Var	64,7	77,3	65,8	77,9	78,1	78,5	72,8	85,4

Tabel 4.3 menunjukkan statistika deskriptif rasio keuangan ITR pada perusahaan yang mengalami *financial distress* dan tidak pada masing-masing *size*. Baik dari rata-rata, median, modus, dan varians ITR pada perusahaan yang sehat lebih besar daripada perusahaan yang mengalami *financial distress* pada masing-masing *size*. Selain itu, besar rata-rata, median, modus, dan varians ITR antar perusahaan yang mengalami *financial distress* dan tidak pada tiap *size* tidak menunjukkan perbedaan yang signifikan. Nilai rata-rata jauh lebih besar daripada nilai modus yaitu pada *size* 0 atau periode tahun yang sama pada perusahaan yang mengalami *financial distress* memiliki rata-rata sebesar 5,63 dan modus sebesar 1,29 dengan rata-rata ITR perusahaan yang tidak mengalami *financial distress* sebesar 6,06 dan modus 2,53, hal ini akan menyebabkan grafik sebaran ITR baik untuk perusahaan yang mengalami *financial distress* dan tidak menjadi landai. Meninjau dari besaran-besaran statistika deskriptif ITR pada Tabel 4.2 diduga ITR tidak memberi pengaruh yang signifikan terhadap kondisi keuangan perusahaan untuk mengalami *financial distress* atau tidak.

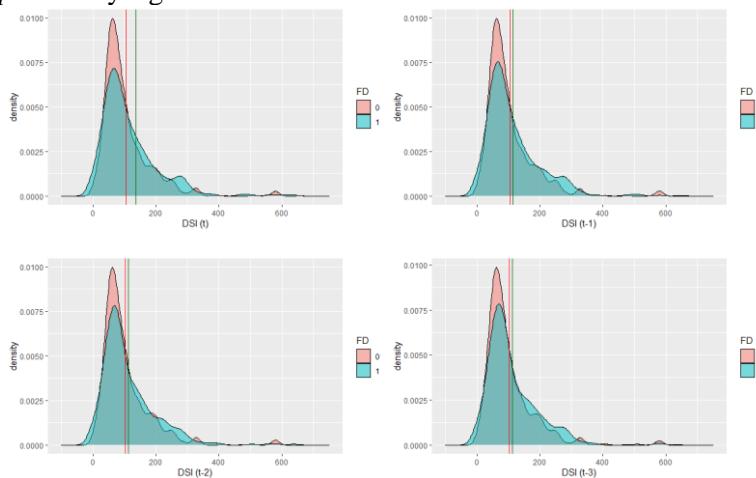
**Gambar 4.3 Density Plot ITR**

Telah dijelaskan pada Tabel 4.3 bahwa rata-rata ITR pada setiap size dan kategori perusahaan yang mengalami *financial distress* maupun tidak, memiliki rata-rata yang lebih besar daripada nilai modus. Akibatnya pada Gambar 4.3 yaitu *density plot* ITR masing-masing size memberikan bentuk condong ke kanan atau *positive skewness*. Namun rata-rata yang diberikan oleh ITR pada perusahaan yang mengalami *financial distress* lebih kecil daripada pada ITR pada perusahaan yang sehat secara finansial. Secara umum dilihat dari *density plot* ITR pada masing-masing size, rasio keuangan ITR tidak memberikan perbedaan yang signifikan terhadap kondisi keuangan perusahaan untuk dikategorikan mengalami *financial distress* atau tidak.

**Tabel 4.4** Statistika Deskriptif DSI

Statistik	DSI							
	Size 0		Size 1		Size 2		Size 3	
	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0
Mean	136,7	104,1	115	103,6	112,2	103,3	113,2	102,1
Med	88,6	80,3	86,2	79,9	84,6	79,3	85,3	78,7
Mod	131,3	144,1	131,3	144,1	131,3	144,1	131,3	131,8
Var	2022	6583,	1116	6728,	7468,	7806,	1117	650
	90,9	1	5,5	9	997	6	3,4	7,3

Statistika deskriptif rasio keuangan DSI diberikan oleh Tabel 4.4. Secara keseluruhan selain modus, yaitu rata-rata, median, dan varians DSI perusahaan yang mengalami *financial distress* lebih besar daripada perusahaan yang sehat secara finansial pada setiap *size* data. Nilai modus DSI perusahaan yang mengalami *financial distress* pada seluruh *size* data bernilai sama yaitu 131,3 begitupula pada perusahaan yang sehat dalam finansial kecuali pada *size* data 3 dimana modus DSI pada *size-size* sebelumnya adalah 144,1 menjadi 131,8 pada data *size* 3. Jika ditinjau, varians DSI pada perusahaan dengan kategori mengalami *financial distress* maupun tidak memiliki nilai yang sangat besar dan selisih yang tinggi. Ditinjau dari statistika deskriptif pada Tabel 4.4 DSI diduga memiliki pengaruh terhadap kondisi keuangan suatu perusahaan. Agar lebih jelas, analisis dilakukan secara visual melalui *density plot* DSI yang diberikan di Gambar 4.4 di bawah ini.



**Gambar 4.4** Density Plot DSI

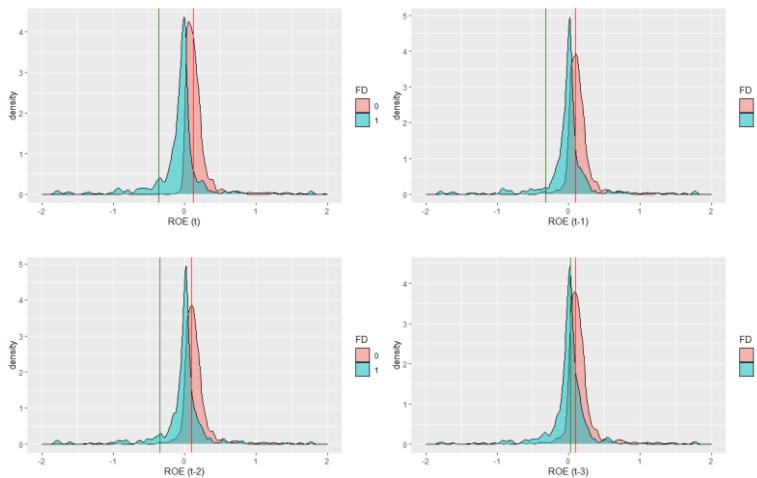
*Density plot* DSI pada Gambar 4.4 di atas menunjukkan bahwa letak perbedaan antara perusahaan yang mengalami *financial distress* dan tidak terdapat di rata-rata data *size* 0. Secara keseluruhan rata-rata DSI perusahaan yang sehat secara finansial lebih kecil daripada rata-rata perusahaan yang mengalami *financial distress*, tapi selisih yang diberikan pada data *size* 0 lebih besar daripada *size* data lainnya. Dilihat dari *density plot*, puncak atau

modus berada dititik yang hamper sama, namun modus DSI perusahaan yang sehat secara finansial lebih tinggi. Untuk membuktikan apakah rasio keuangan DSI memberi pengaruh terhadap kondisi keuangan perusahaan diperlukan analisis dan pengujian yang lebih lanjut.

**Tabel 4.5** Statistika Deskriptif ROE

Statistik	ROE							
	Size 0		Size 1		Size 2		Size 3	
	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0
Mean	-0,36	0,12	-0,32	0,1	-0,35	0,09	0,02	0,09
Med	-0,03	0,11	0,006	0,11	0,01	0,11	0,01	0,11
Mod	-0,07	0,04	0,03	0,07	-0,02	0,07	0,03	0,13
Var	52,6	3,08	56,1	3,38	60,2	3,64	1,58	3,99

Rasio keuangan yang akan dianalisis menggunakan statistika deskriptif selanjutnya adalah ROE yang telah diberikan pada Tabel 4.5 di atas. Rata-rata ROE perusahaan yang mengalami *financial distress* pada data *size* 0, 1, dan 2 bernilai negatif dan memiliki nilai hamper sama. Selanjutnya pada data *size* 3 mengalami kenaikan hingga bernilai positif. Rata-rata ROE perusahaan yang sehat secara finansial sendiri bernilai positif pada seluruh *size* data. Sedangkan nilai median ROE perusahaan yang sehat secara finansial pada seluruh *size* data bernilai sama yaitu 0,11, tetapi untuk median ROE perusahaan yang mengalami *financial distress* pada data *size* 0 bernilai negatif dan naik menjadi positif pada *size* berikutnya. Secara nilai modus ROE perusahaan yang mengalami *financial distress* maupun perusahaan yang sehat secara finansial memiliki selisih yang kecil. Perbedaan yang besar diberikan pada nilai varians, yaitu varians ROE perusahaan yang mengalami *financial distress* jauh lebih besar daripada perusahaan yang sehat secara finansial. Namun pada data *size* 3 varians ROE perusahaan yang mengalami *financial distress* lebih kecil dari perusahaan yang sehat secara finansial. Ditinjau dari besaran nilai statistika deskriptif, diduga variabel ROE memberi perbedaan terhadap kondisi keuangan perusahaan, maka dari itu akan dilakukan analisis lebih lanjut melalui *density plot* ROE pada Gambar 4.5 berikut.



Gambar 4.5 Density Plot ROE

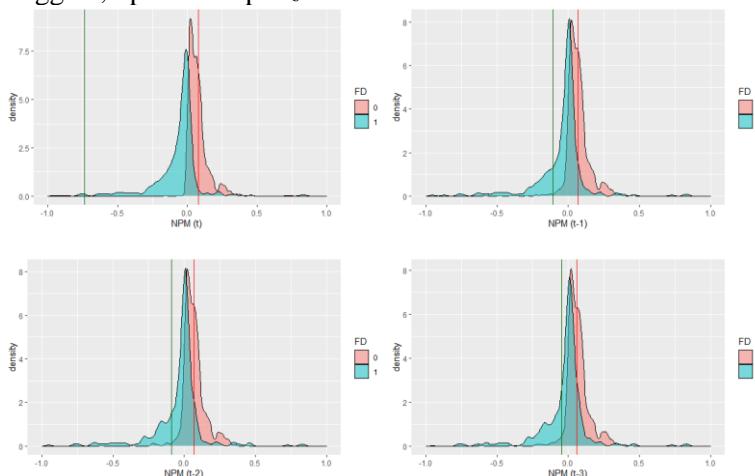
Melalui Gambar 4.5 di atas dapat dilihat bahwa rata-rata ROE perusahaan yang mengalami *financial distress* lebih kecil dari pada perusahaan yang sehat secara finansial pada seluruh *size* data dengan selisih yang tinggi pada data *size* 0, 1, dan 2. Hal tersebut disebabkan banyaknya frekuensi nilai ROE perusahaan yang mengalami *financial distress* lebih dari nilai rata-rata ROE tersebut. Sedangkan untuk rata-rata perusahaan yang sehat secara finansial tidak mengalami bayak perubahan. Ditinjau melalui *density plot*, rasio keuangan ROE memberikan pengaruh terhadap kondisi keuangan perusahaan.

Tabel 4.6 Statistika Deskriptif NPM

Statis -tik	NPM							
	Size 0		Size 1		Size 2		Size 3	
	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0
Mean	-0,74	0,08	-0,11	0,07	-0,09	0,07	-0,05	0,06
Med	-0,03	0,06	-0,01	0,06	0,0001	0,06	0,002	0,055
Mod	0,02	0,08	0,02	0,08	0,017	0,1	0,02	0,1
Var	140,3	0,03	0,56	0,03	0,56	0,02	0,56	0,02

Tabel 4.6 di atas telah menyajikan rata-rata, median, modus, dan varians dari masing-masing *size* data rasio keuangan NPM perusahaan yang mengalami *financial distress* maupun tidak. Rata-

rata NPM perusahaan yang mengalami *financial distress* pada setiap *size* data bernilai negatif dan rata-rata NPM perusahaan yang sehat secara finansial bernilai positif. Sedangkan modus NPM baik perusahaan yang mengalami *financial distress* maupun perusahaan yang sehat secara keuangan bernilai positif dan mendekati nol. Median dari NPM perusahaan yang mengalami *financial distress* pada data *size* nol dan satu bernilai negatif, tetapi mengalami kenaikan hingga bernilai positif pada *size* dua dan tiga. Varians dari NPM perusahaan yang mengalami *financial distress* pada data *size* satu bernilai sangat besar yaitu sebesar 140,3 sedangkan pada data *size* 1,2, dan 3 bernilai sama yaitu 0,56. Begitupula pada varians NPM perusahaan yang sehat secara finansial berkisar antara 0,2 hingga 0,3 pada setiap *size* data.



Gambar 4.6 Density Plot NPM

Gambar 4.6 menunjukkan *density plot* NPM terhadap perusahaan yang mengalami *financial distress* dan perusahaan yang sehat secara finansial pada masing-masing *size* data. Rata-rata NPM perusahaan yang mengalami *financial distress* pada periode yang sama memiliki nilai yang sangat kecil bila dibandingkan dengan rata-rata NPM perusahaan yang mengalami *financial distress* pada *size* data lainnya. Berdasarkan *density plot* yang diberikan terlihat bahwa modus NPM pada perusahaan yang mengalami *financial distress* dan perusahaan yang sehat secara

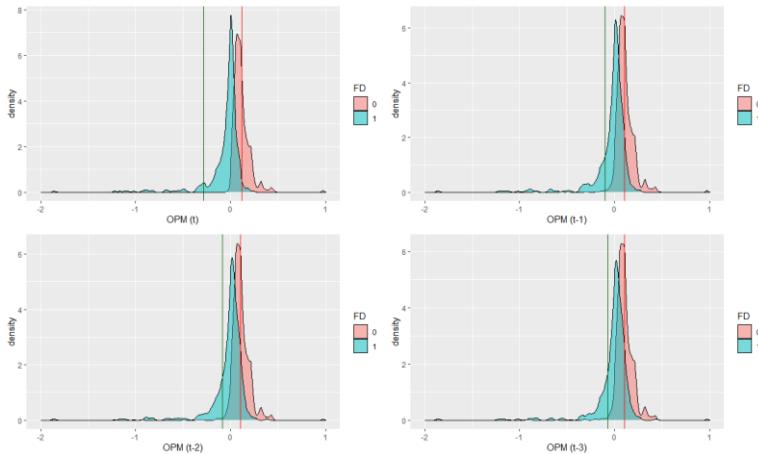
finansial adalah sama atau hamper sama. Secara visual melalui *density plot* diduga NPM memberi perbedaan pada perusahaan yang mengalami *financial distress* dan tidak.

**Tabel 4.7** Statistika Deskriptif OPM

Statistik	OPM							
	Size 0		Size 1		Size 2		Size 3	
	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0
Mean	-0,29	0,12	-0,1	0,11	-0,09	0,1	-0,07	0,1
Med	-0,006	0,1	0,005	0,09	0,01	0,09	0,013	0,09
Mod	0,002	0,18	0,003	0,11	0,003	0,15	0,003	0,15
Var	6,36	0,03	0,52	0,02	0,5	0,01	0,51	0,01

Tabel 4.7 merupakan tabel statistika deskriptif rasio keuangan OPM perusahaan dengan kategori mengalami *financial distress* dan sehat secara finansial pada seluruh *size* data. Ditinjau melalui rata-rata, OPM perusahaan dengan kategori *financial distress* bernilai negatif dan lebih kecil dari OPM perusahaan dengan kategori sehat secara finansial pada setiap *size* data. Begitupula untuk nilai median dan modus OPM pada perusahaan yang mengalami *financial distress* secara keseluruhan bernilai lebih kecil daripada perusahaan yang sehat secara finansial. Namun varians OPM perusahaan yang mengalami *financial distress* bernilai lebih besar pada seluruh *size* data. Khusunya pada data *size* 0, memiliki varians OPM dengan selisih besar antara perusahaan yang mengalami *financial distress* dan tidak mengalami *financial distress*. Secara umum melalui statsitika deskriptif rasio keuangan OPM memberi pengaruh terhadap kondisi keuangan perusahaan untuk dikategorikan dalam perusahaan yang mengalami *financial distress* atau tidak.

Selain menggunakan statistika deskriptif, untuk melihat apakah rasio keuangan DSI memberi pengaruh terhadap kondisi keuangan perusahaan akan dilakukan analisis melalui *density plot*. Gambar 4.7 merupakan *density plot* OPM perusahaan yang mengalami *financial distress* dan tidak pada masing-masing *size* data.



Gambar 4.7 Density Plot OPM

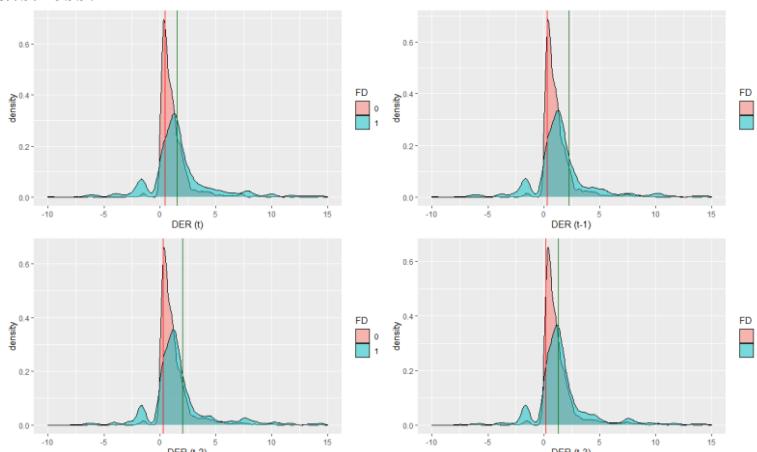
Melalui *density plot* yang diberikan pada Gambar 4.7 dapat dilihat bahwa rata-rata OPM perusahaan yang mengalami *financial distress* negatif dan mengalami kenaikan dari OPM size 0 hingga 3. Sedangkan rata-rata OPM perusahaan yang sehat secara finansial terlihat konstan pada rata-rata OPM sekitar 0,1. Modus OPM perusahaan yang mengalami *financial distress* positif disekitar 0 meskipun memiliki rata-rata negatif. Secara visual melalui *density plot* diduga OPM memberi pengaruh terhadap kondisi keuangan perusahaan.

Tabel 4.8 Statistika Deskriptif DER

Statistik	DER							
	Size 0		Size 1		Size 2		Size 3	
	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0
Mean	1,5	0,46	2,26	0,29	2,05	0,3	1,29	0,19
Med	1,31	0,83	1,28	0,83	1,26	0,84	1,26	0,84
Mod	1,26	2,32	1,26	0,17	1,26	0,17	1,26	0,25
Var	475,4	1053,1	375,8	1146, 9	387, 1	124 9,95	94,1	137 2,4

Tabel 4.8 di atas merupakan table statistika deskriptif rasio keuangan DER pada perusahaan yang mengalami *financial distress* maupun tidak. Rata-rata DER pada setiap size data, perusahaan yang mengalami *financial distress* memiliki rata-rata yang lebih

tinggi daripada perusahaan yang sehat secara finansial. Selain pada rata-rata, median DER juga lebih besar pada perusahaan yang mengalami *financial distress* daripada perusahaan yang sehat berbadning terbaik dengan modus dimana perusahaan yang sehat secara finansial memiliki modus lebih besar daripada perusahaan yang mengalami *financial distress*. Nilai median dan modus DER memiliki besar yang hamper sama pada perusahaan yang mengalami *financial distress*. Sedangkan varians DER untuk perusahaan yang sehat secara finansial sangat lebih tinggi dibandingkan dengan perusahaan yang tidak mengalami *financial distress*.



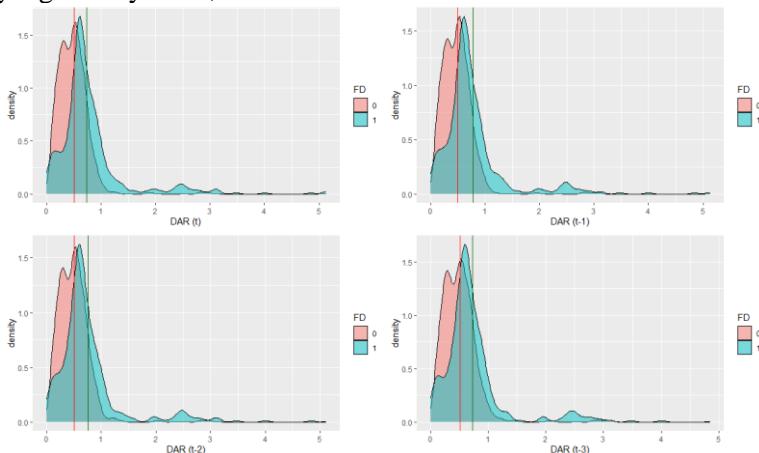
**Gambar 4.8 Density Plot DER**

*Density plot* rasio keuangan DER diberikan pada Gambar 4.8 mulai dari DER *size* 0 hingga 3. Melalui garis rata-rata terlihat bahwa rata-rata DER perusahaan yang mengalami *financial distress* lebih besar daripada perusahaan yang sehat secara finansial. Sebaran rasio keuangan DER perusahaan yang sehat secara finansial lebih tingga daripada perusahaan yang mengalami *financial distress* pada setiap *size* data. Maka dari itu untuk melihat apakah rasio keuangan DER memberi pengaruh terhadap kondisi keuangan perusahaan.

**Tabel 4.9** Statistika Deskriptif DAR

Statistik	DAR							
	Size 0		Size 1		Size 2		Size 3	
	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0
Mean	0,8	0,49	0,78	0,49	0,75	0,5	0,73	0,51
Med	0,64	0,47	0,64	0,48	0,63	0,48	0,62	0,48
Mod	0,55	0,76	0,55	0,11	0,55	0,11	0,55	0,21
Var	0,43	0,1	0,39	0,11	0,37	0,11	0,32	0,12

Tabel 4.9 merupakan statistika deskriptif rasio keuangan DAR pada seluruh *size* data baik pada perusahaan yang mengalami *financial distress* atau tidak. Rata-rata, median, dan varians DAR perusahaan yang mengalami *financial distress* lebih besar daripada perusahaan yang sehat secara finansial. Begitupula pada modus rasio keuangan DAR, tetapi pada DAR *size* 0, perusahaan yang sehat secara finansial memiliki modus yang lebih tinggi. Nilai modus perusahaan yang mengalami *financial distress* pada setiap *size* memiliki besar yang sama yaitu 0,55 dan nilai modus perusahaan yang sehat secara finansial pada DAR *size* 1 dan 2 memiliki besar yang sama yaitu 0,11.

**Gambar 4.9** Density Plot DAR

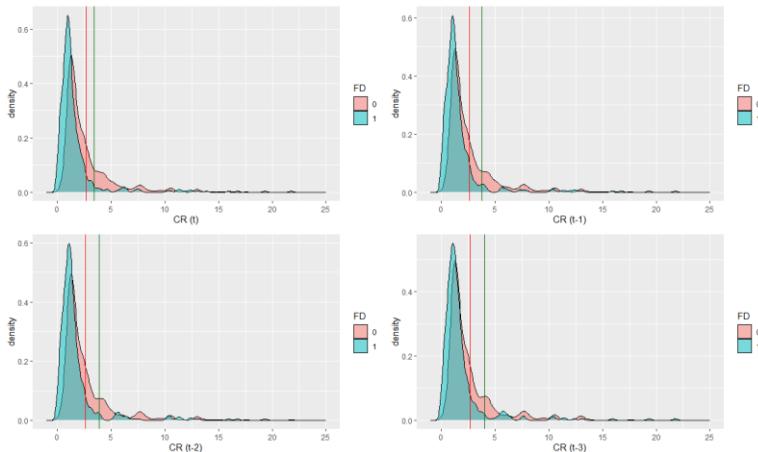
Gambar 4.9 adalah *density plot* rasio keuangan DAR pada masing-masing *size*. Secara visual melalui *density plot* tidak terdapat perbedaan antar *size* data. Namun *density plot* antara DAR

perusahaan yang mengalami *financial distress* dan DAR perusahaan yang sehat secara finansial memiliki perbedaan. Melalui garis rata-rata dapat dilihat bahwa pada setiap *size* data, rata-rata DAR perusahaan yang sehat lebih kecil daripada rata-rata DAR perusahaan yang mengalami *financial distress*. Ditinjau dari *density plot* diduga rasio keuangan DAR memberi pengaruh terhadap kondisi keuangan perusahaan.

