TUGAS AKHIR – KS 184822

PREDIKSI *FINANCIAL DISTRESS* PERUSAHAAN SEKTOR INDUSTRI DI INDONESIA DENGAN METODE KLASIFIKASI DAN MELIBATKAN *SYNTHETIC FEATURES GENERATION*

Muhammad Adlansyah Muda
NRP 062117 4000 0074

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 2021

TUGAS AKHIR – KS 184822

PREDIKSI *FINANCIAL DISTRESS* PERUSAHAAN SEKTOR INDUSTRI DI INDONESIA DENGAN METODE KLASIFIKASI DAN MELIBATKAN *SYNTHETIC FEATURES GENERATION*

Muhammad Adlansyah Muda
NRP 062117 4000 0074

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 2021

(*Halaman ini sengaja dikosongkan*)

FINAL PROJECT – KS 184822

PREDICTION OF FINANCIAL DISTRESS FOR INDUSTRIAL SECTOR COMPANIES IN INDONESIA WITH CLASSIFICATION METHODS AND INVOLVING SYNTHETIC FEATURES GENERATION

Muhammad Adlansyah Muda
SN 062117 4000 0074

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 2021

(*Halaman ini sengaja dikosongkan*)

LEMBAR PENGESAHAN

PREDIKSI *FINANCIAL DISTRESS* PERUSAHAAN SEKTOR INDUSTRI DI INDONESIA DENGAN METODE KLASIFIKASI DAN MELIBATKAN *SYNTHETIC FEATURES GENERATION*

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:
Muhammad Adlansyah Muda
NRP. 062117 4000 0074

Disetujui oleh Pembimbing:
Dr. rer. pol. Dedy Dwi Prastyo, S.Si., M.Si
NIP. 19831204 200812 1 002

KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
FAKULTAS SAINS DAN ANALITIKA DATA
DEPARTEMEN STATISTIKA

Mengetahui,
Kepala Departemen Statistika

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

SURABAYA, JANUARI 2021

(*Halaman ini sengaja dikosongkan*)

PREDIKSI *FINANCIAL DISTRESS* PERUSAHAAN SEKTOR INDUSTRI DI INDONESIA DENGAN METODE KLASIFIKASI DAN MELIBATKAN *SYNTHETIC FEATURES GENERATION*

Nama Mahasiswa : Muhammad Adlansyah Muda
NRP : 062117 4000 0074
Departemen : Statistika
Dosen Pembimbing : Dr.rer.pol. Dedy Dwi Prastyo, S.Si., M.Si

Abstrak

Masalah kondisi financial distress dapat berakhir dengan kebangkrutan apabila tidak segera ditanggulangi. Untuk mengantisipasi dan meminimalisir dampak dari bangkrutnya suatu perusahaan terutama pada sektor industri, maka dilakukan prediksi financial distress untuk menilai kondisi keuangan perusahaan dan perspektif masa depannya. Pada penelitian ini prediksi financial distress dilakukan dengan metode klasifikasi seperti Generalized Extreme Value Regression, Logistic Regression, Support Vector Machine, dan Extreme Gradient Boosting dengan melibatkan synthetic features generation secara serentak dan seleksi variabel. Berdasarkan nilai accuracy, AUC, dan F1-score dari hasil evaluasi model menggunakan data testing didapatkan bahwa metode synthetic features generation tidak selalu memberikan performansi klasifikasi terbaik pada tiap size. Pada size 0 dan size 1 disimpulkan bahwa model Extreme Gradient Boosting dengan melibatkan synthetic features generation dan seleksi variabel merupakan model dengan performansi klasifikasi terbaik, sedangkan pada size 2 dan size 3 didapatkan bahwa model Extreme Gradient Boosting tanpa melibatkan synthetic features generation dengan seleksi variabel merupakan model dengan performansi klasifikasi terbaik dalam memprediksi kondisi keuangan perusahaan sektor industri di Indonesia.

Kata Kunci: Drift, Financial Distress, Klasifikasi, Sektor Industri, Synthetic Features Generation

(*Halaman ini sengaja dikosongkan*)

PREDICTION OF FINANCIAL DISTRESS FOR INDUSTRIAL SECTOR COMPANIES IN INDONESIA WITH CLASSIFICATION METHODS AND INVOLVING SYNTHETIC FEATURES GENERATION

Name : **Muhammad Adlansyah Muda**
Student Number : **062117 4000 0074**
Department : **Statistika**
Supervisor : **Dr.rer.pol. Dedy Dwi Prastyo, S.Si., M.Si**

Abstract

The problem of financial distress can lead to bankruptcy if it is not addressed immediately. To anticipate and minimize the impact of a company bankruptcy, especially in the industrial sector, financial distress predictions are made to assess the company's financial condition and its future perspective. In this study, prediction of financial distress is carried out using classification methods such as Generalized Extreme Value Regression, Logistic Regression, Support Vector Machine, and Extreme Gradient Boosting by involving synthetic features generation simultaneously and variable selection. Based on the accuracy, AUC, and F1-score from the results of model evaluation using data testing, it is found that the synthetic features generation method does not always provide the best classification performance for each size. At size 0 and size 1, it can be concluded that the Extreme Gradient Boosting model involving synthetic features generation with variable selection is the model with the best classification performance, whereas at size 2 and size 3, it is found that the Extreme Gradient Boosting model without involving synthetic features generation with variabel selection is the model with the best classification performance in predicting the financial condition of industrial sector companies in Indonesia.

Keywords: Classification, Drift, Financial Distress, Industrial Sector, Synthetic Features Generation

(*Halaman ini sengaja dikosongkan*)

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat dan hidayahnya, sehingga dapat menyelesaikan laporan Tugas Akhir ini yang berjudul **"Prediksi *Financial Distress* Perusahaan Sektor Industri di Indonesia dengan Metode Klasifikasi dan Melibatkan *Synthetic Features Generation*"** dengan tepat waktu.

Laporan Tugas Akhir ini dapat terselesaikan dengan baik atas bantuan, motivasi, dan dukungan dari berbagai pihak baik secara langsung maupun tidak langsung kepada penulis. Oleh karena itu penulis ingin mengucapkan terima kasih kepada:

1