**Tabel 4.10** Statistika Deskriptif CR

Statistik	CR							
	Size 0		Size 1		Size 2		Size 3	
	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0
Mean	3,4	2,65	3,77	2,59	3,89	2,61	3,98	2,64
Med	1,1	1,75	1,16	1,72	1,18	1,72	1,22	1,71
Mod	1,81	10,6	1,81	10,5	1,81	10,6	1,81	4,39
Var	480,4	8,85	522,7	5,83	556,1	7,08	587,2	9,64

Statistika deskriptif rasio keuangan CR atau *Current Ratio* terhadap kondisi keuangan perusahaan diberikan oleh Tabel 4.10. Rata-rata CR pada perusahaan yang mengalami *financial distress* lebih besar daripada rata-rata yang dimiliki perusahaan yang sehat secara finansial. Sedangkan median dan modus perusahaan yang mengalami *financial distress* lebih kecil daripada perusahaan yang sehat secara finansial. Dari besaran CR tersebut, nilai rata-rata perusahaan yang mengalami *financial distress* lebih tinggi dibandingkan nilai median dan modusnya. Sedangkan untuk perusahaan yang sehat secara finansial, nilai modus lebih tinggi daripada nilai rata-rata dan median. Sedangkan untuk nilai varians CR, pada seluruh *size* data perusahaan yang mengalami *financial distress* memiliki varians yang sangat tinggi dibandingkan dengan perusahaan yang sehat secara finansial. Secara umum dari analisis statistika deskriptif CR diduga memberi pengaruh terhadap kondisi keuangan perusahaan, namun dibutuhkan analisis lainnya untuk membuktikan apakah CR memberi pengaruh terhadap kondisi keuangan perusahaan.



Gambar 4.10 Density Plot CR

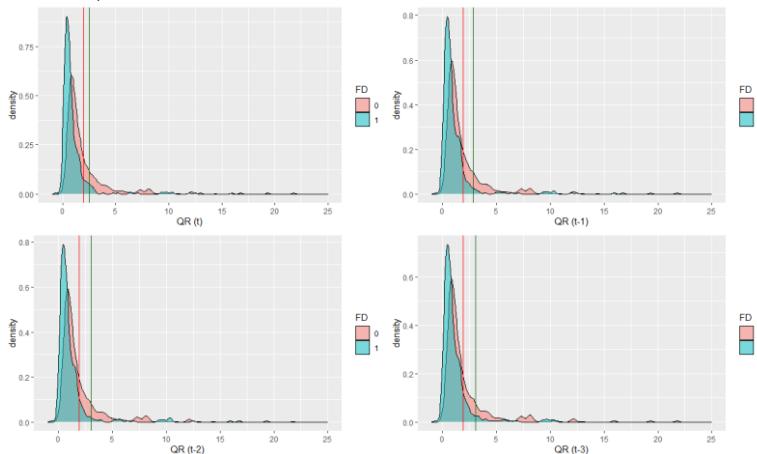
Gambar 4.10 adalah *density plot* rasio keuangan CR pada masing-masing *size* data. Dapat dilihat bahwa pada setiap *size* data, rata-rata CR perusahaan yang mengalami *financial distress* lebih besar daripada perusahaan yang sehat secara finansial. Secara keseluruhan dari *density plot* di atas, tidak terdapat perbedaan yang signifikan dari masing-masing *size* data. Melihat dari *density plot* perusahaan yang mengalami *financial distress* dan perusahaan yang sehat diduga kurang memberi pengaruh. Namun untuk mendukung asumsi tersebut diperlukan analisis atau pengujian lainnya.

Tabel 4.11 Statistika Deskriptif QR

Statistik	QR							
	Size 0		Size 1		Size 2		Size 3	
	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0
Mean	2,5	1,93	2,84	1,88	3,02	1,87	3,06	1,92
Med	0,6	1,18	0,67	1,15	0,69	1,15	0,7	1,15
Mod	1,51	8,22	1,51	8,22	0,12	8,22	1,51	8,22
Var	431,8	6,05	466,6	4,46	0,25	4,19	530,1	5,66

Tabel 4.11 adalah tabel yang menyajikan statistika deskriptif rasio keuangan QR pada seluruh *size* data untuk perusahaan yang mengalami *financial distress* dan tidak. Meninjau dari statistika deskriptif yang diberikan setiap *size* menunjukkan besaran yang

sama, tapi kondisi yang berbeda ditunjukkan pada *size* 2 yaitu varians pada QR perusahaan yang mengalami *financial distress* lebih rendah daripada perusahaan yang sehat secara finansial dimana pada *size* lainnya menunjukkan varians yang lebih tinggi. Kondisi yang unik juga diberikan oleh modus dimana modus QR pada perusahaan yang tidak mengalami *financial distress* adalah sama yaitu 8,22 dan modus QR pada perusahaan yang mengalami *financial distress* adalah 1,51 kecuali pada QR data *size* 2 yaitu sebear 0,12.



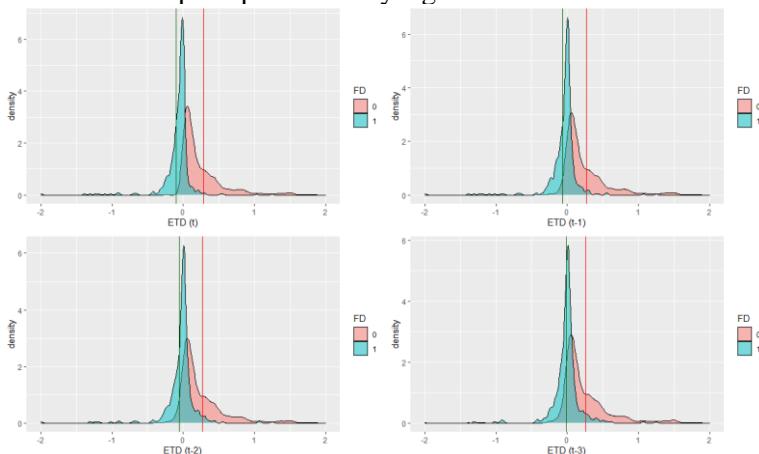
Gambar 4.11 Density Plot QR

Gambar 4.11 adalah *density plot* rasio keuangan QR masing-masing *size* data baik pada perusahaan yang mengalami *financial distress* atau perusahaan yang sehat secara finansial. Sama seperti rasio keuangan sebelumnya yaitu CR bahwa ditinjau dari *size* data, *density plot* tidak memberikan perbedaan yang signifikan. Namun jika dilihat berdasarkan *density plot* perusahaan yang mengalami *financial distress* dan perusahaan yang sehat secara finansial terdapat beberapa perbedaan. Sehingga QR diduga memberi pengaruh terhadap kondisi keuangan perusahaan untuk dikategorikan mengalami *financial distress* atau perusahaan yang sehat secara finansial.

**Tabel 4.12** Statistika Deskriptif ETD

Statistik	ETD							
	Size 0		Size 1		Size 2		Size 3	
	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0
Mean	-0,1	0,28	-0,07	0,27	-0,06	0,27	-0,01	0,26
Med	-0,03	0,13	-0,01	0,13	0,0001	0,13	0,003	0,13
Mod	-0,02	0,08	0,03	0,08	-0,02	0,45	0,03	0,54
Var	1,74	0,18	1,87	0,18	2,03	0,16	2,06	0,21

Rasio keuangan lainnya yang akan digunakan untuk memprediksi klasifikasi perusahaan tergolong sebagai perusahaan yang mengalami *financial distress* atau tidak adalah ETD. Sama halnya dengan sebelumnya, statistika deskriptif ETD akan dianalisis terlebih dahulu yang mana telah diberikan pada Tabel 4.12. Rata-rata ETD setiap *size* data perusahaan yang mengalami *financial distress* bernilai negatif dan lebih kecil daripada perusahaan yang sehat secara finansial. Sedangkan median dari perusahaan yang dikategorikan sehat secara finansial bernilai sama yaitu 0,13. Setiap *size* data memiliki modus ETD perusahaan yang sehat secara finansial lebih besar daripada perusahaan yang mengalami *financial distress*. Begitupula varians ETD perusahaan yang mengalami *financial distress* secara keseluruhan *size* data lebih besar daripada perusahaan yang sehat secara finansial.

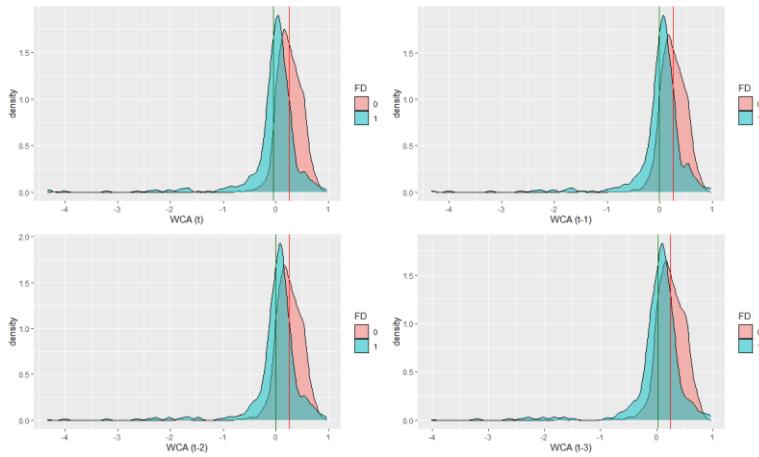
**Gambar 4.12** Density Plot ETD

Melalui Gambar 4.12 yaitu *density plot* rasio keuangan ETD perusahaan yang mengalami *financial distress* dan tidak pada seluruh *size* data. Melalui *density plot* dapat dilihat bahwa tidak terdapat perbedaan yang signifikan antar *size* data. Sedangkan *density plot* antar perusahaan yang mengalami *financial distress* dan perusahaan yang sehat secara finansial terdapat perbedaan khususnya pada puncak plot perusahaan yang mengalami *financial distress* sangat tinggi. Secara umum melalui *density plot* ETD, rasio keuangan tersebut memberikan pengaruh terhadap kondisi keuangan perusahaan untuk dikategorikan pada perusahaan yang mengalami *financial distress* dan tidak.

**Tabel 4.13** Statistika Deskriptif WCA

Statistik	WCA							
	Size 0		Size 1		Size 2		Size 3	
	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0
Mean	-0,05	0,25	-0,01	0,25	0,00	0,25	0,03	0,25
Med	0,03	0,25	0,06	0,24	0,06	0,25	0,09	0,25
Mod	0,12	0,15	0,12	0,55	0,12	0,55	0,12	0,41
Var	0,3	0,07	0,28	0,07	0,25	0,08	0,21	0,08

Statistika deskriptif WCA baik untuk perusahaan yang mengalami *financial distress* dan tidak pada setiap *size* data diberikan pada Tabel 4.13. Nilai rata-rata, median dan modus WCA perusahaan yang tidak mengalami *financial distress* lebih besar daripada WCA perusahaan yang mengalami *financial distress*. Modus WCA pada perusahaan yang mengalami *financial distress* pada seluruh *size* data menunjukkan modus pada WCA 0,12. Sedangkan untuk varians baik pada perusahaan yang mengalami *financial distress* maupun perusahaan yang sehat secara finansial bernilai kurang dari 1, namun varians WCA perusahaan yang mengalami *financial distress* pada seluruh *size* data bernilai lebih besar daripada varians WCA perusahaan yang sehat secara finansial. Secara umum ditinjau dari statistika deskriptif WCA pada Tabel 4.13 dapat diduga bahwa WCA memberi pengaruh terhadap kondisi keuangan perusahaan untuk dikategorikan pada perusahaan yang mengalami *financial distress* atau perusahaan yang sehat secara finansial.



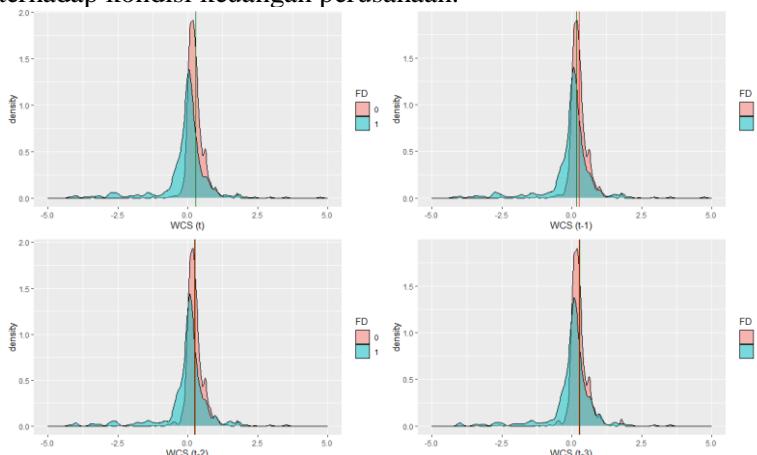
Gambar 4.13 Density Plot WCA

Setelah dilakukan analisis melalui statistika deskriptif, untuk melihat apakah rasio keuangan WCA memberi pengaruh terhadap kondisi keuangan perusahaan juga dilakukan dengan membandingkan *density plot* antara WCA perusahaan yang mengalami *financial distress* dan WCA perusahaan yang sehat secara finansial. Secara keseluruhan dibandingkan antar *size* data, *density plot* tidak memberikan perbedaan yang signifikan. Sedangkan ditinjau dari *density plot* WCA perusahaan yang mengalami *financial distress* dan perusahaan yang sehat secara finansial memiliki perbedaan sebaran. Garis rata-rata menunjukkan bahwa WCA perusahaan yang mengalami *financial distress* kurang dari rata-rata WCA perusahaan yang sehat secara finansial.

Tabel 4.14 Statistik Deskriptif WCS

Statistik	WCS							
	Size 0		Size 1		Size 2		Size 3	
	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0	FD=1	FD=0
Mean	0,29	0,28	0,18	0,26	0,23	0,27	0,28	0,26
Med	0,03	0,21	0,07	0,21	0,07	0,21	0,09	0,21
Mod	0,68	0,63	0,68	0,21	0,68	0,63	0,68	0,63
Var	295,4	0,48	8,65	0,35	8,15	0,32	8,34	0,21

Tabel 4.14 merupakan tabel statistika deskriptif rasio keuangan WCS perusahaan yang mengalami *financial distress* dan tidak pada seluruh *size* data. Rata-rata yang diberikan pada Tabel 4.14 menunjukkan bahwa selisih rata-rata antara WCS perusahaan yang mengalami *financial distress* dan WCS perusahaan yang sehat secara finansial terbilang sedikit dan berada pada rata-rata yang hampir sama. Untuk nilai median WCS sendiri, perusahaan yang sehat secara finansial yaitu 0,21 pada seluruh *size* data lebih besar daripada median WCS perusahaan yang mengalami *financial distress*. Hal serupa yaitu nilai yang sama terdapat pada modus WCS seluruh *size* data dan baik untuk perusahaan yang mengalami *financial distress* maupun perusahaan yang sehat secara finansial, namun pada modus WCS perusahaan yang sehat secara finansial *size* 1 berbeda dan lebih rendah daripada modus pada WCS *size* lainnya. Sedangkan varians WCS perusahaan yang mengalami *financial distress* pada data *size* 0 sangat besar daripada varians lainnya. Secara umum diduga WCS kurang memberi pengaruh terhadap kondisi keuangan perusahaan.



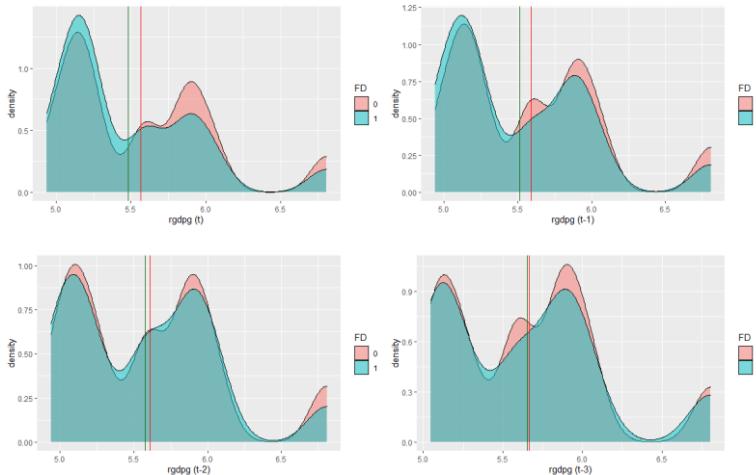
**Gambar 4.14 Density Plot WCS**

Analisis yang dilakukan selanjutnya untuk melihat apakah WCS memberi pengaruh terhadap kondisi keuangan perusahaan adalah melalui *density plot* pada Gambar 4.14. Dapat dilihat bahwa *density plot* antar *size* memiliki bentuk yang hampir seupa. Selain itu garis rata-rata antara WCS perusahaan yang mengalami

*financial distress* dan tidak memiliki selisih yang kecil hingga pada *size* 0, 2, dan 3 terlihat berada pada titik yang sama. Namun bila dilihat bentuk *density plot* WCS perusahaan yang mengalami *financial distress* berbentuk lebih landau daripada *density plot* perusahaan yang sehat secara finansial pada seluruh *size* data. Secara umum melalui *density plot* pada Gambar 4.14 diduga WCS kurang berpengaruh terhadap kondisi keuangan perusahaan sehingga diperlukan analisis atau pengujian lainnya untuk mendapatkan kesimpulannya.

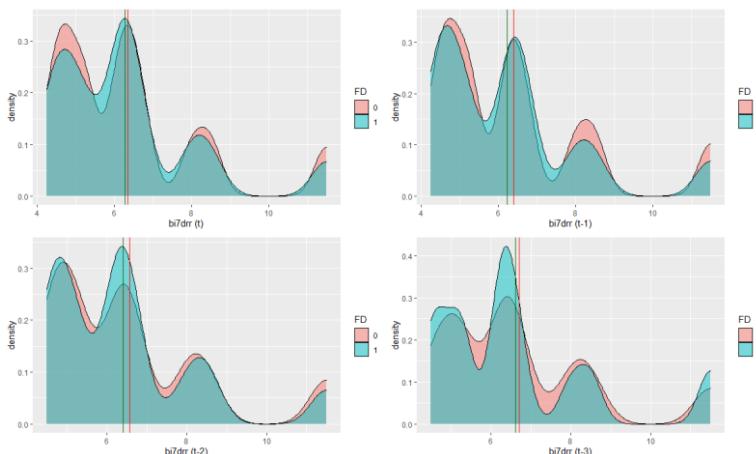
Kondisi keuangan suatu perusahaan dapat dipengaruhi oleh faktor internal dan faktor eksternal. Faktor internal digambarkan oleh rasio-rasio keuangan yang dikeluarkan perusahaan dalam laporan keuangan perusahaan. Sedangkan faktor eksternal digambarkan melalui kondisi perekonomian secara nasional. Seperti yang telah dijelaskan sebelumnya bahwa faktor eksternal yang menggambarkan kondisi keuangan perusahaan dalam penelitian ini adalah faktor makroekonomi diantaranya yaitu *Real GDP Growth*, *BI 7-Day Repo Rate*, *USD/IDR*, dan Inflasi.

Tujuan faktor eksternal yaitu variabel makroekonomi dianalisis dan dimasukkan ke dalam model adalah bukan untuk melihat apakah variabel makroekonomi signifikan atau tidak pada perusahaan yang mengalami kondisi *financial distress* dan tidak, tetapi untuk mengakomodasi dependensi antar perusahaan secara tidak langsung. Model-model yang digunakan mengasumsikan bahwa data antar perusahaan dan antar periode adalah independen padahal seharusnya ada dependensi. Variabel makroekonomi digunakan untuk mengakomodasi dependensi antar perusahaan dan antar periode sehingga model yang dihasilkan menjadi kompleks dan model yang mengasumsikan antar perusahaan dan antar periode adalah independen menjadi tidak sesuai. Maka variabel makroekonomi mengakomodasi dependensi antar perusahaan ditangkap secara tidak langsung karena variabel tersebut memberikan dampak ke seluruh perusahaan.



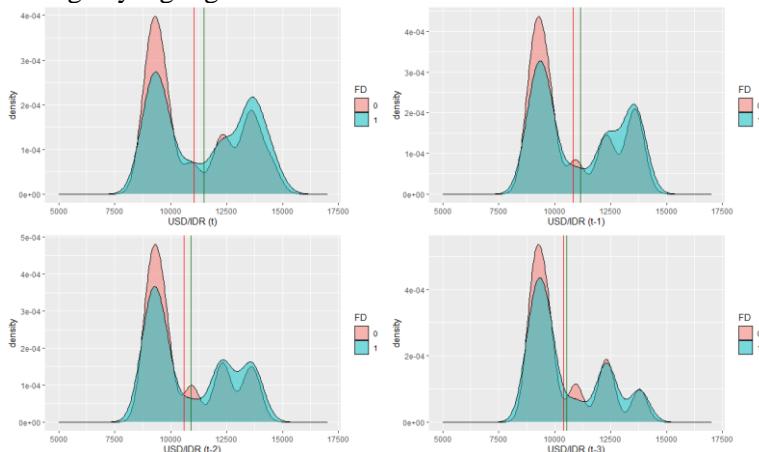
**Gambar 4.15 Density Plot RGDGP**

Melalui *density plot* variabel makroekonomi *Real GDP Growth* di atas yaitu pada Gambar 4.15 dapat dilihat bahwa rata-rata RGDGP perusahaan yang sehat secara finansial lebih besar daripada RGDGP perusahaan yang mengalami *financial distress* dengan rata-rata yang mendekati sama pada data size tiga. *Density plot* RGDGP perusahaan yang mengalami *financial distress* dan perusahaan sehat secara finansial naik pada RGDGP sekitar nol.



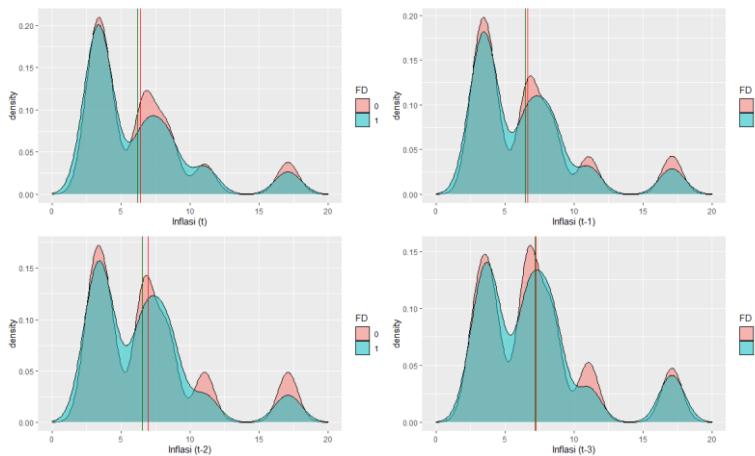
**Gambar 4.16 Density Plot BI7DRR**

*Density plot BI 7 Day Repo Rate* yang merupakan salah satu variabel makroekonomi yang digunakan dalam penelitian ini diberikan pada Gambar 4.16 dapat dilihat bahwa secara umum dari masing-masing *size* data tidak memberikan perbedaan yang signifikan. Rata-rata yang BI7DRR perusahaan yang mengalami *financial distress* lebih kecil daripada rata-rata BI7DRR perusahaan yang sehat secara finansial pada seluruh *size* data. Seperti yang telah dijelaskan sebelumnya bahwa penggunaan variabel makroekonomi adalah untuk mengakomodasi antar observasi independen yang belum diakomodasi oleh rasio kauangan yang digunakan.



**Gambar 4.17 Density Plot USD/IDR**

Gambar 4.17 adalah *density plot* dari salah satu variabel makroekonomi yang digunakan dalam penelitian yaitu nilai tukar mata uang Rupiah dalam *US Dollar*. dapat dilihat melalui *density plot* USD/IDR untuk data *size* 3 dimana rata-rata antara nilai tukar pada perusahaan yang mengalami *financial distress* dan perusahaan yang sehat secara finansial adalah hampir sama. Sedangkan rata-rata nilai tukar perusahaan yang mengalami *financial distress* pada seluruh *size* data lebih besar daripada rata-rata nilai tukar mata uang perusahaan yang sehat secara finansial.



**Gambar 4.18 Density Plot Inflasi**

Variabel makroekonomi yang digunakan lagi dalam penelitian ini adalah tingkat inflasi per tahun. Melalui Gambar 4.18 yaitu *density plot* inflasi terhadap perusahaan yang mengalami *financial distress* dan tidak menggunakan seluruh *size* data. Dapat dilihat bahwa hampir tidak terjadi perubahan antar *size* data pada *density plot* inflasi ini.

## 4.2 Klasifikasi Menggunakan Support Vector Machines (SVM)

Prediksi klasifikasi yang akan dilakukan adalah menggunakan metode *Support Vector Machine* linier dan *Regularized Support Vector Machine* menggunakan penalti *lasso* dan *elastic-net*. Data yang digunakan sendiri terdiri dari 4 *size* data dimana masing-masing data terdiri dari 80% data *training* dan 20% data *testing*. Dalam melakukan prediksi digunakan konsep *full memory time window* seperti yang telah diilustrasikan pada subbab 3.3.

### 4.2.1 Support Vector Machine Linier

Metode *Support Vector Machine* secara linier adalah metode yang memisahkan data secara linier. Metode ini memisahkan data dan menghasilkan prediksi tanpa adanya seleksi variabel. Prediksi yang dilakukan menggunakan metode *Support Vector Machine* secara linier terhadap masing-masing data *size* diberikan pada *Confussion Matrix* pada Tabel 4.15.

**Tabel 4.15a** Confussion Matrix Model SVM Linier *Size 0*

Size 0					
<b>Training</b>	Aktual		<b>Testing</b>	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1193	72	0	295	26
1	20	310	1	9	105

**Tabel 4.15b** Confussion Matrix Model SVM Linier *Size 1**Size 1*

Size 1					
<b>Training</b>	Aktual		<b>Testing</b>	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1065	184	0	279	42
1	36	165	1	25	89

**Tabel 4.15c** Confussion Matrix Model SVM Linier *Size 2**Size 2*

Size 2					
<b>Training</b>	Aktual		<b>Testing</b>	Aktual	
Prediksi	0	1	Prediksi	0	1
0	959	188	0	284	67
1	29	129	1	20	64

**Tabel 4.15d** Confussion Matrix Model SVM Linier *Size 3**Size 3*

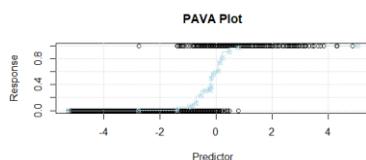
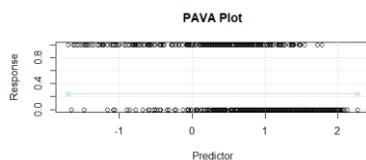
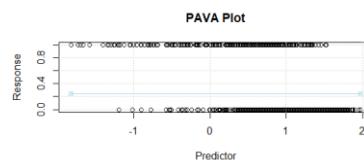
Size 3					
<b>Training</b>	Aktual		<b>Testing</b>	Aktual	
Prediksi	0	1	Prediksi	0	1
0	852	180	0	197	57
1	20	108	1	107	74

Melihat dari *confussion matrix* dari masing-masing hasil prediksi maka dapat ditemukan ketepatan klasifikasi dari masing-masing prediksi. Ukuran yang digunakan dalam menentukan model prediksi terbaik antar *size* data untuk menentukan perusahaan mengalami *financial distress* atau tidak adalah AUC dan Accuracy yang diberikan pada Tabel 4.16.

**Tabel 4.16** Ketepatan Klasifikasi Model SVM Linier

<i>Size</i>	Ukuran	<i>Training</i>	<i>Testing</i>
0	AUC	0,9412	0,9200
	Accuracy	0,9423	0,9195
1	AUC	0,8368	0,8249
	Accuracy	0,8483	0,8460
2	AUC	0,8263	0,7855
	Accuracy	0,8337	0,8000
3	AUC	0,8347	0,5922
	Accuracy	0,8276	0,6230

AUC pada hasil prediksi data *train* dan data *test size 0* merupakan nilai AUC terbesar. Hal serupa terjadi pada Accuracy data *train* dan data *test size 0*. Semakin besar *size* data, baik AUC maupun Accuracy semakin kecil. Pada hasil klasifikasi pada data *testing size 4* memberikan hasil bahwa prediksi klasifikasi kelas 0 yang diklasifikasikan pada kelas 1 lebih besar daripada hasil prediksi kelas 1 yang diklasifikasikan pada kelas 1. Sedangkan nilai AUC yang diberikan lebih kecil daripada Accuracy. Maka dari itu, AUC dianggap lebih menunjukkan ketepatan klasifikasi yang lebih baik daripada Accuracy. Sehingga klasifikasi perusahaan mengalami *financial distress* atau tidak yang terbaik menggunakan metode *Support Vector Machine* secara linier dengan rasio keuangan dan variabel makroekonomi yang memiliki periode sama.

**Gambar 4.19a** SVM Linier Size 0**Gambar 4.19b** SVM Linier Size 1**Gambar 4.19c** SVM Linier Size 2**Gambar 4.19d** SVM Linier Size 3

Melalui Gambar 4.19 dapat dilihat bahwa menggunakan *isotonic regression* dengan prediktor berupa skor *Support Vector Machine* dari masing-masing *size* data diperoleh hasil estimasi yang monoton naik. Prediktor yang digunakan untuk membentuk *isotonic regression* yaitu skor *output SVM* yang diperoleh melalui estimasi dari persamaan  $\mathbf{x}'_i \mathbf{w} + b$ .  $\mathbf{w}$  merupakan vektor koefisien parameter atau bobot yang tegak lurus terhadap *hyperplane* dan  $b$  adalah *intercept* atau posisi-posisi bidang relatif terhadap pusat koordinat. Sedangkan respon yang digunakan adalah respon asli dari data yaitu kondisi keuangan perusahaan apakah mengalami *financial distress* atau tidak. Hasil yang diharapkan dari adanya *isotonic regression* adalah dihasilkan probabilitas yang monoton naik yaitu dengan skor kecil memiliki peluang besar untuk tidak mengalami *financial distress* dan skor tinggi berpeluang mengalami *financial distress* lebih besar. Algoritma yang digunakan adalah *Pair-Adjacent-Violators Algorithm* yang mana jika skor SVM diperengkatkan dengan benar, maka probabilitas  $P(Y_j = 1|x_i)$  dimana  $Y_j \in \{0,1\}$  diasumsikan akan naik seiring kenaikan skor SVM. Namun, terjadi pelanggaran pada *isotonic regression* skor SVM data *size* 1, 2, dan 3. Dilihat dari plot status *financial ditress* bahwa pada *size* 0 perusahaan yang mengalami *financial distress* cenderung memiliki skor yang tinggi. Sedangkan pada data *size* lainnya perusahaan yang mengalami *financial distress* memiliki skor yang lebih kecil dan cenderung sama dengan perusahaan yang sehat secara keuangan. Jika skor SVM ke-(*i*-1) yaitu  $g(x_{(i-1)})$  lebih besar dari  $g(x_{(i)})$  maka nilai  $x_{(i-1)}$  dan  $x_{(i)}$  sebagai *pair-adjacent* melanggar asumsi (*violator*). Nilai dari  $g(x_{(i-1)})$  dan  $g(x_{(i)})$  digantikan dengan masing-masing rata-rata, hingga nilai-nilai tersebut tidak melanggar asumsi *isotonic*. Tahap ini diulang terus-menerus hingga didapat nilai yang *isotonic* dan pada skor SVM data *size* 1, 2, dan 3 diperoleh hasil yang monoton dari observasi ke-1 hingga ke-n.

#### 4.2.2 Lasso Support Vector Machine

Parameter yang diperlukan untuk melakukan permodelan menggunakan *Lasso Support Vector Machine* adalah *regularized parameter*  $\lambda$ . Nilai  $\lambda$  yang mendekati 0 akan membuat variabel terpilih semakin banyak, begitu pula sebaliknya jika nilai  $\lambda$  mendekati tak hingga maka variabel yang terpilih akan semakin sedikit. Parameter  $\lambda$  optimal dipilih berdasarkan hasil prediksi terbaik pada masing-masing *size* data. Tabel 4.17 akan menunjukkan kesalahan klasifikasi dari model yang telah dibuat pada masing-masing data *size*.

**Tabel 4.17a** Confussion Matrix Model Lasso SVM Size 0

Size 0

<b>Training</b>		<b>Aktual</b>		<b>Testing</b>		<b>Aktual</b>	
Prediksi		0	1	Prediksi		0	1
0	1152	61		0	287	17	
1	70	312		1	23	108	

**Tabel 4.17b** Confussion Matrix Model Lasso SVM Size 1

Size 1

<b>Training</b>		<b>Aktual</b>		<b>Testing</b>		<b>Aktual</b>	
Prediksi		0	1	Prediksi		0	1
0	1073	28		0	296	8	
1	246	103		1	93	38	

**Tabel 4.17c** Confussion Matrix Model Lasso SVM Size 2

Size 2

<b>Training</b>		<b>Aktual</b>		<b>Testing</b>		<b>Aktual</b>	
Prediksi		0	1	Prediksi		0	1
0	972	16		0	302	2	
1	247	70		1	104	27	

**Tabel 4.17d Confusion Matrix Model Lasso SVM Size 3***Size 3*

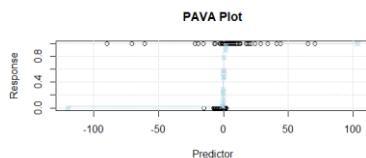
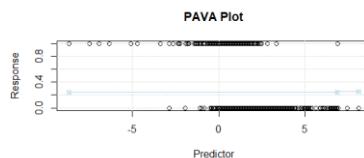
<b>Training</b>		<b>Aktual</b>		<b>Testing</b>		<b>Aktual</b>	
Prediksi		0	1	Prediksi		0	1
0		860	248	0		301	118
1		12	40	1		3	14

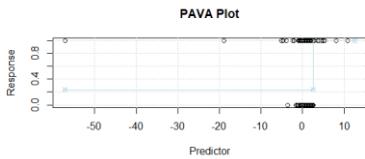
Hasil prediksi di atas diberikan oleh model terbaik yang dapat setelah melakukan *tunning* parameter  $\lambda$ . Tabel 4.18 di bawah akan memberikan nilai AUC dan *Accuracy* data *train* dan data *test* masing-masing data *size* dari model terbaik.

**Tabel 4.18 Ketepatan Klasifikasi Model Lasso SVM**

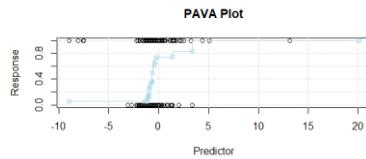
<i>Size</i>	$\lambda$	p*	<i>Training</i>		<i>Testing</i>	
			<i>Accuracy</i>	AUC	<i>Accuracy</i>	AUC
0	0,0143	12	0,9179	0,8832	0,9080	0,8843
1	0,0121	12	0,8262	0,6348	0,7678	0,6319
2	0,0317	7	0,7985	0,6023	0,7563	0,5998
3	0,0023	17	0,7802	0,5759	0,7333	0,5638

Model yang dihasilkan oleh data *size* 3 menghasilkan prediktor yang masuk dalam model paling banyak. Parameter  $\lambda$  yang diberikan untuk masing-masing model menentukan jumlah prediktor yang dapat dimodelkan. Ketepatan klasifikasi tertinggi baik data *train* ataupun data *test* berdasarkan nilai AUC, data *size* 0 memberikan ketepatan klasifikasi tertinggi. Sehingga model prediksi klasifikasi kondisi keuangan perusahaan terbaik menggunakan metode *Lasso Support Vector Machine* adalah model data *size* 0 dengan prediktor yang digunakan sebanyak 12 prediktor.

**Gambar 4.20a Lasso SVM Size 0****Gambar 4.20b Lasso SVM Size 1**



Gambar 4.20c Lasso SVM Size 2



Gambar 4.20d Lasso SVM Size 3

Melalui Gambar 4.20 dapat dilihat bahwa hasil estimasi menggunakan *isotonic regression* diperoleh bahwa probabilitas yang didapat bersifat monoton naik. Menggunakan *isotonic regression* dengan prediktor berupa skor *Lasso Support Vector Machine* dari masing-masing *size* data diperoleh hasil estimasi berupa probabilitas yang monoton naik. Sama halnya dengan metode *Support Vector Machine*, estimasi dengan seleksi variabel *Lasso* dilakukan dengan persamaan  $\mathbf{x}_i' \mathbf{w} + b$  yang mana koefisien parameter diberikan pada Lampiran 14. Pada probabilitas yang dihasilkan dari *isotonic regression* data *size* 1 dan 2 didapat plot yang cenderung monoton. Artinya banyak pelanggaran asumsi oleh skor yang didapat.

#### 4.2.3 Elastic-Net Support Vector Machine

Permodelan menggunakan *Elastic-Net Support Vector Machine* dilakukan dengan melakukan *tunning* parameter  $\lambda_1$  dan  $\lambda_2$  terlebih dahulu dimana bernilai lebih dari nol. Apabila nilai  $\lambda_2$  konstan, maka semakin besar nilai  $\lambda_1$  maka variabel yang terpilih akan semakin sedikit variabel yang terpilih. Sementara apabila  $\lambda_1$  bernilai konstan semakin besar nilai  $\lambda_2$  maka variabel yang terpilih akan semakin sedikit variabel yang terpilih. Tabel 4.19 merupakan tabel hasil prediksi klasifikasi perusahaan yang mengalami *financial distress* dan tidak menggunakan model *Elastic-Net Support Vector Machine* yang didapat.

Tabel 4.19a Confussion Matrix Model Elastic-Net SVM Size 0

Size 0					
<b>Training</b>	Aktual		<b>Testing</b>	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1201	137	0	299	29
1	12	245	1	5	102

**Tabel 4.19b** Confusion Matrix Model Elastic-Net SVM Size 1

Size 1					
<b>Training</b>	Aktual		<b>Testing</b>	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1050	174	0	285	42
1	51	175	1	19	89

**Tabel 4.19c** Confusion Matrix Model Elastic-Net SVM Size 2

Size 2					
<b>Training</b>	Aktual		<b>Testing</b>	Aktual	
Prediksi	0	1	Prediksi	0	1
0	974	253	0	296	101
1	14	64	1	8	30

**Tabel 4.19d** Confusion Matrix Model Elastic-Net SVM Size 3

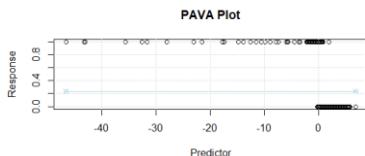
Size 3					
<b>Training</b>	Aktual		<b>Testing</b>	Aktual	
Prediksi	0	1	Prediksi	0	1
0	855	232	0	300	122
1	17	56	1	4	9

Melalui *confussion matrix* hasil prediksi dengan kondisi aktual keuangan perusahaan dapat dihitung ketepatan klasifikasi masing-masing model. Ketepatan klasifikasi akan ditunjukkan melalui AUC dan Accuracy masing-masing data *train* dan data *test* setiap data *size*. Pada Tabel 4.20 akan diberikan masing-masing AUC dan Accuracy serta parameter  $\lambda_1$  dan  $\lambda_2$  dan jumlah prediktor yang digunakan untuk mendapatkan model.

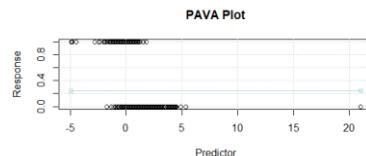
**Tabel 4.20** Ketepatan Klasifikasi Model Elastic-Net SVM

Size	$\lambda_1$	$\lambda_2$	p*	<b>Training</b>		<b>Testing</b>	
				Accuracy	AUC	Accuracy	AUC
0	4,365	0,125	15	0,9066	0,9255	0,9218	0,9324
1	0,184	0,008	18	0,8448	0,8161	0,8598	0,8478
2	26,02	0,25	13	0,7954	0,8072	0,7494	0,7675
3	16,34	0,25	15	0,7853	0,7768	0,7103	0,7016

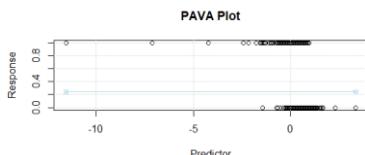
Berdasarkan hasil ketepatan klasifikasi yang diberikan oleh masing-masing hasil prediksi oleh data *train* atau data *test*, data *size* 0 memberikan ketepatan klasifikasi terbesar baik dari *Accuracy* atau AUC. Parameter  $\lambda_1$  dan  $\lambda_2$  terkecil diberikan oleh model data *size* 1 sehingga mendapatkan prediktor yang masuk ke dalam model terbanyak. Disebabkan oleh ketepatan klasifikasi yang tinggi diberikan oleh model prediksi data *size* 0 maka model terbaik diprediksi menggunakan rasio keuangan pada periode yang sama dengan kondisi keuangan perusahaan.



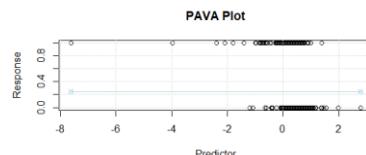
**Gambar 4.21a** Elastic-Net SVM Size 0



**Gambar 4.21b** Elastic-Net SVM Size 1



**Gambar 4.21c** Elastic-Net SVM Size 2



**Gambar 4.21d** Elastic-Net SVM Size 3

Melalui Gambar 4.21 dapat dilihat penggunaan *isotonic regression* memperlihatkan hasil estimasi berupa nilai probabilitas yang monoton naik. Prediktor yang digunakan untuk melakukan *isotonic regression* adalah skor yang dihasilkan oleh *Elastic-Net Support Vector Machine* dimana koefisien parameter yang digunakan telah diberikan pada Lampiran 15. Permodelan menggunakan *Elastic-Net SVM* memberikan skor yang cenderung sama antar perusahaan yang mengalami *financial distress* dan tidak. Akibatnya pada *isotonic regression* terjadi pelanggaran asumsi dan memberikan probabilitas yang monoton pada seluruh data *size*.

#### 4.3 Klasifikasi Menggunakan *Generalized Extreme Value Regression*

Prediksi klasifikasi menggunakan metode *Generalized Extreme Value Regression* dilakukan secara serentak dan menggunakan

seleksi variabel. Permodelan dengan seleksi variabel menggunakan *backward elimination*. Data yang digunakan dalam melakukan prediksi baik data *training* dan data *testing* sama dengan data yang digunakan untuk klasifikasi *financial distress* perusahaan sektor industri pada klasifikasi menggunakan metode *Support Vector Machine*.

#### 4.3.1 Generalized Extreme Value Regression secara Serentak

Permodelan dilakukan menggunakan data *training* sehingga didapatkan model untuk dilakukan klasifikasi. Sebelum dilakukan permodelan, terlebih dahulu ditentukan parameter  $\tau$ . Dalam menentukan nilai  $\tau$  yang paling optimum dilakukan pengujian menggunakan 20 nilai dari -0,1 hingga 0,1. Dari seluruh nilai  $\tau$  yang diberikan didapat bahwa  $\tau=0,05$  menghasilkan nilai log likelihood terbesar, namun  $\tau=0,1$  memberikan nilai yang lebih baik sehingga permodelan dilakukan menggunakan  $\tau=0,1$ . Tabel 4.21 berikut adalah model dari permodelan menggunakan *Generalized Extreme Value Regression*.

**Tabel 4.21** Model Prediksi *Financial Distress* GEVR secara Serentak

Size	Model
0	$\hat{z} = -2,053 - 17,33\text{EBITA} - 0,3299\text{STA} + 0,008014\text{ITR} + 0,001402\text{DSI} - 0,08246\text{ROE} - 4,444\text{NPM} - 9,6630\text{PM} + 0,004357\text{DER} + 0,9226\text{DAR} + 0,06317\text{CR} - 0,1172\text{QR} - 1,954\text{ETD} - 0,4504\text{WCA} + 0,1207\text{WCS} + 0,2078\text{RGDPG} + 0,1166\text{BI7DRR} + 0,00009453\text{USD\_IDR} - 0,0094\text{Inflasi}$
1	$\hat{z} = 1,765 - 23,52\text{EBITA} - 0,31\text{STA} + 0,0005081\text{ITR} + 0,0002761\text{DSI} + 0,3044\text{ROE} + 0,6622\text{NPM} + 0,9032\text{OPM} + 0,03378\text{DER} + 1,419\text{DAR} + 0,009809\text{CR} - 0,01919\text{QR} - 0,1958\text{ETD} - 0,1315\text{WCA} + 0,021487\text{WCS} - 0,3203\text{RGDPG} - 0,1296\text{BI7DRR} + 0,00006149\text{USD\_IDR} - 0,04523\text{Inflasi}$
2	$\hat{z} = -3,031 - 4,757\text{EBITA} - 0,1645\text{STA} + 0,0006135\text{ITR} + 0,001066\text{DSI} + 0,1636\text{ROE} - 0,7203\text{NPM} - 2,861\text{OPM} + 0,03704\text{DER} + 0,47\text{DAR} - 0,047\text{CR} + 0,1512\text{QR} - 0,2879\text{ETD} - 0,5813\text{WCA} + 0,374\text{WCS} + 0,245\text{RGDPG} + 0,03407\text{BI7DRR} + 0,0001701\text{USD\_IDR} - 0,03212\text{Inflasi}$

**Tabel 4.21** Model Prediksi *Financial Distress* GEVR secara Serentak (Lanjutan)

Size	Model
3	$\hat{z} = 3,856 - 3,864\text{EBITA} - 0,1\text{STA} + 0,000164\text{ITR} + 0,0008\text{DSI} + 0,03311\text{ROE} - 0,1167\text{NPM} - 2,913\text{OPM} + 0,00888\text{DER} + 0,4062\text{DAR} - 0,00636\text{CR} + 0,0352\text{QR} - 0,0388\text{ETD} - 0,6535\text{WCA} - 0,0619\text{WCS} - 0,165\text{RGDPG} - 0,139\text{BI7DRR} - 0,000254\text{USD_IDR} + 0,0689\text{Inflasi}$

Estimasi parameter dan uji signifikansi parameter untuk mendapatkan model-model pada Tabel 4.21 dapat dilihat pada Lampiran 17. Permodelan menggunakan metode *Generalized Extreme Value Regression* secara serentak ini didapatkan hasil bahwa tidak seluruh rasio keuangan maupun variabel makroekonomi berpengaruh terhadap kondisi keuangan perusahaan. Tabel 4.22 berikut menunjukkan prediktor-prediktor yang signifikan pada masing-masing model.

**Tabel 4.22** Variabel Signifikan Model Serentak GEVR

Size	Variabel
0	EBITA, STA, ITR, NPM, OPM, DAR, ETD
1	EBITA, OPM, DER, DAR
2	EBITA, STA, OPM, DAR, QR, WCA, WCS, USD/IDR
3	EBITA, OPM, DAR, WCA, RGDPG, USD/IDR, Inflasi

Permodelan menggunakan metode *Generalized Extreme Value Regression* secara serentak menunjukkan bahwa hanya terdapat beberapa variabel yang signifikan dalam. Namun analisis menggunakan model secara serentak tetap dilakukan. Tabel 4.23 berikut adalah nilai *Confussion Matrix* hasil klasifikasi pada masing-masing size data baik untuk data *training* dan *testing*.

**Tabel 4.23a** *Confussion Matrix* Model Serentak GEVR Size 0

Size 0			Size 0		
<i>Training</i>	Aktual		<i>Testing</i>	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1169	66	0	286	20
1	44	316	1	18	111

**Tabel 4.23b** Confussion Matrix Model Serentak GEVR Size 1

Size 1

<b><i>Training</i></b>		<b>Aktual</b>		<b><i>Testing</i></b>		<b>Aktual</b>	
Prediksi		0	1	Prediksi		0	1
0		1026	139	0		267	28
1		75	210	1		37	103

**Tabel 4.23c** Confussion Matrix Model Serentak GEVR Size 2

Size 2

<b><i>Training</i></b>		<b>Aktual</b>		<b><i>Testing</i></b>		<b>Aktual</b>	
Prediksi		0	1	Prediksi		0	1
0		950	173	0		250	30
1		38	144	1		54	101

**Tabel 4.23d** Confussion Matrix Model Serentak GEVR Size 3

Size 3

<b><i>Training</i></b>		<b>Aktual</b>		<b><i>Testing</i></b>		<b>Aktual</b>	
Prediksi		0	1	Prediksi		0	1
0		833	176	0		300	109
1		39	112	1		4	22

Prediksi klasifikasi yang telah diberikan pada Tabel 4.23 memiliki tingkat ketepatan klasifikasi yang berbeda-beda. Ketepatan klasifikasi pada masing-masing hasil prediksi ditentukan dengan *Accuracy* dan AUC dimana untuk gambar ROC diberikan pada Lampiran 12. Tabel 4.24 adalah tabel yang menunjukkan *Accuracy* dan AUC masing-masing hasil prediksi.

**Tabel 4.24** Ketepatan Klasifikasi Model Serentak GEVR

<i>Size</i>	<i>Ukuran</i>	<i>Training</i>	<i>Testing</i>
0	AUC	0,9760	0,9708
	Accuracy	0,9367	0,9241

**Tabel 4.24** Ketepatan Klasifikasi Model Serentak GEVR (Lanjutan)

<i>Size</i>	Ukuran	<i>Training</i>	<i>Testing</i>
1	AUC	0,8947	0,9280
	<i>Accuracy</i>	0,8586	0,8667
2	AUC	0,8721	0,8782
	<i>Accuracy</i>	0,8414	0,8207
3	AUC	0,8494	0,8288
	<i>Accuracy</i>	0,8233	0,7954

AUC dan *Accuracy* tertinggi diberikan oleh prediksi klasifikasi data *training* pada data *size* 0. Begitu pula untuk AUC dan *Accuracy* data *testing* tertinggi diberikan pada data *testing size* 0. AUC dan *Accuracy* hasil prediksi menggunakan metode *Generalized Extreme Value Regression* pada data *size* 0 lebih dari 0,9 sehingga termasuk dalam kategori *excellent classification*. Sedangkan baik AUC ataupun *Accuracy* pada data *size* 1,2, dan 3 terus menurun seiring bertambahnya *size* data.

#### 4.3.2 Generalized Extreme Value Regression dengan Backward Elimination

Prosedur yang digunakan dalam penelitian ini adalah *backward elimination* yaitu dengan mengeliminasi satu per satu prediktor (rasio keuangan dan varabel makroekonomi) yang tidak signifikan dilihat dari *p-value* terbesar. Koefisien parameter model yang didapat dengan *backward elimination* menggunakan metode *Generalized Extreme Value Regression* diberikan pada Lampiran 18. Penggunaan *backward elimination* pada metode *Generalized Extreme Value Regression* adalah dengan membentimbangkan hasil dari analisis secara serentak dimana memiliki ketepatan klasifikasi yang tinggi namun terdapat prediktor yang belum signifikan. Sama halnya dengan permodelan yang dilakukan menggunakan metode sebelumnya yaitu data yang digunakan untuk membangun model adalah data *training* dari masing-masing *size* data. Prediktor-prediktor yang terpilih dengan prosedur *backward elimination* diberikan pada Tabel 4.25 berikut.

**Tabel 4.25** Variabel Signifikan Model *Backward GEVR*

<i>Size</i>	Variabel	<i>Step</i>
0	EBITA, STA, ITR, NPM, OPM, DAR, QR, ETD, BI7DRR	9
1	EBITA, STA, OPM, DER, DAR, QR, RGDGP, Inflasi	10
2	EBITA, STA, NPM, OPM, DAR, QR, WCA, USD/IDR	10
3	EBITA, OPM, DAR, WCA	14

Prediktor yang terpilih dimodelkan kembali dan model baru yang terbentuk digunakan untuk membuat klasifikasi baru untuk data *training* dan data *testing*. Model yang dihasilkan dengan prediktor yang telah signifikan secara keseluruhan diberikan pada Tabel 4.26 berikut.

**Tabel 4.26** Model Prediksi *Financial Distress* GEVR dengan *Backward Elimination*

<i>Size</i>	Model
0	$\hat{z} = 0,5054 - 18,2381\text{EBITA} - 0,3662\text{STA} + 0,00774\text{ITR} - 4,5315\text{NPM} - 8,65780\text{PM} + 1,0408\text{DAR} - 0,0536\text{QR} - 1,9861\text{ETD} + 0,066\text{BI7DRR}$
1	$\hat{z} = 1,5041 - 24,0229\text{EBITA} - 0,323\text{STA} + 0,58570\text{PM} + 0,03\text{DER} + 1,4987\text{DAR} + 0,0485\text{QR} - 0,282\text{RGDPG} + 0,063\text{Inflasi}$
2	$\hat{z} = -6,273 - 4,248\text{EBITA} - 0,192\text{STA} - 0,8804\text{NPM} - 3,640\text{PM} + 0,5861\text{DAR} + 0,0554\text{QR} - 0,7346\text{WCA} + 0,0000787\text{USD_IDR}$
3	$\hat{z} = 0,1168 - 4,8853\text{EBITA} - 2,2349\text{OPM} + 0,3381\text{DAR} - 0,6507\text{WCA}$

Melalui model yang baru, maka dilakukan prediksi klasifikasi baik pada data *training* maupun data *testing*. Kesalahan yang terjadi pada klasifikasi menggunakan model baru adalah perusahaan yang sehat diklasifikasikan pada kategori perusahaan yang mengalami *financial distress*. Perbandingan hasil klasifikasi dengan data aktual diberikan pada Tabel 4.27 berikut.

**Tabel 4.27a Confussion Matrix Model Backward GEVR Size 0**

Size 0					
<b>Training</b>	Aktual		<b>Testing</b>	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1172	67	0	290	20
1	41	315	1	14	111

**Tabel 4.27b Confussion Matrix Model Backward GEVR Size 1**

Size 1					
<b>Training</b>	Aktual		<b>Testing</b>	Aktual	
Prediksi	0	1	Prediksi	0	1
0	992	107	0	276	31
1	109	242	1	28	100

**Tabel 4.27c Confussion Matrix Model Backward GEVR Size 2**

Size 2					
<b>Training</b>	Aktual		<b>Testing</b>	Aktual	
Prediksi	0	1	Prediksi	0	1
0	940	171	0	277	48
1	48	146	1	27	83

**Tabel 4.27c Confussion Matrix Model Backward GEVR Size 3**

Size 3					
<b>Training</b>	Aktual		<b>Testing</b>	Aktual	
Prediksi	0	1	Prediksi	0	1
0	834	189	0	294	77
1	38	99	1	10	54

Meninjau *Confussion Matrix* pada Tabel 4.27 didapat ketepatan klasifikasi pada masing-masing prediksi. Ukuran ketepatan klasifikasi yang digunakan pada metode *Generalized Extreme Value Regression* dengan *backward elimination* adalah AUC dan Accuracy. Tabel 4.28 berikut memberikan masing-masing nilai AUC dan Accuracy pada tiap prediksi yang dilakukan.

**Tabel 4.28** Ketepatan Klasifikasi Model *Backward GEVR*

<i>Size</i>	Ukuran	<i>Training</i>	<i>Testing</i>
0	AUC	0,9754	0,9705
	Accuracy	0,9354	0,9356
1	AUC	0,8886	0,9263
	Accuracy	0,8531	0,8667
2	AUC	0,8699	0,8913
	Accuracy	0,8398	0,8299
3	AUC	0,8415	0,8500
	Accuracy	0,8241	0,8092

Secara keseluruhan pada *size* data dan data *training* maupun *testing*, AUC lebih besar daripada *Accuracy*. Nilai AUC dan *Accuracy* prediksi model pada data *size* 1 baik untuk data *training* atau *testing* memberikan lebih besar daripada 0,9 sehingga tergolong pada kategori *excellent classification*. Baik AUC maupun *Accuracy* data *training* dan *testing* terbesar adalah pada data *size* 1. Sehingga diduga kondisi keuangan perusahaan dipengaruhi oleh rasio keuangan pada periode yang sama.

#### 4.4 Klasifikasi Menggunakan *Logistic Regression*

Prediksi klasifikasi yang akan dilakukan yaitu menggunakan metode *Logistic Regression*. *Logistic Regression* pada penelitian ini diterapkan secara serentak dan dengan seleksi variabel yaitu melalui prosedur *stepwise* dan menggunakan metode regularisasi *lasso* dan *elastic-net*. Masing-masing prosedur diterapkan untuk data *training* data *size* 0 hingga 4. Ketepatan klasifikasi dihitung melalui AUC dan *Accuracy* dari hasil prediksi klasifikasi data *training* dan data *testing*.

##### 4.4.1 *Binary Logistic Regression* secara Serentak

Estimasi parameter dilakukan menggunakan data *training* masing-masing *size* data. Setelah didapat model masing-masing *size* data dan melalui uji signifikansi parameter didapat bahwa ada prediktor yang tidak signifikan. Hal ini dapat terjadi disebabkan oleh adanya multikolinieritas dimana untuk memprediksi kondisi keuangan perusahaan telah diwakilkan oleh prediktor lainnya.

Model-model yang didapatkan adalah sebagai berikut pada Tabel 4.29. Model pada Tabel 4.29 mengikutsertakan seluruh prediktor baik yang signifikan dan tidak.

**Tabel 4.29** Model Prediksi *Financial Distress* LOGIT secara Serentak

Size	Model
0	$g(x) = 2,638 + 27,75\text{EBITA} + 0,7154\text{STA} - 0,014\text{ITR} - 0,0034\text{DSI} + 0,119\text{ROE} + 8,984\text{NPM} + 26,5320\text{OPM} - 0,00642\text{DER} - 1,632\text{DAR} - 0,0568\text{CR} - 0,061\text{QR} + 5,85\text{ETD} + 0,966\text{WCA} - 0,1293\text{WCS} - 0,236\text{RGDPG} - 0,1525\text{BI7DRR} - 0,0001363\text{USD\_IDR} + 0,01297\text{Inflasi}$
1	$g(x) = -1,765 + 23,52\text{EBITA} + 0,31\text{STA} + 0,00051\text{ITR} - 0,000276\text{DSI} - 0,3044\text{ROE} + 0,6622\text{NPM} - 0,90320\text{OPM} - 0,03378\text{DER} - 1,419\text{DAR} - 0,00981\text{CR} - 0,019\text{QR} + 0,196\text{ETD} + 0,132\text{WCA} - 0,02148\text{WCS} + 0,32\text{RGDPG} + 0,1296\text{BI7DRR} - 0,000009\text{USD\_IDR} + 0,04523\text{Inflasi}$
2	$g(x) = 5,649 + 14,76\text{EBITA} + 0,248\text{STA} - 0,00045\text{ITR} - 0,00061\text{DSI} - 0,072\text{ROE} + 0,5539\text{NPM} + 1,1510\text{OPM} - 0,004314\text{DER} - 0,812\text{DAR} + 0,08274\text{CR} - 0,1936\text{QR} + 0,496\text{ETD} + 0,81\text{WCA} + 0,1056\text{WCS} - 0,4203\text{RGDPG} - 0,1167\text{BI7DRR} - 0,0002629\text{USD\_IDR} + 0,07053\text{Inflasi}$
3	$g(x) = -5,166 + 11,31\text{EBITA} + 0,183\text{STA} - 0,00025\text{ITR} - 0,001119\text{DSI} + 0,1916\text{ROE} - 0,1533\text{NPM} + 1,2130\text{OPM} - 0,01517\text{DER} - 0,5982\text{DAR} + 0,0197\text{CR} - 0,0412\text{QR} + 0,1365\text{ETD} + 0,82\text{WCA} + 0,0254\text{WCS} + 0,1556\text{RGDPG} + 0,254\text{BI7DRR} + 0,0004262\text{USD\_IDR} + 0,1464\text{Inflasi}$

Seperti yang sudah dijelaskan bahwa model serentak yang dibangun menggunakan metode *Binary Logistic Regression* secara serentak memiliki beberapa prediktor yang tidak berpengaruh signifikan di dalam model. Diantara prediktor-prediktor yang digunakan, pada Tabel 4.30 akan ditunjukkan prediktor yang memiliki pengaruh signifikan dalam masing-masing model.

**Tabel 4.30** Variabel Signifikan Model Serentak LOGIT

Size	Variabel
0	EBITA, STA, NPM, OPM, DAR, ETD

**Tabel 4.30** Variabel Signifikan Model Serentak LOGIT (Lanjutan)

<i>Size</i>	Variabel
1	EBITA, STA, OPM, DER, DAR
2	EBITA, DAR, QR, USD/IDR
3	EBITA, DAR, WCA, BI7DRR, Inflasi

Meninjau dari Tabel 4.30 bahwa hanya terdapat sedikit rasio keuangan dan variabel makroekonomi yang memiliki pengaruh terhadap kondisi keuangan perusahaan, namun analisis tetap dilanjutkan menggunakan model penuh untuk mengetahui keakuratan metode dalam menghasilkan prediksi kondisi keuangan perusahaan. Tabel 4.31 berikut menunjukkan hasil prediksi baik data *training* maupun data *testing* menggunakan model dengan kondisi keuangan aktual.

**Tabel 4.31a** Confussion Matrix Model Serentak LOGIT *Size 0*

Size 0					
<b>Training</b>	Aktual		<b>Testing</b>	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1174	39	0	291	13
1	68	314	1	21	110

**Tabel 4.31b** Confussion Matrix Model Serentak LOGIT *Size 1*

Size 1					
<b>Training</b>	Aktual		<b>Testing</b>	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1037	64	0	281	23
1	160	189	1	37	94

**Tabel 4.31c** Confussion Matrix Model Serentak LOGIT *Size 2*

Size 2					
<b>Training</b>	Aktual		<b>Testing</b>	Aktual	
Prediksi	0	1	Prediksi	0	1
0	945	43	0	256	48
1	171	146	1	37	94

**Tabel 4.31d Confussion Matrix Model Serentak LOGIT Size 3***Size 3*

<b>Training</b>		<b>Aktual</b>		<b>Testing</b>		<b>Aktual</b>	
Prediksi		0	1	Prediksi		0	1
0		834	38	0		300	4
1		178	110	1		115	16

*Confussion matrix* di atas adalah perbandingan kondisi keuangan perusahaan aktual dengan hasil prediksi menggunakan model dari metode *Binary Logistic Regression* secara serentak. Melalui hasil tersebut dapat dihitung nilai-nilai ketepatan klasifikasi menggunakan AUC dan Accuracy. Tabel 4.32 memberikan AUC dan Accuracy masing-masing hasil prediksi.

**Tabel 4.32 Ketepatan Klasifikasi Model Serentak LOGIT**

Size	Ukuran	Training	Testing
0	AUC	0,8949	0,8985
	Accuracy	0,9329	0,9218
1	AUC	0,7417	0,8209
	Accuracy	0,8455	0,8621
2	AUC	0,7085	0,7798
	Accuracy	0,8360	0,8046
3	AUC	0,6692	0,5545
	Accuracy	0,8138	0,7264

AUC data *training* pada *size* 0 lebih kecil daripada data *testing* pada data *size* yang sama. Sedangkan pada data *size* 0 juga, *Accuracy* data *training* lebih besar daripada data *testing*. Seperti ketepatan klasifikasi pada metode *Generalized Extreme Value Regression*, pada metode *Logistic Regression* juga menunjukkan penurunan AUC maupun *Accuracy* pada penambahan *size* data. Sehingga dengan menggunakan metode *Binary Logistic Regression* secara serentak, kondisi keuangan perusahaan dapat dijelaskan menggunakan model dengan rasio keuangan dan variabel makroekonomi pada periode yang sama.

#### 4.4.2 *Binary Logistic Regression* secara *Stepwise*

Evaluasi dari permodelan menggunakan metode *Binary Logistic Regression* secara serentak adalah terdapat prediktor yang tidak signifikan masuk ke dalam model sehingga tidak efisien untuk memprediksi kondisi keuangan perusahaan. Upaya untuk membentuk model terbaik pada penelitian ini adalah dengan mengaplikasikan prosedur *stepwise* pada metode *Binary Logistic Regression*. Menggunakan model lengkap yang telah disebutkan pada Tabel 4.29 dilakukan eliminasi sehingga membentuk model yang memiliki nilai kebaikan model AUC paling rendah. Tabel 4.33 memperlihatkan prediktor-prediktor yang masuk dalam model menggunakan metode *Binary Logistic Regression* dengan prosedur *stepwise*.

**Tabel 4.33** Variabel Signifikan Model *Stepwise* LOGIT

Size	Variabel	Step
0	EBITA, STA, ITR, DSI, NPM, OPM, DAR, ETD, BI7DRR	9
1	EBITA, STA, ITR, ROE, NPM, OPM, DER, DAR, QR, RGDPG	8
2	EBITA, STA, DAR, CR, QR, ETD, WCA, RGDPG, USD/IDR	9
3	EBITA, STA, OPM, DER, DAR, QR, WCA, BI7DRR, USD/IDR, Inflasi	14

Menggunakan rasio-rasio keuangan dan variabel makroekonomi yang dapat memprediksi kondisi keuangan perusahaan didapatkan model pada Tabel 4.34. Model yang digunakan dibentuk menggunakan data *training* pada setiap *size* data.

**Tabel 4.34** Model Prediksi *Financial Distress* LOGIT secara *Stepwise*

Size	Model
0	$g(x) = -0,229 + 28,755\text{EBITA} + 0,746\text{STA} - 0,0145\text{ITR} - 0,002686\text{DSI} + 8,443\text{NPM} + 25,20\text{OPM} - 1,896\text{DAR} + 6,8155\text{ETD} - 0,09852\text{BI7DRR}$
1	$g(x) = -1,039 + 24,08\text{EBITA} + 0,314\text{STA} - 0,00048\text{ITR} - 0,3065\text{ROE} + 0,8352\text{NPM} + 1,0150\text{OPM} - 0,03328\text{DER} - 1,462\text{DAR} - 0,02883\text{QR} - 0,2661\text{RGDPG}$

**Tabel 4.34** Model Prediksi *Financial Distress* LOGIT secara *Stepwise* (Lanjutan)

Size	Model
2	$g(x) = 4,175 + 16,67\text{EBITA} + 0,2612\text{STA} - 0,833\text{DAR} + 0,07436\text{CR} - 0,1931\text{QR} + 0,4742\text{ETD} + 0,9842\text{WCA} - 0,3614\text{RGDPG} - 0,0001858\text{USD_IDR}$
3	$g(x) = -3,345 + 11,39\text{EBITA} + 0,2472\text{STA} + 1,4870\text{PM} - 0,01966\text{DER} - 0,6481\text{DAR} - 0,02442\text{QR} + 0,7934\text{WCA} + 0,2167\text{BI7DRR} + 0,0003313\text{USD_IDR} - 0,132\text{Inflasi}$

Model-model pada Tabel 4.34 didapat menggunakan data *training* untuk setiap *size* data. Prediksi kondisi keuangan perusahaan dilakukan baik untuk data *training* maupun data *testing*. *Confussion matrix* hasil klasifikasi dan kondisi keuangan perusahaan aktual diberikan pada Tabel 4.35.

**Tabel 4.35a** *Confussion Matrix* Model *Stepwise* LOGIT *Size 0*

Size 0					
<b>Training</b>	Aktual		<b>Testing</b>	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1173	40	0	295	9
1	70	312	1	22	109

**Tabel 4.35b** *Confussion Matrix* Model *Stepwise* LOGIT *Size 1*

Size 1					
<b>Training</b>	Aktual		<b>Testing</b>	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1040	61	0	283	21
1	154	195	1	42	89

**Tabel 4.35c** *Confussion Matrix* Model *Stepwise* LOGIT *Size 2*

Size 2					
<b>Training</b>	Aktual		<b>Testing</b>	Aktual	
Prediksi	0	1	Prediksi	0	1
0	941	47	0	275	29
1	170	147	1	50	81

**Tabel 4.35d Confussion Matrix Model Stepwise LOGIT Size 3**  
**Size 3**

<b>Training</b>		<b>Aktual</b>		<b>Testing</b>		<b>Aktual</b>	
Prediksi		0	1	Prediksi		0	1
0		829	43	0		299	1
1		181	107	1		111	20

Berdasarkan *confussion matrix* prediksi kondisi keuangan dengan kondisi keuangan aktual perusahaan didapatkan ketepatan klasifikasi yang dihitung melalui AUC dan Accuracy dari model *Binary Logistic Regression* menggunakan prosedur *stepwise*. Berikut pada Tabel 4.36 akan diberikan AUC dan Accuracy masing-masing model pada data *training* dan data *testing*.

**Tabel 4.36 Ketepatan Klasifikasi Model Stepwise LOGIT**

<b>Size</b>	<b>Ukuran</b>	<b>Training</b>	<b>Testing</b>
0	AUC	0,8919	0,9012
	Accuracy	0,9310	0,9287
1	AUC	0,7517	0,8052
	Accuracy	0,8517	0,8552
2	AUC	0,7081	0,7615
	Accuracy	0,8337	0,8184
3	AUC	0,6611	0,5681
	Accuracy	0,8069	0,7333

Data *training* pada *size 0* memiliki AUC lebih kecil dan Accuracy yang lebih besar daripada data *testing*. AUC dan Accuracy pada data *size 0* adalah terbesar daripada pada data *size* lainnya. Sehingga dengan menggunakan model *Binary Logistic Regression* secara *stepwise*, model yang paling baik dalam memprediksi kondisi keuangan perusahaan adalah model dengan rasio keuangan dan cariabel makroekonomi pada periode yang sama.

#### 4.4.3 Lasso Logistic Regression

Nilai parameter *regularized parameter* yaitu  $\lambda$  diperoleh dengan melakukan percobaan hingga didapat model terbaik. Hasil model terbaik pada masing-masing *size* data disajikan pada Tabel 4.39 yang diukur berdasarkan AUC dan *Accuracy* pada data *training* dan data *testing*. Prediktor yang masuk dalam model menggunakan metode *Lasso Logistic Regression* ditunjukkan pada Tabel 4.37.

**Tabel 4.37** Variabel Model *Lasso LOGIT*

Size	Variabel
0	EBITA, STA, ITR, DSI, NPM, OPM, DAR, QR, ETD, WCA, BI7DRR, USD/IDR
1	EBITA, STA, ITR, OPM, DAR, CR, ETD, WCS, RGDPG, USD/IDR
2	EBITA, STA, ITR, NPM, DAR, QR, ETD, WCA, USD/IDR
3	EBITA, STA, ITR, DAR, WCA

Koefisien dari masing-masing prediktor diberikan pada Lampiran 21. Setiap model dibentuk menggunakan data *training*. Hasil prediksi menggunakan model *Lasso Logistic Regression* baik pada data *training* maupun data *testing* ditunjukkan melalui *Confussion Matrix* pada Tabel 4.38. Sehingga melalui hasil prediksi dapat diketahui model dan data *size* yang paling tepat untuk mengetahui kondisi keuangan perusahaan.

**Tabel 4.38a** *Confussion Matrix* Model *Lasso LOGIT Size 0*

Size 0					
<b>Training</b>	Aktual		<b>Testing</b>	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1181	32	0	291	13
1	71	311	1	19	112

**Tabel 4.38b** *Confussion Matrix* Model *Lasso LOGIT Size 1*

Size 1					
<b>Training</b>	Aktual		<b>Testing</b>	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1047	54	0	283	21
1	178	171	1	40	91

**Tabel 4.38c** Confussion Matrix Model Lasso LOGIT Size 2

Size 2			
<b>Training</b>	Aktual	<b>Testing</b>	Aktual
Prediksi	0 1	Prediksi	0 1
0	951 37	0	288 16
1	194 123	1	68 63

**Tabel 4.38d** Confussion Matrix Model Lasso LOGIT Size 3

Size 3			
<b>Training</b>	Aktual	<b>Testing</b>	Aktual
Prediksi	0 1	Prediksi	0 1
0	848 24	0	298 6
1	219 69	1	97 34

**Tabel 4.39** Ketepatan Klasifikasi Model Lasso LOGIT

Size	$\lambda$	<i>Training</i>		<i>Testing</i>	
		Accuracy	AUC	Accuracy	AUC
0	0,00097	0,9354	0,8939	0,9264	0,9061
1	0,00693	0,8400	0,7205	0,8598	0,8128
2	0,0104	0,8230	0,6753	0,8069	0,7141
3	0,02685	0,7905	0,6060	0,7632	0,6199

Accuracy yang dihasilkan data *training* pada data *size 0* lebih besar daripada Accuracy data *testing size 0*. Sedangkan untuk AUC, data *testing* lebih besar daripada data *training* pada data *size 0*. Accuracy yang dihasilkan pada *size 0* baik pada data *training* ataupun data *testing* lebih besar daripada 0,9 sehingga untuk Accuracy termasuk dalam kategori *excellent classification*. Seperti pada metode-metode sebelumnya, bahwa permodelan menggunakan *lasso logistic regression* menghasilkan AUC dan Accuracy tertinggi pada data *size 0* baik untuk data *training* maupun data *testing*.

#### 4.4.4 Elastic-Net Logistic Regression

Penelitian ini menggunakan nilai parameter yang dicobakan adalah  $\gamma=[0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9]$  dan parameter  $\lambda$  dicoba hingga mendapat nilai AUC tertinggi. Kedua parameter tersebut dikombinasi untuk mendapatkan nilai AUC pada data *training* paling tinggi sehingga didapatkan parameter yang paling optimum. Permodelan dilakukan menggunakan data *training* pada masing-masing data *size*. Tabel 4.40 menunjukkan rasio keuangan dan variabel makroekonomi yang masuk dalam model.

**Tabel 4.40** Variabel Model *Elastic-Net LOGIT*

Size	Variabel
0	EBITA, STA, ITR, DSI, ROE, NPM, OPM, DAR, QR, ETD, WCA, BI7DRR, USD/IDR
1	EBITA, STA, ITR, DSI, NPM, OPM, DER, DAR, CR, ETD, WCS, RGDPG, BI7DRR, USD/IDR
2	EBITA, STA, ITR, DSI, NPM, DER, DAR, CR, QR, ETD, WCA, WCS, RGDPG, BI7DRR, USD/IDR, Inflasi
3	EBITA, STA, ITR, DSI, OPM, DER, DAR, CR, QR, ETD, WCA, RGDPG, BI7DRR, USD/IDR/ Inflasi

Predksi menggunakan masing-masing model menghasilkan ketepatan klasifikasi yang berbeda-beda.. *Confussion matrix* dari hasil prediksi menggunakan model dari metode *elastic-net logistic regression* diberikan pada Tabel 4.41.

**Tabel 4.41a** *Confussion Matrix Model Elastic-Net LOGIT Size 0*

Size 0					
<b>Training</b>	Aktual		<b>Testing</b>	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1182	31	0	291	13
1	69	313	1	19	112

**Tabel 4.41b** *Confussion Matrix Model Elastic-Net LOGIT Size 1*

Size 1					
<b>Training</b>	Aktual		<b>Testing</b>	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1046	55	0	283	21
1	179	170	1	40	91

**Tabel 4.41c Confussion Matrix Model Elastic-Net LOGIT Size 2  
Size 2**

<b>Training</b>		<b>Aktual</b>		<b>Testing</b>		<b>Aktual</b>	
Prediksi		0	1	Prediksi		0	1
0		948	40	0		274	30
1		182	135	1		50	81

**Tabel 4.41d Confussion Matrix Model Elastic-Net LOGIT Size 3  
Size 3**

<b>Training</b>		<b>Aktual</b>		<b>Testing</b>		<b>Aktual</b>	
Prediksi		0	1	Prediksi		0	1
0		842	30	0		300	4
1		194	94	1		100	31

Berdasarkan prediksi klasifikasi yang telah diberikan maka dapat dihitung ketepatan klasifikasinya menggunakan AUC dan Accuracy yang diberikan di Tabel 4.42. Prediksi dilakukan menggunakan data *training* dan data *testing* untuk setiap size data.

**Tabel 4.42 Ketepatan Klasifikasi Model Elastic-Net LOGIT**

<b>Size</b>	$\lambda$	$\gamma$	<b>Training</b>		<b>Testing</b>	
			Accuracy	AUC	Accuracy	AUC
0	0,0013	0,7	0,9373	0,8928	0,9264	0,9061
1	0,0047	0,4	0,8386	0,7186	0,8690	0,8128
2	0,0058	0,1	0,8299	0,6927	0,8161	0,7598
3	0,0116	0,3	0,8069	0,6460	0,7609	0,6117

AUC dan Accuracy prediksi dari model *elastic-net logistic regression* pada data *size 0* adalah yang paling tinggi diantara data *size* lainnya baik pada data *training* maupun data *testing*. Sehingga dengan menggunakan metode *elastic-net logistic regression*, prediksi kondisi keuangan terbaik diberikan oleh model dengan

rasio keuangan perusahaan dan variabel makroekonomi pada periode tahun yang sama dengan kondisi keuangan.

## 4.5 Perbaikan Kebaikan Model

Melalui hasil prediksi dapat dilihat besar keakuratan model dalam melakukan prediksi baik untuk data *training* maupun data *testing*. Model prediksi terbaik dapat dilihat melalui besar AUC dan *Accuracy* yang dihasilkan dari setiap hasil prediksi data *training* dan data *testing*.

### 4.5.1 Perbandingan Kebaikan Model Serentak

Permodelan serentak dilakukan menggunakan metode *Support Vextor Machine*, *Generalized Extreme Value Regression*, dan *Logistic Regression* dengan mengikutsertakan seluruh prediktor. Tabel 4.43 di bawah adalah AUC dan *Accuracy* dari masing-masing model.

**Tabel 4.43** Perbandingan Akurasi Model Serentak

Metode	Size	AUC		Accuracy	
		Training	Testing	Training	Testing
SVM	0	0,9412	0,9200	<b>0,9423</b>	0,9195
	1	0,9368	0,8249	0,8483	0,8460
	2	0,8263	0,7855	0,8337	0,8000
	3	0,8347	0,5922	0,8276	0,6230
GEVR	0	<b>0,9760</b>	<b>0,9708</b>	0,9367	<b>0,9241</b>
	1	0,8947	0,9280	0,8586	0,8667
	2	0,8721	0,8782	0,8414	0,8207
	3	0,8494	0,8288	0,8233	0,7954
LOGIT	0	0,8949	0,8985	0,9329	0,9218
	1	0,7417	0,8209	0,8455	0,8621
	2	0,7085	0,7798	0,8360	0,8046
	3	0,6692	0,5545	0,8138	0,7264

Telah ditunjukkan pada Tabel 4.43 bahwa pada model serentak menggunakan metode-metode yang digunakan memiliki AUC dan *Accuracy* tertinggi pada data *size* 0. Agar dapat dilihat lebih mudah pada Gambar 4.22 ditunjukkan perbedaan ketepatan klasifikasi

menggunakan AUC dan *Accuracy* pada masing-masing metode yang digunakan.



**Gambar 4.22a** AUC Model Serentak



**Gambar 4.22b** Accuracy Model Serentak

Dapat dilihat bahwa baik AUC maupun *Accuracy* data *train* lebih besar daripada data *test* kecuali pada AUC metode *logistic regression*. Ketepatan klasifikasi yang dihasilkan relatif tinggi. Metode *Generalized Extreme Value Regression* memberikan ketepatan klasifikasi tertinggi berdasarkan nilai AUC dan berdasarkan *Accuracy*, ketepatan klasifikasi terbaik diberikan oleh metode *Support Vector Machine*. Namun, pada model yang dibentuk terdapat variabel-variabel yang tidak signifikan berpengaruh terhadap model.

Model secara serentak klasifikasi perusahaan yang mengalami *financial distress* dan tidak paling baik diberikan oleh model klasifikasi data *size 0* dengan metode *Generalized Extreme Value Regression*. Pemilihan model didasarkan pada AUC data *training*

maupun data *testing*, dimana model data *size* 0 dengan metode *Generalized Extreme Value Regression* memberikan AUC paling tinggi. Rasio keuangan yang signifikan pada model *Generalized Extreme Value Regression* secara serentak adalah EBITA, STA, ITR, NPM, OPM, DAR, dan ETD sehingga pada model ini tidak terdapat variabel makroekonomi yang signifikan. Pemilihan model terbaik menggunakan nilai AUC dikarenakan AUC lebih robust dibandingkan dengan *Accuracy* karena diukur dengan mempertimbangkan kesalahan klasifikasi pada kelas minoritas pada kelas mayoritas.

#### 4.5.2 Perbandingan Kebaikan Model dengan Seleksi Variabel

Permodelan secara serentak dengan menggunakan seluruh prediktor pada model menghasilkan adanya prediktor yang tidak signifikan, maka dari itu dilakukan permodelan dengan seleksi variabel. Seleksi variabel yang dilakukan pada metode *Support Vector Machine* adalah dengan dilakukannya regularisasi dengan metode *Lasso* dan *Elastic-Net* yang juga diterapkan pada metode *Logistic Regression*. Sedangkan untuk metode *Generalized Extreme Value Regression* adalah menggunakan *Backward Elimination* yang juga diterapkan pada metode *Logistic Regression*. Berikut Tabel 4.44 merupakan tabel ketepatan klasifikasi menggunakan model dengan seleksi variabel.

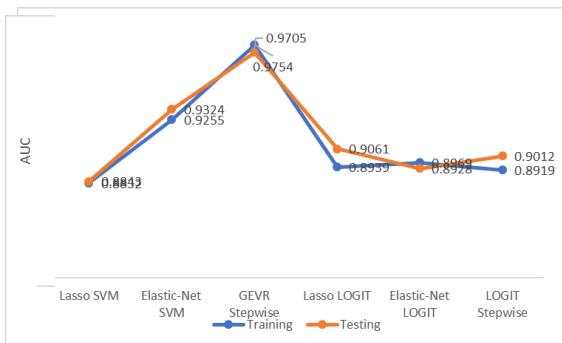
**Tabel 4.44** Perbandingan Akurasi Model dengan Seleksi Variabel

Metode	Size	AUC		Accuracy	
		Training	Testing	Training	Testing
<i>Lasso SVM</i>	0	0,8832	0,8843	0,9179	0,9080
	1	0,6348	0,6319	0,8262	0,7678
	2	0,6023	0,5998	0,7985	0,7563
	3	0,5759	0,5638	0,7802	0,7333
<i>Elastic-Net SVM</i>	0	0,9255	0,9324	0,9066	0,9218
	1	0,8161	0,8478	0,8448	0,8598
	2	0,8072	0,7675	0,7954	0,7494
	3	0,7768	0,7016	0,7853	0,7103

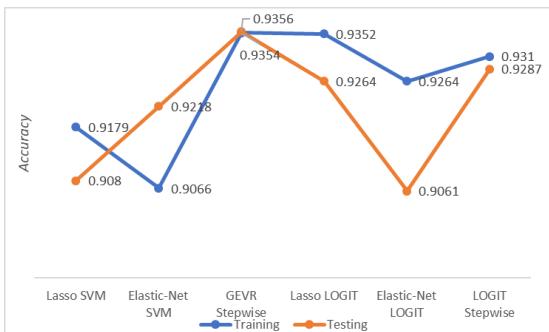
**Tabel 4.44** Perbandingan Akurasi Model dengan Seleksi Variabel (Lanjutan)

Metode	Size	AUC		Accuracy	
		Training	Testing	Training	Testing
<i>Backward GEVR</i>	0	<b>0,9754</b>	<b>0,9705</b>	<b>0,9354</b>	<b>0,9356</b>
	1	0,8886	0,9263	0,8531	0,8667
	2	0,8699	0,8913	0,8398	0,8299
	3	0,8415	0,8500	0,8241	0,8092
<i>Lasso LOGIT</i>	0	0,8939	0,9061	0,9352	0,9264
	1	0,7205	0,8128	0,8400	0,8598
	2	0,6753	0,7141	0,8230	0,8069
	3	0,6060	0,6199	0,7905	0,7632
<i>Elastic-Net LOGIT</i>	0	0,8969	0,8928	0,9264	0,9061
	1	0,7186	0,8302	0,8386	0,8690
	2	0,6927	0,7598	0,8299	0,8161
	3	0,6460	0,6117	0,8069	0,7609
<i>Stepwise LOGIT</i>	0	0,8919	0,9012	0,9310	0,9287
	1	0,7517	0,8052	0,8517	0,8552
	2	0,7081	0,7615	0,8337	0,8184
	3	0,6611	0,5681	0,8069	0,7333

Sama dengan yang terjadi pada model serentak, akurasi pada model dengan seleksi variabel memiliki nilai AUC dan Accuracy tertinggi pada data size 0 untuk model menggunakan setiap metode. Metode yang memberikan ketepatan klasifikasi terbaik akan ditunjukkan pada Gambar 4.23 berikut.



Gambar 4.23a AUC Model dengan Seleksi Variabel



Gambar 4.23a Accuracy Model dengan Seleksi Variabel

Ketepatan klasifikasi berdasarkan nilai AUC menghasilkan kesimpulan bahwa klasifikasi terbaik baik menggunakan data *train* maupun data *test* diberikan oleh metode *Backward Elimination Generalized Extreme Value Regression*. Hal yang sama yaitu menggunakan nilai *Accuracy* ketepatan klasifikasi tertinggi data *train* diberikan dan data *test* adalah dengan metode *Backward Elimination Generalized Extreme Value Regression*.

Sama halnya dengan pemilihan model serentak, pada model dengan seleksi variabel dipilih model terbaik adalah model dengan AUC data *testing* dan data *training* tertinggi yaitu metode *Generalized Extreme Value Regression* dengan *backward elimination* pada data *size* 0. Rasio keuangan dan variabel makroekonomi yang signifikan pada model tersebut adalah EBITA, STA, ITR, NPM, OPM, DAR, QR, ETD, dan BI7DRR.

## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1 Kesimpulan**

Berdasarkan analisis yang telah dilakukan, diperoleh kesimpulan sebagai berikut:

1. Beberapa rasio keuangan perusahaan menunjukkan bahwa perusahaan yang mengalami *financial distress* memiliki rasio keuangan yang berbeda dengan perusahaan yang tidak mengalami *financial distress*. Perbedaan yang ditunjukkan untuk data pada masing-masing *size* pada beberapa rasio keuangan terlihat dengan jelas dan pada beberapa rasio keuangan tidak menunjukkan perbedaan antar *size* data. Akibatnya pada permodelan menggunakan metode yang sama akan ditemukan rasio keuangan signifikan yang berbeda antar *size* data. Sedangkan variabel makroekonomi digunakan pada penelitian untuk mengakomodasi dependensi antar perusahaan dan antar periode.
2. Prediksi klasifikasi perusahaan yang mengalami *financial distress* dan tidak dilakukan menggunakan permodelan secara serentak dan dengan seleksi variabel. Prediksi dengan model serentak menggunakan metode *Support Vector Machine*, *Generalized Extreme Value Regression*, dan *Logistic Regression* masing-masing menunjukkan bahwa data *size 0* merupakan skenario terbaik untuk memprediksi *financial distress* pada perusahaan. Menggunakan AUC baik data *train* atau data *test*, metode *Generalized Extreme Value Regression* memberikan ketepatan klasifikasi tertinggi. Sedangkan dengan menggunakan *Accuracy*, ketepatan klasifikasi data *train* tertinggi adalah klasifikasi yang dihasilkan metode *Logistic Regression* dan ketepatan klasifikasi data *test* tertinggi ditunjukkan oleh metode *Support Vector Machine*. Permodelan dengan seleksi variabel dilakukan dengan enam metode yaitu *Lasso Support Vector Machine*, *Elastic-Net Support Vector Machine*, *Backward Generalized Extreme Value Regression*, *Lasso Logistic Regression*, *Elastic-Net Logistic Regression*, dan *Stepwise Logistic Regression*. Sama halnya dengan permodelan

secara serentak, bahwa model terbaik diberikan pada data *size* 0. Berdasarkan AUC dan *Accuracy* model dengan seleksi variabel terbaik baik data *train* dan data *test* diberikan oleh metode *Backward Generalized Extreme Value Regression*.

3. Rasio keuangan yang signifikan pada model *Generalized Extreme Value Regression* secara serentak adalah EBITA, STA, ITR, NPM, OPM, DAR, dan ETD sehingga tidak terdapat variabel makroekonomi yang signifikan terhadap model. Sedangkan dengan adanya seleksi variabel pada metode *Generalized Extreme Value Regression* menggunakan prosedur *backward elimination* hingga 9 step menghasilkan 10 prediktor signifikan dalam model, diantaranya adalah EBITA, STA, ITR, NPM, OPM, DAR, QR, ETD, dan BI7DRR.

## 5.2 Saran

Setelah dilakukan penelitian prediksi *financial distress* perusahaan sektor industri di Indonesia menggunakan beberapa metode terdapat hal-hal yang harus diperhatikan dan dijadikan saran kepada peneliti selanjutnya. Hal tersebut adalah pada beberapa metode yaitu *Generalized Extreme Value Regression*, *Lasso Support Vector Machine*, *Elastic-Net Support Vector Machine*, *Backward Generalized Extreme Value Regression*, *Lasso Logistic Regression*, dan *Elastic-Net Logistic Regression* memerlukan *tunning* parameter lebih awal untuk mendapatkan model terbaik. Maka akan lebih baik jika menggunakan nilai parameter yang lebih banyak dicobakan untuk melakukan permodelan. Selain itu, pada penelitian selanjutnya disarankan untuk melakukan perbandingan prediksi *financial distress* yang diindikasikan dengan ICR kurang dari satu atau ROA negatif dengan *financial distress* diindikasikan oleh ICR kurang dari satu dan ROA negatif.

Berdasarkan rasio keuangan yang signifikan mempengaruhi *financial distress* didapat bahwa seluruh rasio keuangan baik rasio aktivitas, profitabilitas, solvabilitas, dan likuiditas terwakili oleh tiap rasio keuangan yang termasuk. Sehingga seluruh rasio keuangan tersebut memiliki andil untuk menentukan kondisi keuangan perusahaan, maka dari itu perusahaan harusnya selalu memperhatikan seluruh rasio keuangan yang ada. Selain itu pada

model dengan seleksi variabel didapat bahwa salah satu variabel makroekonomi berpengaruh terhadap kondisi keuangan perusahaan industri, maka kondisi perubahan keuangan global seharusnya diberikan perhatian yang lebih agar perusahaan mampu menghadapi perubahan kondisi keuangan global yang akan terjadi.

*(Halaman ini sengaja dikosongkan)*

## **DAFTAR PUSTAKA**

- Abdulkadir, M. (2000). *Hukum Perusahaan Indonesia*. Bandung: Citra Aditya Bakti.
- Agresti, A. (2002). *Categorical Data Analysis (2nd ed)*. New Jersey: John Wiley & Sons Inc.
- Agusti, C. P. (2013). *Analisis Faktor yang Mempengaruhi Kemungkinan Terjadinya Financial Distress*. Semarang.
- Almilia, L. S., & Kritjadi. (2003). Analisis Rasio Keuangan untuk Memprediksi Kondisi Financial Distress Perusahaan Manufaktur yang Terdaftar di Bursa Efek Jakarta. *Jurnal Akutansi dan Auditing Indonesia*, 7(2).
- Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance*, 23, 589-609.
- Anggraeni, R. (2020, January 6). *Industri Manufaktur Penyumbang Terbesar Ekspor Nasional*. Diambil kembali dari Sindonews.com: <https://ekbis.sindonews.com/read/1488817/34/industri-manufaktur-penyumbang-terbesar-ekspor-nasional-1578305339>
- Argo, A. (2010). *Analisis Pengaruh Rasio Keuangan terhadap Prediksi Kondisi Bermasalah pada Bank (Studi Kasus pada Bank Umum Swasta Nasional Periode 2004-2007)*. Semarang: Jurusan Akutansi Fakultas Ekonomi Universitas Diponegoro.
- Asquith, P., Gertner, R., & Scharfstein, D. (1994). Anatomy of Financial Distress: An Examination of Junk-Bond Issuers. *Quarterly Journal of Economics*, 109, 1189-1222.
- Astuti, P., & Pamudji, S. (2015). Analisis Pengaruh Opini Going Concern, Likuiditas, Solvabilitas, Arus Kas, Umur Perusahaan dan Ukuran Perusahaan terhadap

- Kemungkinan Financial Distress. *Diponegoro Journal of Accounting*, 4(1).
- Bank Indonesia. (2017). *Metadata: Indikator Pengedaran Uang*. Jakarta: Bank Indonesia.
- Bradley, P. S., & Mangasarian, O. L. (1998). Feature Selection via Concave Minimization and Support Vector Machine. *The Fifteenth International Conference on Machine Learning (ICML)*.
- Calabrese, R., & Giudici, P. (2015). Estimating Bank Default with Generalised Extreme Value Regression Models. *Journal of The Operational Research*.
- Calabrese, R., & Osmetti, S. A. (2013). Modelling Small and Medium Enterprise Loan Defaults as Rare Even: The Generalized Extreme Value Regression Model. *Journal of Applied Statistics*.
- Classens, R., Djankov, S., & Klapper. (1999). Resolution of Corporate Distress: Evidence from East Asia's Financial Crisis. *The Annual World Bank Group-Brooking*. New York: Institution Conference Palisades.
- Dewi, S. (2016). Komparasi 5 Metode Algoritma Klasifikasi Data Mining pada Prediksi Keberhasilan Pemasaran Produk Layanan Perbankan. *Jurnal Techno Nusa Mandiri*, 8(1).
- Dharmawan, K. (2012). Estimasi Nilai VaR Dinamis Indeks Saham Menggunakan PeakOver Threshold dan Block Maxima. *Jurnal Matematika*, 2(2), 1-12.
- Escobar, C. E., & Menedez, R. M. (2019). Process Monitoring for Quality a Model Selection Criterion for Shallow Neural Networks. *Proceedings of the Annual Conference of the PHM Society, At Scottsdale, AZ*, 11.
- Firdaus, A. (2012). *Pengaruh Rasio Likuiditas, Rasio Aktivitas, Rasio Profitabilitas dan Rasio Pasar terhadap Harga*

- Saham Syariah pada Industri Barang Konsumsi yang Terdaftar di Daftar Efek Syariah Tahun 2008-2011.* Universitas Islam Sunan Kalijaga, Program Studi Keuangan Islam, Yogyakarta.
- Fitriyah, I., & Hariyati. (2013). Pengaruh rasio keuangan terhadap financial distress pada perusahaan properti dan real estate. *Jurnal Akutansi Kontemporer*, 1(3), 760-773.
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Path for Generalized Linear Models via Coordinate Descent. *Journal of Statistics Software*, 1-22.
- Gilli, M., & Kellezi, E. (2006). An Application of Extreme Value Theory for Measuring Financial Risk. *Computational Economic*, 207-228.
- Gorunescu, F. (2011). *Classification Performance Evaluation. In: Data Mining*. New York: Springer, Berlin, Heidelberg.
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concept and Techniques Third Edition*. New York: Elsevier.
- Hapsari, A. R. (2015). *Financial Distress, Corporate Governance dan Karakteristik Perusahaan terhadap Pengungkapan Sukarela pada Laporan Tahunan Perusahaan*. Jakarta.
- Härdle, W. K., & Prastyo, D. D. (2014). Embedded Predictor Selection for Default Risk Calculation: a Shoutheast Asian Industry Study. In D.L. Chuen, & G.N. Grgoriou (Eds). *Financial Markets and Sovereign Wealth Funds*, 131-148.
- Härdle, W. K., Prastyo, D. D., & Hafner, C. M. (2014). Support Vector Machines with Evolutionary Model Selection for Default Prediction, In J.S. Racine, L.Su, & A. Ullah (Eds.). *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, 346-373.
- Härdle, W., Moro, R. A., & Schäfer, D. (2005). *Predicting Bankruptcy with Support Vector Machine*.

- Hill, N. T., Perry, S. E., & Andes, S. (1996). Evaluating Firms in Financial Distress: An Event Histiry Analysis. *Journal of Applied Bussiness Research*, 12(3), 60-71.
- Hosmer, D. W., & Lemshow, S. (2000). *Applied Logistic Regression*, 2nd ed. New York: John Wiley and Sons.
- Huang, J., & Ling, C. X. (2005). Using AUC and Accuracy in Evaluating Learning Algorithms. *IEEE Transactions on Knowldge and Data Enginering*, 299-310.
- IDX. (2016, February 15). *Indeks Harga Saham Gabungan*. Diambil kembali dari PT Bursa Efek Indonesia: <http://www.idx.co.id/id-id/beranda/informasi/bagiinvestor/indeks.aspx>
- IDX. (2020, January 10). *Profil Perusahaan Tercatat*. Diambil kembali dari PT Bursa Efek Indonesia: <https://www.idx.co.id/perusahaan-tercatat/profil-perusahaan-tercatat/>
- Indasari, B. (2018). *Studi Simulasi dan Analisis Survival Delisting Time di Bursa Efek Indonesia untuk Perusahaan Manufaktur dengan Metode Multiperiod Generalized Extreme Value Regression*. Insitut Teknologi Sepuluh Nopmeber, Program Studi Sarjana Departemen Statistika, Surabaya.
- Irianto, Agus. (2012). *Statistik Konsep Dasar, Aplikasi, dan Pengembangannya*. Jakarta: Kencana.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Application in R*. USA: Springer.
- Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis Sixth Edition*. New Jersey: Pearson Prentice Hall.
- Klinkenberg, R. (2013). Using Labeled and Unlabeled Data to Learn Drifting Concepts.

- Kowarik, A., & Templ, M. (2016). Imputation with the R Package VIM. *Journal of Statistical Software*, 74(7), 1-16.
- Kurniasari, C., & Ghozali, I. (2013). Analisis Pengaruh Rasio CAMEL dalam Memprediksi Financial Distress Perbankan di Indonesia. *Diponegoro Journal of Accounting*, 2.
- Lambrou, A., Papadopoulos, H., Nouretdinov, I., & Gammerman, A. (2012). Reliable Probability Estimates Based on Support Vector Machines for Large Multiclass Datasets. *8th International Conference on Artificial Intelligence Applications and Innovations (AIAI)*, 182-191.
- Lau, A. H. (1987). A Five State Financial Distress Prediction Model. *Journal of Accounting Research*, 25, 127-138.
- Leeuw, J. d., Hornik, K., & Mair, P. (2009). Isotone Optimization in R: POOL-Adjacent-Violators Algorithm (PAVA) and Active Set Methods. *Journal of Statistical Software*.
- Prastyo, D. D. (2015). Regularized Support Vector Machines with Application to Corporate Default Prediction. *Manuscript*.
- Prihadi, T. (2010). *Analisis Laporan Keuangan Teori dan Aplikasinya*. Jakarta: PPM Manajemen.
- Ross, Westerfield, & Jaffe. (2002). *Corporate Finance Sixth Edition*. The McGraw-Hill Companies.
- Sawir, A. (2005). *Analisis Kinerja Keuangan dan Perencanaan Keuangan Perusahaan*. Jakarta: PT Gramedia Pustaka Utama.
- Schölkopf, B., & Smola, A. J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. Cambridge: MIT Press.

- SirElKhatim, M., & Salim, N. (2017). Prediction of Banks Financial Distress. *Journal of Engineering and Computer Science*, 1.
- Sjam, A. A., & Dharma, A. G. (2010). *Rasio-rasio Keuangan, Analisa Diskriminan, dan Prediksi Probabilitas Kegagalan Perusahaan-perusahaan pada Industri Properti dan Real Estate yatu Listing di Bursa Efek Indonesia*. Bandung: Universitas Kristen Maranatha.
- Sujarweni, V. W. (2016). *Pengantar Akutansi*. Yogyakarta: Pustaka Baru Press.
- Theodorus, S., & Artini, L. S. (2018). Studi Financial Distress pada Perusahaan Perbankan di BEI. *E-Journal Manajemen Unud*, 7(5), 2710-2732.
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, 267-288.
- Tobing, S. (2019, May 6). *Sektor Industri Masih Penyumbang Terbesar Pertumbuhan Ekonomi*. Diambil kembali dari katadata.co.id:<https://katadata.co.id/berita/2019/05/06/sektor-industri-masih-penyumbang-terbesar-pertumbuhan-ekonomi>
- Wang, X., & Dey, D. K. (2010). Generalized Extreme Value Regression for Binary Response Data: An Application to B2B Electronic Payments System Adoption. *The Annals of Applied Statistics*, 4(4), 2000-2021.
- Warren, C. S., Reeve, J. M., Duchac, J. E., Suhardianto, N., Kalanjati, D. S., Jusuf, A. A., & Djakman, C. D. (2017). *Pengantar Akutansi Adaptasi Indonesia Edisi 25*. Jakarta Selatan: Penerbit Salemba Empat.

- Whardani, R. (2006). Mekanisme Corporate Governance daam Perusahaan yang Mengalami Permasalahan Keuangan (Financial Distressed Firms). *Makalah SNA IX*.
- Whitaker, R. B. (1999). The Early Stages of Financial Distress. *Journal of Economics and Finance*, 23, 123-133.
- Widyarani, A. (2018). *Prediksi Financial Distress Bank Umum di Indonesia dengan Metode Generalized Extreme Value Regression, Regresi Logistik, dan Analisis Diskriminan Kernel*. Institut Teknologi Sepuluh Nopember, Program Studi Departemen Statistika, Surabaya.
- Witten, I. H., Fraink, E., & Hall, M. A. (2011). *Data Mining Practical Machine Learning Tools and Technique Third Edition*. New York: Elsavier.
- Wu, Q., & Wang, W. (2013). Piecewise-Smooth Support Vector Machine for Clasification. *Mathematical Problems in Engineering*(7).
- Zhu, J., Rosset, S., Hastie, T., & Tibshirani, R. (2004). 1-norm Support Vector Machines. *Advantages in Neural Information Processing System 16*.
- Zou, H., & Hastie, T. (2005). Regularization and Vaeriable Selection via The Elastic Net. *Journal of The Royal Statistics Society*, 301-320.

*(Halaman ini sengaja dikosongkan)*

## LAMPIRAN

**Lampiran 1.** Data Penelitian

PERIODE	KODE	EBITA	STA	...	USD/IDR	Inflasi	FD
2005Q4	ABBA	-0.05	0.72	...	9830	17.11	1
2006Q4	ABBA	0.06	0.68	...	9020	6.6	0
2007Q4	ABBA	0.07	0.72	...	9419	6.59	0
2008Q4	ABBA	0.06	0.73	...	10950	11.06	0
2009Q4	ABBA	0.03	0.66	...	9400	2.78	0
2010Q4	ABBA	0.01	0.56	...	8991	6.96	0
2011Q4	ABBA	0.01	0.60	...	9068	3.79	1
2012Q4	ABBA	0.03	0.61	...	9670	4.3	0
2013Q4	ABBA	0.02	0.69	...	12189	8.38	1
2014Q4	ABBA	-0.02	0.71	...	12440	8.36	1
2015Q4	ABBA	-0.04	0.67	...	13795	3.35	1
2016Q4	ABBA	-0.05	0.70	...	13436	3.02	1
2017Q4	ABBA	-0.02	0.63	...	13548	3.61	1
2018Q4	ABBA	-0.08	0.44	...	14481	3.13	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2005Q4	YPAS	0.17	1.67	...	9830	17.11	0
2006Q4	YPAS	0.17	1.70	...	9020	6.6	0
2007Q4	YPAS	0.17	1.70	...	9419	6.59	0
2008Q4	YPAS	0.19	1.67	...	10950	11.06	0
2009Q4	YPAS	0.17	1.52	...	9400	2.78	0
2010Q4	YPAS	0.17	1.74	...	8991	6.96	0
2011Q4	YPAS	0.12	1.68	...	9068	3.79	0
2012Q4	YPAS	0.10	1.37	...	9670	4.3	0
2013Q4	YPAS	0.06	0.88	...	12189	8.38	0
2014Q4	YPAS	0.02	1.19	...	12440	8.36	1
2015Q4	YPAS	0.00	0.93	...	13795	3.35	1
2016Q4	YPAS	0.00	0.95	...	13436	3.02	1
2017Q4	YPAS	-0.01	1.04	...	13548	3.61	1
2018Q4	YPAS	0.01	1.30	...	14481	3.13	1

**Lampiran 2.** *Syntax* Imputasi Data

```
library(VIM)
#Imputation
ABBA=kNN(data[1:56,-8])
ADES=kNN(data[57:112,-8])
ADMG=kNN(data[113:168,-8])
AKPI=kNN(data[167:224,-8])
ALDO=kNN(data[225:280,-8])
:
WTON=kNN(data[9073:9128,-8])
YPAS=kNN(data[9129:9184,-8])
WIIM=kNN(data[9185:9240,-8])
WOOD=kNN(data[9241:9296,-8])
VOKS=kNN(data[9297:9352,-8])

#EXPORT DATA
write.csv(ABBA,"ABBA.csv")
write.csv(ADES,"ADES.csv")
write.csv(ADMG,"ADMG.csv")
write.csv(AKPI,"AKPI.csv")
write.csv(ALDO,"ALDO.csv")
:
write.csv(VOKS,"VOKS.csv")
write.csv(WIIM,"WIMM.csv")
write.csv(WOOD,"WOOD.csv")
write.csv(WTON,"WTON.csv")
write.csv(YPAS,"YPAS.csv")

#merge data
setwd("D:/Tugas Akhir")
filenames=list.files(full.names = TRUE)
All=lapply(filenames,function(i)
{
  read.csv(i,header = FALSE)
})
df=do.call(rbind.data.frame,All)
write.csv(df,"All.csv")
```

### Lampiran 3. Syntax Eksplorasi Data

```

library(dplyr)
library(ggplot2)
FD=filter(Data_TA,FD==1)
noFD=filter(Data_TA,FD==0)
Data_TA$FD=as.factor(Data_TA$FD)

#EBITA
avgEBITA.FD=mean(FD$EBITA)
avgEBITA.noFD=mean(noFD$EBITA)
ggplot(Data_TA, aes(EBITA, fill=FD)) +
  geom_density(alpha=.5) +
  geom_vline(data=Data_TA,
             mapping=aes(xintercept=avgEBITA.noFD), color="red") +
  geom_vline(data=Data_TA,
             mapping=aes(xintercept=avgEBITA.FD),           color="dark
green") +xlab("EBITA (t)")

::

#Inflasi
avgInflasi.FD=mean(FD$Inflasi)
avgInflasi.noFD=mean(noFD$Inflasi)
ggplot(Data_TA, aes(Inflasi, fill=FD))+ 
  geom_density(alpha=0.5) + xlab("Inflasi (t)")+
  geom_vline(data=Data_TA,
             mapping=aes(xintercept=avgInflasi.noFD), color="red") +
  geom_vline(data=Data_TA,
             mapping=aes(xintercept=avgInflasi.FD),           color="dark
green") +scale_x_continuous(limits = c(0,20))

```

### Lampiran 4. Syntax SVM Linier

```

rm(list=ls(all=TRUE))
library(e1071)
library(readxl)
datatrain <- read_excel("D:/Tugas Akhir/DataSplit.xlsx",
                        sheet = "train0")

```

```
datatrain_x=datatrain[,-1]
FD=as.factor(datatrain$FD)
dataSim=data.frame(FD,datatrain_x)
xtrain=model.matrix(FD~.,data=dataSim)[,-1]
datatest <- read_excel("D:/Tugas Akhir/DataSplit.xlsx",
                      sheet = "test0")
datatest_x=datatest[,-1]
FDt=as.factor(datatest$FD)
datTest=data.frame(FDt,datatest_x)
xtest=model.matrix(FDt~,data=datTest,kernel=linear)[,-1]
set.seed(123)
svmfit=svm(FD ~ ., data = dataSim)
print(svmfit)
predtr=predict(svmfit,dataSim[,-1])
predtest=predict(svmfit,datTest[,-1])

library(data.table)
library(caret)
library(pROC)
library(ROCR)

#akurasi data train
auc.train=auc(predtr,as.numeric(FD))
auc.train
cm.train=table(actual=FD,prediction=predtr)
cm.train
acc.train=(sum(diag(cm.train))/sum(cm.train))
acc.train
predtrain=prediction(as.numeric(predtr),FD)
perftrain=performance(predtrain,"tpr","fpr")
plot(perftrain,col="red",main="ROC Curve")

#akurasi data test
auc.test=auc(predtest,as.numeric(FDt))
auc.test
cm.test=table(actual=FDt,prediction=predtest)
cm.test
```

```

acc.test=(sum(diag(cm.test))/sum(cm.test))
acc.test
predtest=prediction(as.numeric(predtest),FDt)
perftest=performance(predtest,"tpr","fpr")
plot(perftest,col="red",main="ROC Curve")
#isotonic regression
library(isotone)
svmscore0=svmfit$decision.values
isoreg=gpava(z=NULL,y=svmscore0, ties = "primary")
plot(isoreg)

```

### Lampiran 5. Syntax Lasso SVM

```

rm(list=ls(all=TRUE))
set.seed(123)
library(readxl)
datatrain <- read_excel("D:/Tugas Akhir/DataSplit.xlsx",
                        sheet = "train0")
datatrain_x=datatrain[,-1]
FD=as.factor(datatrain$FD)
dataSim=data.frame(FD,datatrain_x)
xtrain=model.matrix(FD~.,data=dataSim)[,-1]
datatest <- read_excel("D:/Tugas Akhir/DataSplit.xlsx",
                        sheet = "test0")
datatest_x=datatest[,-1]
FDt=as.factor(datatest$FD)
datTest=data.frame(FDt,datatest_x)
xtest=model.matrix(FDt~.,data=datTest)[,-1]
library(sparseSVM)
library(data.table)
library(caret)
library(pROC)
library(ROCR)
Lasso=sparseSVM(xtrain, FD, gamma = 0.1, nlambda=10,
                 lambda.min = ifelse(nrow(xtrain)>ncol(xtrain), 0.01, 0.05),
                 preprocess = c("standardize", "rescale", "none"),
                 screen = "none", max.iter =500, eps = 1e-5,
                 dfmax=ncol(xtrain)+1, penalty.factor=rep(1, ncol(xtrain)),
                 message = FALSE)

```

```

pred.train=as.numeric(predict(Lasso,xtrain,lambda=
0.0051,type="class"))
(misclasstrain <- table(predict =pred.train , actual = dataSim$FD))
auc.train=auc(FD,pred.train)
auc.train
acc.train=(sum(diag(misclasstrain))/sum(misclasstrain))
acc.train
predtrain=prediction(pred.train,FD)
perftrain=performance(predtrain,"tpr","fpr")
plot(perftrain,col="red",main="ROC Curve")

pred.test=as.numeric(predict(Lasso,xtest,lambda=
0.0051,type="class"))
(misclasstest <- table(predict =pred.test , actual = datTest$FD))
auc.test=auc(FDt,pred.test)
auc.test
acc.test=(sum(diag(misclasstest))/sum(misclasstest))
acc.test
predtest=prediction(pred.test,FDt)
perftest=performance(predtest,"tpr","fpr")
plot(perftest,col="red",main="ROC Curve")

isoreg=gpava(z=NULL,y=ScoreLasso0, ties = "primary")
plot(isoreg)

```

## Lampiran 6. Syntax Elastic-Net SVM

```

rm(list=ls(all=TRUE))
set.seed(123)
library(penalizedSVM)
library(e1071)
library(readxl)
datatrain <- read_excel("D:/Tugas Akhir/DataSplit.xlsx",
sheet = "train0")
datatrain_x=datatrain[,-1]
FD=as.numeric(datatrain$FD)
dataSim=data.frame(FD,datatrain_x)
xtrain=model.matrix(FD~.,data=dataSim)[,-1]
datatest <- read_excel("D:/Tugas Akhir/DataSplit.xlsx",
sheet = "test0")
datatest_x=datatest[,-1]

```

```

FDt=as.numeric(dataTest$FD)
dataTest=data.frame(FDt,dataTest_x)
lambda1.1norm<-seq(0.1,0.6,0.1)
bounds=seq(0.0001,0.0009,0.0002)
fit.elasticnet=svmfs(x=xtrain,y=dataSim$FD,
                      fs.method="DrHSVM",
                      cross.outer= 0, grid.search = "discrete",
                      lambda1.set=lambda1.1norm,bounds = bounds,
                      parms.coding = "none", show="none",
                      maxIter = 700, inner.val.method = "cv", cross.inner= 5,
                      verbose=FALSE)
library(data.table)
library(caret)
library(pROC)
library(ROCR)
pred.train=predict(fit.elasticnet,newdata=xtrain, newdata.labels=FD)
pred.train$tab
fit.train=pred.train[["pred.class"]]
auc.train=auc(fit.train,FD)
auc.train
acc.train=(sum(diag(pred.train$tab))/sum(pred.train$tab))
acc.train
predtrain=prediction(as.numeric(fit.train),FD)
perftrain=performance(predtrain,"tpr","fpr")
plot(perftrain,col="red",main="ROC Curve")
pred.test=predict(fit.elasticnet,    newdata=as.matrix(dataTest[,-1]),
newdata.labels=FDt)
pred.test$tab
fit.test=pred.test[["pred.class"]]
auc.test=auc(fit.test,FDt)
auc.test
acc.test=(sum(diag(pred.test$tab))/sum(pred.test$tab))
acc.test
predtest=prediction(as.numeric(fit.test),FDt)
perftest=performance(predtest,"tpr","fpr")
plot(perftest,col="red",main="ROC Curve")
library(isotone)
scoreelasticnet=pred.train[["fitted"]]
isoreg=gpava(z=NULL,y=scoreelasticnet, ties = "primary")
plot(isoreg)

```

**Lampiran 7. Syntax GEVR**

```
rm(list=ls())

library(data.table)
library(caret)
#library(pROC)
library(ROCR)
library(bgeva)

EBITA = datatrain$EBITA
STA = datatrain$STA
ITR = datatrain$ITR
DSI = datatrain$DSI
ROE = datatrain$ROE
NPM = datatrain$NPM
OPM = datatrain$OPM
DER = datatrain$DER
DAR = datatrain$DAR
CR = datatrain$CR
QR = datatrain$QR
ETD = datatrain$ETD
WCA = datatrain$WCA
WCS = datatrain$WCS
rgdpg = datatrain$rgdpg
bi7dr = datatrain$bi7dr
USD_IDR = datatrain`USD/IDR`
Inflasi = datatrain$Inflasi

FD = factor(datatrain$FD)
dataSim = data.frame(FD, EBITA, STA, ITR, DSI, ROE, NPM,
                     OPM, DER, DAR, CR, QR, ETD, WCA, WCS, rgdpg,
                     bi7dr, USD_IDR, Inflasi)
EBITAt = datatest$EBITA
STAt = datatest$STA
ITRt = datatest$ITR
DSIt = datatest$DSI
ROEt = datatest$ROE
```

```

NPMt = datatest$NPM
OPMt = datatest$OPM
DERt = datatest$DER
DARt = datatest$DAR
CRt = datatest$CR
QRt = datatest$QR
ETDt = datatest$ETD
WCAt = datatest$WCA
WCSt = datatest$WCS
rgdpgt = datatest$rgdpg
bi7drrt = datatest$bi7dr
USD_IDRt = datatest$`USD/IDR`
Inflasit = datatest$Inflasi

FDt = factor(datatest$FD)
datTest = data.frame(FDt, EBITAt, STAt, ITRt, DSIt, ROEt, NPMt,
                     OPMt, DERt, DARt, CRt, QRt, ETDt, WCAt, WCSt,
                     rgdpgt, bi7drrt, USD_IDRt, Inflasit)

out = bgeva(FD ~ EBITA + STA + ITR + DSIt + ROE + NPM +
             OPM + DER + DAR + CR + QR + ETD + WCA + WCS +
             rgdpg + bi7dr + USD_IDR + Inflasi, data = dataSim, pr.tol
             = 1e-5, rmax = 500, tau=0.1, control = list(tol=1e-5), Hes =
             FALSE)

parameter = out$fit$argument

peluang = function(parameter, covariate, tau){
  n = nrow(covariate)
  tau = tau
  beta = as.matrix(parameter)
  xgab = as.matrix(cbind(rep(1,n), covariate))
  peluang1 = array(0,n)
  kelas = array(0,n)

  mu = (xgab %*% beta)
  for (i in 1:n){
    peluang1[i] = exp(-1*((1+(tau*mu[i]))^(-1/tau)))
    if(peluang1[i] > 0.5)
    {
  
```

```

kelas[i] = 1
} else {
  kelas[i] = 0
}
}
list(kelas=kelas, peluang1=peluang1)
}

hasil_train1 = peluang(parameter, subset(dataSim, select = -FD), 0.1)
pred.train1 = hasil_train1$kelas
hasil_test1 = peluang(parameter,subset(datTest, select=-FDt),0.1)
pred.test1 = hasil_test1$kelas
#str(pred.train1)
pred.train = factor(pred.train1)
pred.test = factor(pred.test1)
dataSim2=cbind(datatrain[,c("periode","kode")],pred.train, dataSim)
View(dataSim2)
dataTest2=cbind(datatest[,c("periode","kode")], pred.test, datTest)
View(dataTest2)
# Confusion matrix
FD.true = dataSim2$FD
FD.pred = dataSim2$pred.train
table(FD.true, FD.pred)
FDt.true=dataTest2$FD
FDt.pred=dataTest2$pred.test
table(FDt.true, FDt.pred)
# ROC Curve
pred <- prediction(as.numeric(hasil_train1$peluang1), FD.true)
perf <- performance(pred, measure = "tpr", x.measure = "fpr")
plot(perf, col=rainbow(10), main = "ROC curve")
predt <- prediction(as.numeric(hasil_test1$peluang1),FDt.true)
perft <- performance(predt, measure = "tpr", x.measure = "fpr")
plot(perft,col=rainbow(10), main = "ROC Curve")
# AUC value
auc_ROCR <- performance(pred, measure = "auc")
auc_ROCR <- auc_ROCR@y.values[[1]]
cat(paste("AUC = ", auc_ROCR), "\n")
auct_ROCR <- performance(predt, measure = "auc")
auct_ROCR <- auct_ROCR@y.values[[1]]
cat(paste("AUC = ", auct_ROCR), "\n")

```

```
# Accuracy
acc_ROCR <- performance(pred, measure = "acc")
acc_ROCR <- max(acc_ROCR@y.values[[1]])
cat(paste("ACC = ", acc_ROCR), "\n")

acct_ROCR <- performance(predt, measure = "acc")
acct_ROCR <- max(acct_ROCR@y.values[[1]])
cat(paste("ACC = ", acct_ROCR), "\n")

summary(out,s.meth="svd",sig.lev = 0.05)
```

### Lampiran 8. Syntax Regresi Logistik

```
rm(list=ls(all=TRUE))
library(readxl)
datatrain <- read_excel("D:/Tugas Akhir/DataSplit.xlsx",
                        sheet = "train1")
datatrain_x=datatrain[,-1]
FD=as.factor(datatrain$FD)
dataSim=data.frame(FD,datatrain_x)
datatest <- read_excel("D:/Tugas Akhir/DataSplit.xlsx",
                       sheet = "test1")
datatest_x=datatest[,-1]
FDt=as.factor(datatest$FD)
datTest=data.frame(FDt,datatest_x)
library(data.table)
library(caret)
library(pROC)
library(ROCR)

model=glm(FD~.,data=dataSim,family=binomial(link='logit'),contr
ol = list(epsilon=1e-6))
OR=exp(coef(model))

#Modelstep=step(model)

pred.train=round(predict(model,subset(dataSim,select=-FD),
type="response"))
pred.test=round(predict(model,subset(datTest,select=-FDt),
type="response"))
```

```
cm.train=table(actual=FD,prediction=pred.train)
cm.train
acc.train=(sum(diag(cm.train))/sum(cm.train))
acc.train
auc.train=auc(FD,pred.train)
auc.train
cm.test=table(actual=FDt,prediction=pred.test)
cm.test
acc.test=(sum(diag(cm.test))/sum(cm.test))
acc.test
auc.test=auc(FDt,pred.test)
auc.test

#ROC
predtrain=prediction(pred.train,FD)
perftrain=performance(predtrain,"tpr","fpr")
plot(perftrain,col="red",main="ROC Curve")
predtest=prediction(pred.test,FDt)
perftest=performance(predtest,"tpr","fpr")
plot(perftest,col="red",main="ROC Curve")

summary(model)
ModelStep=step(model)
summary(ModelStep)
ORstep=exp(coef(ModelStep))
pred.train.step=round(predict(ModelStep,subset(dataSim,select = -FD),type="response"))
pred.test.step=round(predict(ModelStep,subset(datTest,select = -FDt),type="response"))

cm.train.step=table(actual=FD,prediction=pred.train.step)
cm.train.step
acc.train.step=(sum(diag(cm.train.step))/sum(cm.train.step))
acc.train.step
auc.train.step=auc(FD,pred.train.step)
auc.train.step
cm.test.step=table(actual=FDt,prediction=pred.test.step)
cm.test.step
acc.test.step=(sum(diag(cm.test.step))/sum(cm.test.step))
acc.test.step
```

```

auc.test.step=auc(FDt,pred.test.step)
auc.test.step

#ROC
predtrainstep=prediction(pred.train.step,FD)
perftrainstep=performance(predtrainstep,"tpr","fpr")
plot(perftrainstep,col="red",main="ROC Curve")
predteststep=prediction(pred.test.step,FDt)
perfteststep=performance(predteststep,"tpr","fpr")
plot(perfteststep,col="red",main="ROC Curve")

```

### Lampiran 9. Syntax Lasso Regresi Logistik

```

rm(list=ls(all=TRUE))
set.seed(123)
library(readxl)
datatrain <- read_excel("D:/Tugas Akhir/DataSplit.xlsx",
                        sheet = "train2")
datatrain_x=datatrain[,-1]
FD=as.factor(datatrain$FD)
dataSim=data.frame(FD,datatrain_x)
xtrain=model.matrix(FD~.,data=dataSim)[,-1]
datatest <- read_excel("D:/Tugas Akhir/DataSplit.xlsx",
                        sheet = "test2")
datatest_x=datatest[,-1]
FDt=as.factor(datatest$FD)
datTest=data.frame(FDt,datatest_x)
xtest=model.matrix(FDt~.,data=datTest)[,-1]

library(glmnet)
library(ROCR)

cv.lasso=cv.glmnet(xtrain,FD,alpha=1, family= "binomial",
type.measure ="auc")
plot(cv.lasso)
cv.lasso[["lambda.min"]]
#regression coefficients
coef(cv.lasso,s=cv.lasso$lambda.min)
lasso_pred.train <- round(predict(cv.lasso,newx =
xtrain,s=cv.lasso$lambda.min, type="response"))

```

```

cm.train=table(actual=FD,predict=lasso_pred.train>0.5)
lasso_pred.test <- round(predict(cv.lasso,newx
=xtest,s=cv.lasso$lambda.min, type="response"))
cm.test=table(actual=FDt,predict=lasso_pred.test>0.5)
cm.train
cm.test

mutrain=mean(lasso_pred.train == FD)
pred_gbp1 <- prediction(lasso_pred.train, FD)
auctrain <- performance(pred_gbp1, measure = "auc")
auctrain@y.values[1]
mutest=mean(lasso_pred.test == FDt)
pred_gbp2 <- prediction(lasso_pred.test, FDt)
auctest <- performance(pred_gbp2, measure = "auc")
auctest@y.values[1]

acc.train=(sum(diag(cm.train))/sum(cm.train))
acc.train
acc.test=(sum(diag(cm.test))/sum(cm.test))
acc.test

predtrain=prediction(lasso_pred.train,FD)
perftrain=performance(predtrain,"tpr","fpr")
plot(perftrain,col="red",main="ROC Curve")
predtest=prediction(lasso_pred.test,FDt)
perftest=performance(predtest,"tpr","fpr")
plot(perftest,col="red",main="ROC Curve")

```

#### Lampiran 10. Syntax Elastic-Net Regresi Logistik

```

rm(list=ls(all=TRUE))
set.seed(123)

library(readxl)
datatrain <- read_excel("D:/Tugas Akhir/DataSplit.xlsx",
                        sheet = "train2")
datatrain_x=datatrain[,-1]
FD=as.factor(datatrain$FD)
dataSim=data.frame(FD,datatrain_x)
xtrain=model.matrix(FD~.,data=dataSim)[,-1]

```

```
datatest <- read_excel("D:/Tugas Akhir/DataSplit.xlsx",
                      sheet = "test2")
datatest_x=datatest[,-1]
FDt=as.factor(datatest$FD)
datTest=data.frame(FDt,datatest_x)
xtest=model.matrix(FDt~.,data=datTest)[,-1]

cv.elastic_net_0.1 <- cv.glmnet(xtrain,FD,alpha=0.1, family=
"binomial",type.measure ="auc")
predtr.cv.elastic_net_0.1 <- round(predict(cv.elastic_net_0.1,newx =
xtrain,s=cv.elastic_net_0.1$lambda.min, type="response"))
predtr_cv.elastic_net_0.1 <-
prediction(as.numeric(predtr.cv.elastic_net_0.1), FD)
auc.cv.elastic_net_0.1 <- performance(predtr_cv.elastic_net_0.1,
measure = "auc")
auc.cv.elastic_net_0.1@y.values[1]

cv.elastic_net_0.2 <- cv.glmnet(xtrain,FD,alpha=0.2, family=
"binomial",type.measure ="auc")
predtr.cv.elastic_net_0.2 <- round(predict(cv.elastic_net_0.2,newx =
xtrain,s=cv.elastic_net_0.2$lambda.min, type="response"))
predtr_cv.elastic_net_0.2 <-
prediction(as.numeric(predtr.cv.elastic_net_0.2), FD)
auc.cv.elastic_net_0.2 <- performance(predtr_cv.elastic_net_0.2,
measure = "auc")
auc.cv.elastic_net_0.2@y.values[1]

cv.elastic_net_0.3 <- cv.glmnet(xtrain,FD,alpha=0.3, family=
"binomial",type.measure ="auc")
predtr.cv.elastic_net_0.3 <- round(predict(cv.elastic_net_0.3,newx =
xtrain,s=cv.elastic_net_0.3$lambda.min, type="response"))
predtr_cv.elastic_net_0.3 <-
prediction(as.numeric(predtr.cv.elastic_net_0.3), FD)
auc.cv.elastic_net_0.3 <- performance(predtr_cv.elastic_net_0.3,
measure = "auc")
auc.cv.elastic_net_0.3@y.values[1]

cv.elastic_net_0.4 <- cv.glmnet(xtrain,FD,alpha=0.4, family=
"binomial",type.measure ="auc")
```

```

predtr.cv.elastic_net_0.4 <- round(predict(cv.elastic_net_0.4,newx
= xtrain,s=cv.elastic_net_0.4$lambda.min, type="response"))
predtr_cv.elastic_net_0.4 <-
prediction(as.numeric(predtr.cv.elastic_net_0.4), FD)
auc.cv.elastic_net_0.4 <- performance(predtr_cv.elastic_net_0.4,
measure = "auc")
auc.cv.elastic_net_0.4@y.values[1]

cv.elastic_net_0.5 <- cv.glmnet(xtrain,FD,alpha=0.5, family=
"binomial",type.measure ="auc")
predtr.cv.elastic_net_0.5 <- round(predict(cv.elastic_net_0.5,newx
= xtrain,s=cv.elastic_net_0.5$lambda.min, type="response"))
predtr_cv.elastic_net_0.5 <-
prediction(as.numeric(predtr.cv.elastic_net_0.5), FD)
auc.cv.elastic_net_0.5 <- performance(predtr_cv.elastic_net_0.5,
measure = "auc")
auc.cv.elastic_net_0.5@y.values[1]

cv.elastic_net_0.6 <- cv.glmnet(xtrain,FD,alpha=0.6, family=
"binomial",type.measure ="auc")
predtr.cv.elastic_net_0.6 <- round(predict(cv.elastic_net_0.6,newx
= xtrain,s=cv.elastic_net_0.6$lambda.min, type="response"))
predtr_cv.elastic_net_0.6 <-
prediction(as.numeric(predtr.cv.elastic_net_0.6), FD)
auc.cv.elastic_net_0.6 <- performance(predtr_cv.elastic_net_0.6,
measure = "auc")
auc.cv.elastic_net_0.6@y.values[1]

cv.elastic_net_0.7 <- cv.glmnet(xtrain,FD,alpha=0.7, family=
"binomial",type.measure ="auc")
predtr.cv.elastic_net_0.7 <- round(predict(cv.elastic_net_0.7,newx
= xtrain,s=cv.elastic_net_0.7$lambda.min, type="response"))
predtr_cv.elastic_net_0.7 <-
prediction(as.numeric(predtr.cv.elastic_net_0.7), FD)
auc.cv.elastic_net_0.7 <- performance(predtr_cv.elastic_net_0.7,
measure = "auc")
auc.cv.elastic_net_0.7@y.values[1]

cv.elastic_net_0.8 <- cv.glmnet(xtrain,FD,alpha=0.8, family=
"binomial",type.measure ="auc")

```

```

predtr.cv.elastic_net_0.8 <- round(predict(cv.elastic_net_0.8,newx
= xtrain,s=cv.elastic_net_0.8$lambda.min, type="response"))
predtr_cv.elastic_net_0.8 <-
prediction(as.numeric(predtr.cv.elastic_net_0.8), FD)
auc.cv.elastic_net_0.8 <- performance(predtr_cv.elastic_net_0.8,
measure = "auc")
auc.cv.elastic_net_0.8@y.values[1]

cv.elastic_net_0.9 <- cv.glmnet(xtrain,FD,alpha=0.9, family=
"binomial",type.measure ="auc")
predtr.cv.elastic_net_0.9 <- round(predict(cv.elastic_net_0.9,newx
= xtrain,s=cv.elastic_net_0.9$lambda.min, type="response"))
predtr_cv.elastic_net_0.9 <-
prediction(as.numeric(predtr.cv.elastic_net_0.9), FD)
auc.cv.elastic_net_0.9 <- performance(predtr_cv.elastic_net_0.9,
measure = "auc")
auc.cv.elastic_net_0.9@y.values[1]

cv.elastic_net_0.1[["lambda.min"]]
#regression coefficients
coef(cv.elastic_net_0.1,s=cv.elastic_net_0.1$lambda.min)
predtest.cv.elastic_net_0.1 <- round(predict(cv.elastic_net_0.1,newx
= xtest,s=cv.elastic_net_0.1$lambda.min, type="response"))
predtest_cv.elastic_net_0.1 <-
prediction(as.numeric(predtest.cv.elastic_net_0.1), FDt)
auctest.cv.elastic_net_0.1 <-
performance(predtest_cv.elastic_net_0.1, measure = "auc")
auctest.cv.elastic_net_0.1@y.values[1]

cm.train=table(actual=FD,predict=predtr.cv.elastic_net_0.1>0.5)
cm.train
cm.test=table(actual=FDt,predict=predtest.cv.elastic_net_0.1>0.5)
cm.test

acc.train=(sum(diag(cm.train))/sum(cm.train))
acc.train
acc.test=(sum(diag(cm.test))/sum(cm.test))

```

```
acc.test
```

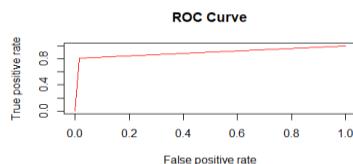
```
predtrain=prediction(predtr.cv.elastic_net_0.1,FD)
perftrain=performance(predtrain,"tpr","fpr")
plot(perftrain,col="red",main="ROC Curve")
predtest=prediction(predtest.cv.elastic_net_0.1,FDt)
perftest=performance(predtest,"tpr","fpr")
plot(perftest,col="red",main="ROC Curve")
```

## Lampiran 11. ROC SVM

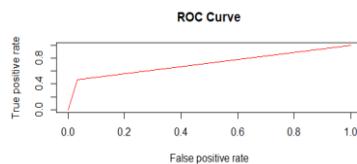
### 1. SVM Linier

Data train

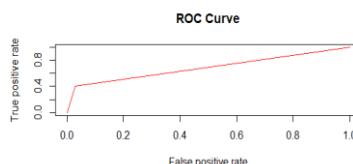
*Size 0*



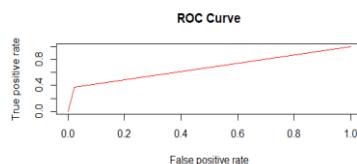
*Size 1*



*Size 2*

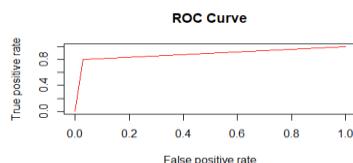


*Size 3*

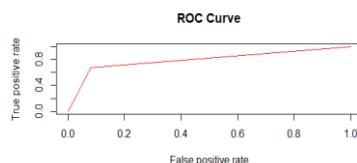


Data test

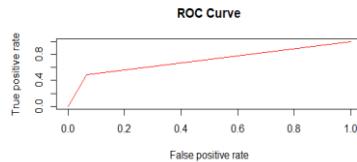
*Size 0*



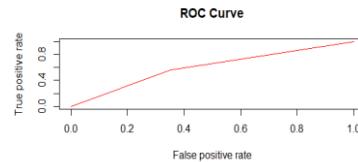
*Size 1*



*Size 2*



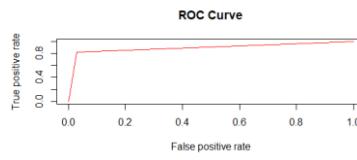
*Size 3*



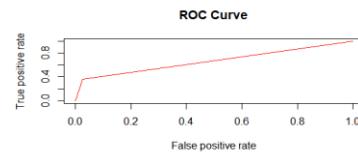
## 2. Lasso SVM

Data train

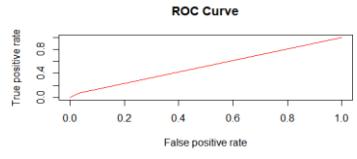
*Size 0*



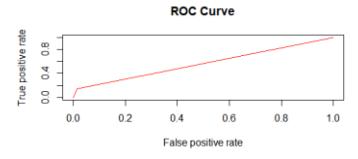
*Size 1*



*Size 2*

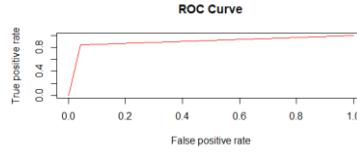


*Size 3*

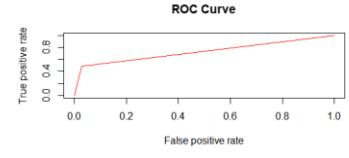


Data test

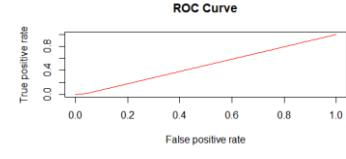
*Size 0*



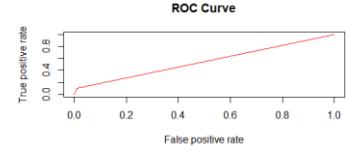
*Size 1*



*Size 2*



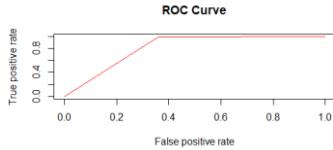
*Size 3*



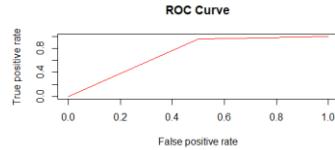
### 3. Elastic-Net SVM

Data *train*

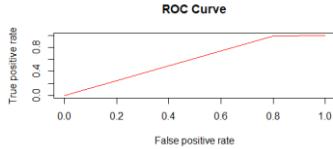
*Size 0*



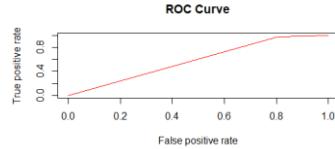
*Size 1*



*Size 2*

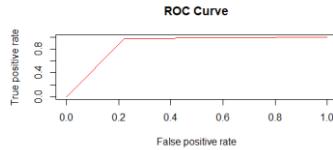


*Size 3*

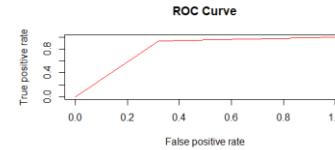


Data *test*

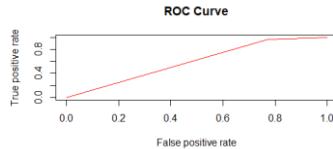
*Size 0*



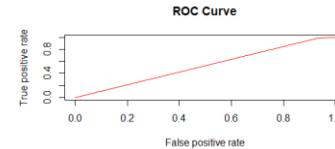
*Size 1*



*Size 2*



*Size 3*

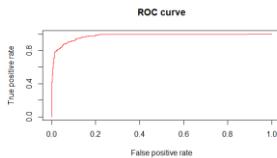


## Lampiran 12. ROC GEVR

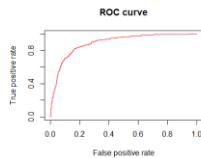
### 1. GEVR Serentak

Data *train*

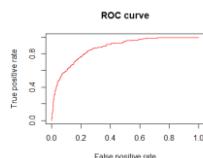
*Size 0*



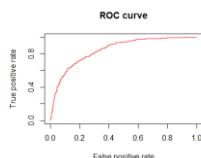
*Size 1*



*Size 2*

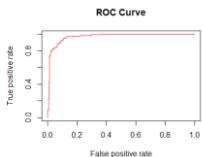


*Size 3*

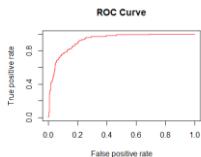


Data *test*

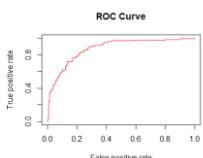
*Size 0*



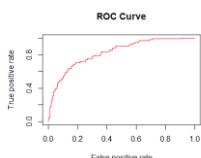
*Size 1*



*Size 2*



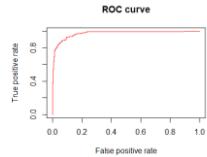
*Size 3*



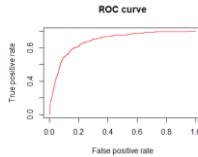
## 2. Backward GEVR

Data *train*

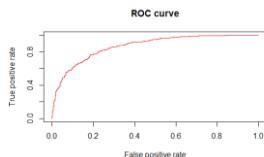
*Size 0*



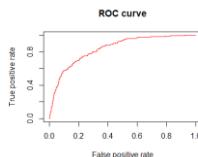
*Size 1*



*Size 2*

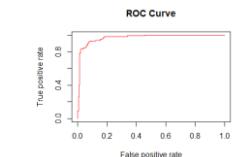


*Size 3*

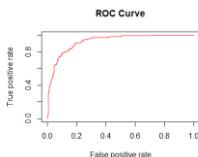


Data *test*

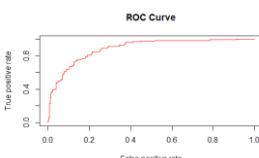
*Size 0*



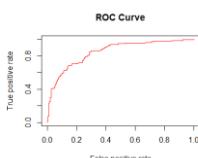
*Size 1*



*Size 2*



*Size 3*

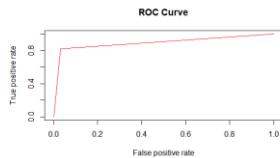


### Lampiran 13. ROC LOGIT

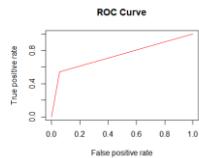
#### 1. LOGIT Serentak

Data *train*

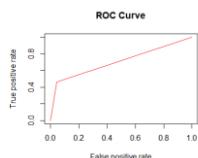
*Size 0*



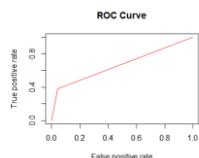
*Size 1*



*Size 2*

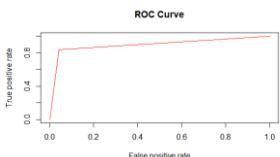


*Size 3*

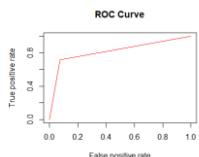


Data *test*

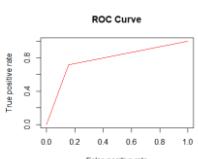
*Size 0*



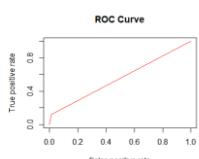
*Size 1*



*Size 2*



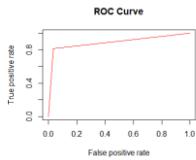
*Size 3*



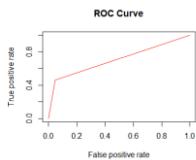
## 2. Stepwise LOGIT

Data train

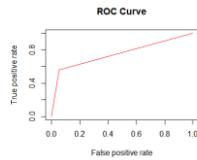
Size 0



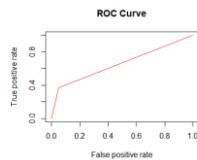
Size 2



Size 1

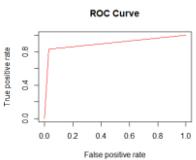


Size 3

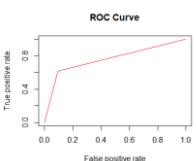


Data test

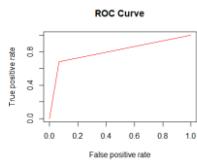
Size 0



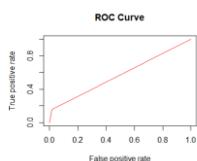
Size 2



Size 1



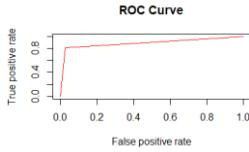
Size 3



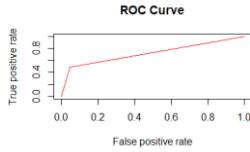
### 3. Lasso LOGIT

#### Data train

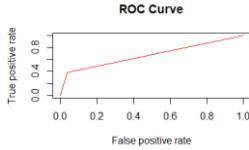
*Size 0*



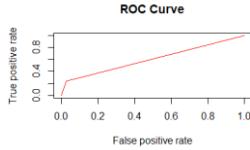
*Size 1*



*Size 2*

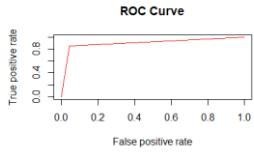


*Size 3*

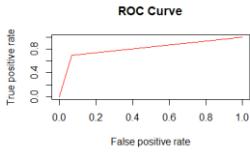


#### Data test

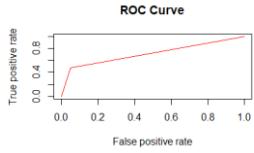
*Size 0*



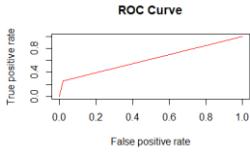
*Size 1*



*Size 2*



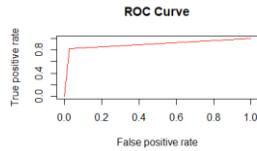
*Size 3*



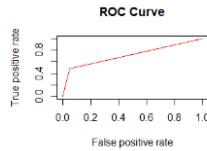
#### 4. Elastic-Net LOGIT

Data train

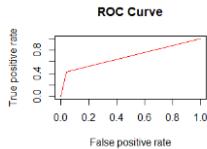
Size 0



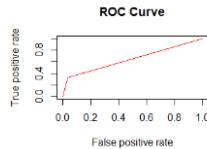
Size 1



Size 2

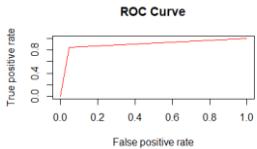


Size 3

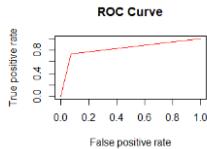


Data test

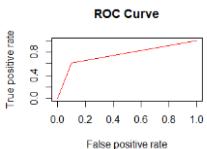
Size 0



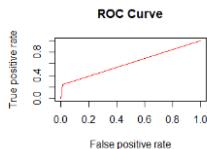
Size 1



Size 2



Size 3



**Lampiran 14.** Koefisien *Lasso SVM*

Prediktor	<i>Size 0</i>	<i>Size 1</i>	<i>Size 2</i>	<i>Size 3</i>
<i>Intercept</i>	0,2660949	0,2661	1,2270	-0,9021545
EBITA	10,07979	10,0798	3,8182	-0,54969905
STA	0,1089795	0,10898	0,0341	-0,1209854
ITR	0	0	0	-0,0002011
DSI	-0,0010754	-0,0011	0	0,000207097
ROE	0	0	0	-0,00085046
NPM	0,04753503	0,0475	0	-0,6713109
OPM	0,6088416	0,6088	-1,1204	-1,347988
DER	0	0	0	0,000173856
DAR	-0,3143695	-0,3144	-0,1219	0,4157625
CR	-0,002758	-0,0028	-0,0406	0,01407667
QR	-0,01548	-0,0155	-0,0876	0
ETD	0,03906778	0,0391	-0,9948	-0,07328798
WCA	0,5188433	0,5188	0,5284	-0,2556101
WCS	0	0	0	-0,2696003
RGDPG	0,06116144	0,0612	0	0,02819316
BI7DRR	0	0	0	-0,000522
USD/IDR	0,00002	-0,00002	0	-0,000006
Inflasi	0	0	0	0,0007505

**Lampiran 15.** Koefisien *Elastic-Net SVM*

Prediktor	<i>Size 0</i>	<i>Size 1</i>	<i>Size 2</i>	<i>Size 3</i>
<i>Intercept</i>	0,3776	-0,4991	0,7757	-0,3903
EBITA	1,07104	7,7046	0	0,0008
STA	0,2247	0,1035	0,1764	0,1437
ITR	-0,0039	-0,00017	-0,00028	-0,0002
DSI	-0,0007	-0,000081	-0,000071	-0,0004
ROE	0,02499	-0,09104	0,0489	0,1052
NPM	0	0,2175	0	0
OPM	4,3156	-0,27401	0,3887	0,7902

Prediktor	Size 0	Size 1	Size 2	Size 3
DER	-0,0015	-0,0112	-0,0029	-0,0061
DAR	-0,0236	-0,4977	-0,2919	-0,2570
CR	0	-0,0074	0	0
QR	-0,0026	-0,0033	-0,0264	-0,0064
ETD	1,9741	0,0515	0,2554	0,0728
WCA	0,0693	0,0268	0,2677	0,2998
WCS	0,0941	-0,0084	0,0415	0,0129
RGDPG	0	0,1077	0	0
BI7DRR	-0,0056	0,0399	0	0,0367
USD/IDR	-0,00003	$-5,81 \times 10^{-6}$	-0,00004	0,00007
Inflasi	-0,00041	-0,0133	0,0086	-0,0230

**Lampiran 16.** Nilai Log-Likelihood Parameter  $\tau$  untuk GEVR

$\tau$	Log-Likelihood	$\tau$	Log-Likelihood
0,01	0,01903826	-0,01	$5,346834 \times 10^{-12}$
0,02	$4,156009 \times 10^{-8}$	-0,02	0
0,03	$5,113489 \times 10^{-7}$	-0,03	0
0,04	$1,981411 \times 10^{-8}$	-0,04	0
0,05	$1,516542 \times 10^{-8}$	-0,05	0
0,06	12,24807	-0,06	0
0,07	$5,005455 \times 10^{-9}$	-0,07	0
0,08	$2,875511 \times 10^{-8}$	-0,08	0
0,09	0	-0,09	0
0,1	$9,724724 \times 10^{-9}$	-0,1	0
0	0		

**Lampiran 17.** Estimasi Paramter Model dengan GEVR Serentak  
Size 0

Prediktor	Estimate	Std. Error	z value	Pr(> z )
Intercept	-2,053	1,767	-1,161	0,245419
EBITA	-17,33	2,558	-6,776	$1,23 \times 10^{-11}$

Prediktor	<i>Estimate</i>	<i>Std. Error</i>	<i>z value</i>	<i>Pr(&gt; z )</i>
STA	-0,3299	0,09756	-3,382	0,00072
ITR	0,008014	0,001729	4,635	3,57 x 10 <sup>-6</sup>
DSI	0,001402	0,001038	1,351	0,176632
ROE	-0,08246	0,1437	-0,574	0,566188
NPM	-4,444	1,261	-3,525	0,000423
OPM	-9,663	2,223	-4,347	1,38 x 10 <sup>-5</sup>
DER	0,004357	0,007726	0,564	0,572785
DAR	0,9226	0,2172	4,248	2,15 x 10 <sup>-5</sup>
CR	0,06317	0,05662	1,116	0,264531
QR	-0,1172	0,06311	-1,858	0,063223
ETD	-1,954	0,9372	-2,085	0,037062
WCA	-0,4504	0,4155	-1,084	0,278406
WCS	0,1207	0,2877	0,420	0,674732
RGDPG	0,2078	0,1967	1,056	0,290793
BI7DRR	0,1166	0,06334	1,841	0,065581
USD/IDR	9,453 x 10 <sup>-5</sup>	5,466 x 10 <sup>-5</sup>	1,729	0,083740
Inflasi	-0,0094	0,02832	-0,332	0,739979

Size 1

Prediktor	<i>Estimate</i>	<i>Std. Error</i>	<i>z value</i>	<i>Pr(&gt; z )</i>
Intercept	1,765	1,906	0,926	0,354468
EBITA	-23,52	1,847	-12,733	< 2 x 10 <sup>-16</sup>
STA	-0,31	0,09371	-3,308	0,000939
ITR	0,0005081	0,0003169	1,603	0,108921
DSI	0,0002761	0,0007316	0,377	0,705867
ROE	0,003044	0,1803	1,688	0,091336
NPM	-0,6622	0,4293	-1,543	0,122928
OPM	0,9032	0,3742	2,413	0,015803
DER	0,03378	0,009454	3,573	0,000353
DAR	1,419	0,2227	6,372	1,86 x 10 <sup>-10</sup>
CR	0,009809	0,05274	0,186	0,852446
QR	0,01919	0,06963	0,276	0,782800

Prediktor	<i>Estimate</i>	<i>Std. Error</i>	<i>z value</i>	<i>Pr(&gt; z )</i>
ETD	-0,1958	0,2464	-0,795	0,426882
WCA	-0,1315	0,3521	-0,374	0,708766
WCS	0,02148	0,1159	0,185	0,852925
RGDPG	-0,3203	0,1909	-1,678	0,093443
BI7DRR	-0,1296	0,07798	-1,662	0,096455
USD/IDR	6,149 x 10 <sup>-5</sup>	7,239 x 10 <sup>-5</sup>	0,849	0,395679
Inflasi	0,04523	0,03766	1,201	0,229715

*Size 2*

Prediktor	<i>Estimate</i>	<i>Std. Error</i>	<i>z value</i>	<i>Pr(&gt; z )</i>
<i>Intercept</i>	-3,031	1,314	-2,307	0,021081
EBITA	-4,757	1,26	-3,776	0,000159
STA	-0,1645	0,07027	-2,341	0,019248
ITR	0,0006135	0,0003202	1,916	0,055364
DSI	0,001066	0,0005479	1,945	0,0518
ROE	0,1636	0,1741	0,940	0,347457
NPM	-0,7203	0,5431	-1,326	0,184771
OPM	-2,861	1,085	-2,636	0,008388
DER	0,003704	0,008528	0,434	0,664042
DAR	0,47	0,1765	2,662	0,007759
CR	-0,047	0,02438	-1,928	0,053836
QR	0,1512	0,04242	3,564	0,000366
ETD	-0,2879	0,2287	-1,259	0,207969
WCA	-0,5813	0,2903	-2,002	0,045265
WCS	-0,374	0,1436	-2,605	0,009195
RGDPG	0,245	0,1306	1,875	0,060739
BI7DRR	0,03407	0,05102	0,668	0,504258
USD/IDR	0,0001701	5,801 x 10 <sup>-5</sup>	2,932	0,003364
Inflasi	-0,03212	0,02539	-1,265	0,205817

*Size 3*

Prediktor	Estimate	Std. Error	z value	Pr(> z )
Intercept	3,8560811	1,9829122	1,945	0,05182
EBITA	-3,8638164	1,1807778	-3,272	0,00107
STA	-0,1002231	0,0658787	-2,521	0,12818
ITR	0,0001640	0,0003997	0,410	0,68157
DSI	0,0007966	0,0005583	1,427	0,15366
ROE	0,0331084	0,1878186	0,176	0,86008
NPM	-0,1166669	0,340998	-0,342	0,73225
OPM	-2,9130211	1,0553497	-2,760	0,00578
DER	0,0088839	0,0088782	1,001	0,31700
DAR	0,4061793	0,1753080	2,317	0,02051
CR	-0,0063613	0,0331820	-0,192	0,84797
QR	0,0351881	0,0461215	0,763	0,44550
ETD	-0,0388334	0,1540257	-0,252	0,80095
WCA	-0,6535245	0,2694610	-2,425	0,01530
WCS	-0,0619362	0,0860048	-0,720	0,47143
RGDPG	-0,1648952	0,1508089	-1,093	0,27422
BI7DRR	-0,1393615	0,0525522	-2,652	0,00800
USD/IDR	-0,0002542	0,0001186	-2,143	0,03211
Inflasi	0,0689465	0,0264664	2,605	0,00919

**Lampiran 18.** Estimasi Paramter Model Akhir dengan *Backward GEVR*

*Size 0*

Prediktor	Estimate	Std. Error	z value	Pr(> z )
Intercept	0,505383	0,264769	1,909	0,056292
EBITA	-18,238119	2,447738	-7,451	9,26 x 10 <sup>-14</sup>
STA	-0,366221	0,089670	-4,084	4,42 x 10 <sup>-5</sup>
ITR	0,007739	0,001424	5,436	5,45 x 10 <sup>-8</sup>
NPM	-4,531492	1,196432	-3,788	0,000152
OPM	-8,657801	1,958409	-4,421	9,83 x 10 <sup>-6</sup>

Prediktor	<i>Estimate</i>	<i>Std. Error</i>	<i>z value</i>	<i>Pr(&gt; z )</i>
DAR	1,040819	0,188048	5,535	3,11 x 10 <sup>-8</sup>
QR	-0,053623	0,013035	-4,114	3,89 x 10 <sup>-5</sup>
ETD	-1,986125	0,960770	-2,067	0,038713
BI7DRR	0,066043	0,029925	2,207	0,027320

*Size 1*

Prediktor	<i>Estimate</i>	<i>Std. Error</i>	<i>z value</i>	<i>Pr(&gt; z )</i>
<i>Intercept</i>	1,504072	0,79798	1,885	0,05945
EBITA	-24,0229	1,649411	-14,565	< 2 x 10 <sup>-16</sup>
STA	-0,323058	0,082401	-3,921	8,83 x 10 <sup>-5</sup>
OPM	0,58573	0,20716	2,827	0,00469
DER	0,030456	0,007547	4,036	5,45 x 10 <sup>-5</sup>
DAR	1,498677	0,186297	8,045	8,66 x 10 <sup>-16</sup>
QR	0,04853	0,022801	2,128	0,03330
RGDPG	-0,282062	0,125505	-2,247	0,02461
Inflasi	0,063064	0,016076	3,923	8,75 x 10 <sup>-5</sup>

*Size 2*

Prediktor	<i>Estimate</i>	<i>Std. Error</i>	<i>z value</i>	<i>Pr(&gt; z )</i>
<i>Intercept</i>	-0,6273	0,4142	-1,515	0,129886
EBITA	-4,248	1,192	-3,562	0,000368
STA	-0,1922	0,06951	-2,765	0,005698
NPM	-0,8804	0,3229	-2,727	0,006396
OPM	-3,64	0,9472	-3,842	0,000122
DAR	-0,5861	0,1759	3,332	0,000862
QR	-0,05539	0,02285	2,423	0,015375
WCA	-0,7346	0,2529	-2,905	0,003672
USD/IDR	7,872 x 10 <sup>-5</sup>	4,003 x 10 <sup>-5</sup>	1,966	0,049244

*Size 3*

Prediktor	Estimate	Std. Error	z value	Pr(> z )
Intercept	0,1168	0,1332	0,877	0,38055
EBITA	-4,8853	0,9157	-5,335	9,54 x 10 <sup>-8</sup>
OPM	-2,2349	0,7134	-3,133	0,00173
DAR	0,3381	0,1594	2,121	0,0339
WCA	-0,6507	0,2152	-3,024	0,00249

**Lampiran 19.** Estimasi Paramter Model dengan LOGIT Serentak

*Size 0*

Prediktor	Estimate	Odds Ratio	Std. Error	z value	Pr(> z )
Intercept	2,638	0,072	3,461	-0,762	0,44587
EBITA	27,75	8,88 x 10 <sup>-13</sup>	6,996	-3,966	0,00007
STA	0,7154	0,489	0,242	-2,956	0,00312
ITR	-0,01402	1,014	0,01765	0,794	0,42704
DSI	-0,0034	1,003	0,001962	1,734	0,08286
ROE	0,1189	0,888	0,2841	-0,418	0,67574
NPM	8,984	0,0001	3,044	-2,952	0,00316
OPM	26,53	3,01 x 10 <sup>-12</sup>	6,027	-4,402	0,00001
DER	-0,00642	1,006	0,015	0,428	0,66879
DAR	-1,632	5,114	0,4171	3,913	0,00009
CR	0,0568	0,945	0,1128	-0,504	0,61453
QR	-0,0615	1,063	0,114	0,539	0,58993
ETD	5,86	0,003	2,717	-2,157	0,031
WCA	0,966	0,381	0,9299	-1,039	0,29886
WCS	-0,1293	1,138	0,617	0,209	0,83407
RGDPG	-0,2358	1,266	0,3894	0,605	0,54493
BI7DRR	-0,1525	1,165	0,1273	1,198	0,23083
USD/IDR	-0,00014	1,0001	0,0001035	1,317	0,18792
Inflasi	0,013	0,987	0,05611	-0,231	0,81724

*Size 1*

Prediktor	<i>Estimate</i>	<i>Odds Ratio</i>	<i>Std. Error</i>	<i>z value</i>	<i>Pr(&gt; z )</i>
<i>Intercept</i>	-1,765	5,839	2,751	0,641	0,5213
EBITA	23,52	6,12 x 10 <sup>-11</sup>	1,889	-12,45	< 2 x 10 <sup>-16</sup>
STA	0,31	0,7334	0,1334	-2,324	0,0201
ITR	-0,00051	1,0005	0,00032	1,596	0,1104
DSI	-0,00028	1,0003	0,00092	0,301	0,7636
ROE	-0,3044	1,3558	0,231	1,318	0,1874
NPM	0,6622	0,5157	0,476	-1,392	0,1640
OPM	-0,9032	2,4674	0,367	2,458	0,0140
DER	-0,0338	1,0344	0,01334	2,532	0,0114
DAR	-1,419	4,1321	0,2821	5,030	4,91 x 10 <sup>-7</sup>
CR	-0,00981	1,0099	0,05121	0,192	0,8481
QR	-0,0192	1,0194	0,07174	0,268	0,7890
ETD	0,1958	0,8222	0,2452	-0,798	0,4248
WCA	0,1315	0,8768	0,4514	-0,291	0,7708
WCS	-0,0215	1,0217	0,1044	0,206	0,8370
RGDPG	0,3203	0,7260	0,2752	-1,164	0,2445
BI7DRR	0,1296	0,8784	0,1121	-1,156	0,2476
USD/IDR	-9,34 x 10 <sup>-6</sup>	1,000009	0,00011	0,089	0,9293
Inflasi	-0,045	1,0463	0,05431	0,833	0,4049

*Size 2*

Prediktor	<i>Estimate</i>	<i>Odds Ratio</i>	<i>Std. Error</i>	<i>z value</i>	<i>Pr(&gt; z )</i>
<i>Intercept</i>	5,649	0,00352	2,536	-2,228	0,0259
EBITA	14,76	3,89 x 10 <sup>-7</sup>	2,101	-7,024	2,15 x 10 <sup>-12</sup>
STA	0,2478	0,7805	0,1322	-1,875	0,0608
ITR	-0,00045	1,0005	0,00039	1,167	0,24308
DSI	-0,00061	1,0006	0,0009	0,679	0,49693
ROE	-0,0721	1,0747	0,2454	0,294	0,76902
NPM	0,5539	0,5747	0,5895	-0,940	0,34743

Prediktor	<i>Estimate</i>	<i>Odds Ratio</i>	<i>Std. Error</i>	<i>z value</i>	<i>Pr(&gt; z )</i>
OPM	1,151	0,3163	1,244	-0,926	0,35466
DER	-0,0043	1,0043	0,0124	0,349	0,72729
DAR	-0,812	2,2524	0,256	3,172	0,00152
CR	0,0827	0,9206	0,0599	-1,381	0,16729
QR	-0,194	1,2136	0,0883	2,192	0,02837
ETD	0,496	0,6091	0,2925	-1,695	0,09006
WCA	0,814	0,4429	0,4477	-1,819	0,06888
WCS	0,106	0,8998	0,093	-1,135	0,25641
RGDPG	-0,4203	1,5225	0,2555	1,645	0,10001
BI7DRR	-0,1167	1,1238	0,0991	1,178	0,23876
USD/IDR	-0,00026	1,0003	0,00011	2,410	0,01597
Inflasi	0,0705	0,9319	0,0495	-1,424	0,15458

Size 3

Prediktor	<i>Estimate</i>	<i>Odds Ratio</i>	<i>Std. Error</i>	<i>z value</i>	<i>Pr(&gt; z )</i>
<i>Intercept</i>	-5,166	175,166	3,879	1,332	0,18298
EBITA	11,31	1,23 x 10 <sup>-5</sup>	1,941	-5,824	5,75 x 10 <sup>-9</sup>
STA	0,1829	0,833	0,1296	-1,411	0,15821
ITR	-0,00025	1,0002	0,000366	0,677	0,49832
DSI	-0,00112	1,0011	0,00093	1,202	0,22926
ROE	0,192	0,8256	0,2217	-0,865	0,38729
NPM	-0,1533	1,1657	0,3051	0,502	0,61536
OPM	1,213	0,2972	1,137	-1,067	0,28588
DER	-0,0152	1,0153	0,0115	1,318	0,18762
DAR	-0,598	1,8189	0,2416	2,476	0,01327
CR	0,0197	0,9805	0,03779	-0,521	0,60254
QR	-0,0411	1,042	0,05498	0,749	0,45413
ETD	0,1365	0,8724	0,1787	-0,764	0,44512
WCA	0,8199	0,4405	0,4105	-1,997	0,04579
WCS	0,0179	0,9823	0,0892	-0,200	0,84122

Prediktor	<i>Estimate</i>	<i>Odds Ratio</i>	<i>Std. Error</i>	<i>z value</i>	<i>Pr(&gt; z )</i>
RGDPG	0,1556	0,8559	0,2904	-0,536	0,59212
BI7DRR	0,2539	0,7758	0,1051	-2,416	0,01571
USD/IDR	0,00043	0,9996	0,0002	-1,825	0,06797
Inflasi	-0,1464	1,1577	0,0521	2,812	0,00492

**Lampiran 20.** Estimasi Paramter Model dengan *Stepwise LOGIT Size 0*

Prediktor	<i>Estimate</i>	<i>Odds Ratio</i>	<i>Std. Error</i>	<i>z value</i>	<i>Pr(&gt; z )</i>
<i>Intercept</i>	-0,229	1,257	0,5564	0,412	0,68070
EBITA	28,7552	$3,25 \times 10^{-13}$	0,6448	-4,327	$1,51 \times 10^{-5}$
STA	0,746	0,474	0,2342	-3,186	0,00144
ITR	-0,0146	1,015	0,0167	0,872	0,38343
DSI	-0,0027	1,003	0,0018	1,477	0,13968
NPM	8,443	0,0002	3,0315	-2,785	0,00535
OPM	25,1997	$1,14 \times 10^{-11}$	5,5012	-4,581	$4,63 \times 10^{-6}$
DAR	-1,896	6,659	0,3729	5,085	$3,69 \times 10^{-7}$
ETD	6,8155	0,001	2,8187	-2,418	0,01561
BI7DRR	-0,0985	1,104	0,0578	1,7004	0,08847

*Size 1*

Prediktor	<i>Estimate</i>	<i>Odds Ratio</i>	<i>Std. Error</i>	<i>z value</i>	<i>Pr(&gt; z )</i>
<i>Intercept</i>	-1,039	2,827	0,9521	1,091	0,27507
EBITA	24,08	$3,47 \times 10^{-11}$	1,8	-13,38	$< 2 \times 10^{-16}$
STA	0,3144	0,7302	0,1226	-2,564	0,01034
ITR	-0,0005	1,0005	0,00024	1,994	0,04613
ROE	-0,3065	1,3587	0,2319	1,322	0,18624
NPM	0,8352	0,4338	0,4411	-1,893	0,05829
OPM	-1,015	2,7582	0,3413	2,972	0,00296
DER	-0,0333	1,0338	0,0133	2,505	0,01223
DAR	-1,462	4,3152	0,2339	6,250	$4,1 \times 10^{-10}$

Prediktor	<i>Estimate</i>	<i>Odds Ratio</i>	<i>Std. Error</i>	<i>z value</i>	<i>Pr(&gt; z )</i>
QR	-0,0288	1,0292	0,0233	1,237	0,21617
RGDPG	0,2661	0,7664	0,1644	-1,618	0,10556

*Size 2*

Prediktor	<i>Estimate</i>	<i>Odds Ratio</i>	<i>Std. Error</i>	<i>z value</i>	<i>Pr(&gt; z )</i>
<i>Intercept</i>	4,175	0,0154	1,777	-2,450	0,018795
EBITA	16,67	$5,76 \times 10^{-8}$	1,535	-10,86	$< 2 \times 10^{-16}$
STA	0,2612	0,7701	0,1245	-2,098	0,035948
DAR	-0,8334	2,3011	0,2505	3,327	0,000879
CR	0,0744	0,9283	0,0531	-1,401	0,161297
QR	-0,1931	1,2130	0,0772	2,501	0,012375
ETD	0,4742	0,6224	0,1934	-2,452	0,014208
WCA	0,9842	0,3737	0,4319	-2,279	0,022675
RGDPG	-0,3614	1,4353	0,1995	1,812	0,070044
USD/IDR	-0,0002	1,0002	0,0001	2,111	0,034737

*Size 3*

Prediktor	<i>Estimate</i>	<i>Odds Ratio</i>	<i>Std. Error</i>	<i>z value</i>	<i>Pr(&gt; z )</i>
<i>Intercept</i>	-3,345	28,3606	1,552	2,154	0,03120
EBITA	11,39	0,00001	1,789	-6,369	$1,9 \times 10^{-10}$
STA	0,2472	0,7810	0,1254	-1,971	0,04873
OPM	1,487	0,2260	0,8861	-1,679	0,09322
DER	-0,0197	1,0199	0,0131	1,499	0,13388
DAR	-0,6481	1,9119	0,2422	2,679	0,00744
QR	-0,0244	1,0247	0,0213	1,144	0,25251
WCA	0,7934	0,4523	0,3855	-2,058	0,03958
BI7DRR	0,2167	0,8052	0,0841	-2,576	0,00998
USD/IDR	0,0003	0,9997	0,0002	-2,070	0,03848
Inflasi	-0,132	1,1411	0,0466	2,834	0,00460

**Lampiran 21.** Estimasi Paramter Model dengan *Lasso* LOGIT

Prediktor	<i>Size</i> 0	<i>Size</i> 1	<i>Size</i> 2	<i>Size</i> 3
<i>Intercept</i>	-0,981	-0,8892	-0,6702	-0,4421
EBITA	-39,19	-20,0159	-14,8551	-10,1025
STA	-0,423	-0,2358	-0,1564	-0,0103
ITR	0,0051	0,0002	0,0001	0,00003
DSI	0,0023	0	0	0
ROE	0	0	0	0
NPM	-8,7615	0	-0,13998	0
OPM	-12,0332	0,0845	0	0
DER	0	0	0	0
DAR	1,7454	1,1936	0,6705	0,271
CR	0	0,0004	0	0
QR	0,003	0	0,0036	0
ETD	-1,0785	-0,0744	-0,0662	0
WCA	-0,7899	0	-0,4771	-0,3965
WCS	0	0,0194	0	0
RGDPG	0	-0,038	0	0
BI7DRR	0,091	0	0	0
USD/IDR	0,00009	0,00007	0,00004	0
Inflasi	0	0	0	0

**Lampiran 22.** Estimasi Paramter Model dengan *Elastic-Net* LOGIT

Prediktor	<i>Size</i> 0	<i>Size</i> 1	<i>Size</i> 2	<i>Size</i> 3
<i>Intercept</i>	-1,1492	0,0142	-3,5642	-0,3029
EBITA	-40,172	-19,165	-12,487	-9,638
STA	-0,3375	-0,2974	-0,2856	-0,1804
ITR	0,0038	0,0003	0,0004	0,0002
DSI	0,002	0,0002	0,0002	0,0007
ROE	-0,0044	0	0	0
NPM	-6,3812	-0,1819	-0,5468	0

Prediktor	<i>Size 0</i>	<i>Size 1</i>	<i>Size 2</i>	<i>Size 3</i>
OPM	-7,4051	0,3242	-1,0854	-0,8221
DER	0	0,0074	0,0016	0,0012
DAR	1,6707	1,2347	0,8069	0,5662
CR	0	0,009	0,0015	0,0031
QR	0,0033	0	0,0292	0,0033
ETD	-0,8058	-0,1752	-0,1899	-0,061
WCA	-0,7223	-0,2575	-0,7667	-0,7581
WCS	0	0,0408	-0,0241	0
RGDPG	0	-0,1701	0,2362	0,1074
BI7DRR	0,082	-0,0223	0,0464	-0,0289
USD/IDR	0,00009	0,00006	0,0002	-0,00009
Inflasi	0	0	-0,0371	0,0267

**Lampiran 23.** Surat Keterangan Data**SURAT PERNYATAAN**

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

Nama :Rizki Nanda Savera

NRP :06211640000084

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 :Bank Indonesia

Keterangan :Data rasio keuangan perusahaan publik dan data makroekonomi

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

Mengetahui

Pembimbing Tugas Akhir

Surabaya, 12 Juni 2020



(Dr. rer. pol. Dedy Dwi Prastyo, S.Si., M.Si.)  
NIP. 19831204 200812 1 002

(Rizki Nanda Savera)  
NRP. 06211640000084

\*(coret yang tidak perlu)

## BIODATA PENULIS



Penulis yaitu Rizki Nanda Savera atau akrab dipanggil dengan Vera adalah mahasiswa Statistika angkatan 2016. Pendidikan yang telah ditempuh oleh penulis adalah MI HM Pakel (2004-2010), MTsN 1 Blitar (2010-2013), dan MAN 3 Blitar (2013-2016). Kemudian dilanjutkan dengan menempuh pendidikan di Institut Teknologi Sepuluh Nopember Departemen Statistika melalui jalur penerimaan SBMPTN 2016. Selama perkuliahan berlangsung, penulis yang memiliki hobi dibidang olahraga dan menulis ini menyalurkan hobinya melalui organisasi dengan bergabung di Divisi Futsal Putri UKM Sepakbola ITS sebagai staf pada tahun kepengurusan 2017-2018 dan dilanjutkan sebagai Ketua Divisi pada tahun kepengurusan 2018-2019. Hobi dibidang menulis, penulis menyalurkan melalui bergabungnya dengan Divisi Pers HIMASTA-ITS tahun 2017-2018 dan dilanjutkan menjadi Ketua Divisi pada 2018-2019. Selain itu penulis juga aktif dalam bidang keagamaan pada tahun 2017-2018 dengan bergabung pada organisasi FORSIS-ITS. Untuk menyalurkan bakat dalam bidang olahraga penulis aktif sebagai anggota Futsal Putri ITS dalam kejuaraan Liga Mahasiswa cabang olahraga futsal dan UGMFC pada tahun 2017, 2018, dan 2019 serta kejuaran-kejuaraan lainnya. Selain mengikuti kejuaraan nasional, penulis juga aktif mengikuti turnamen dilingkungan ITS pada cabang olahraga Futsal dan Basket. Selain itu dibidang akademik, penulis dengan rekan tim berhasil menjadi semifinalis *National Statistics Challenge* di Universitas Brawijaya pada tahun 2019. Jika ingin memberikan saran, kritik, dan diskusi lebih lanjut, dapat menghubungi melalui email: rizkin49@gmail.com atau nomor telepon 081232240299.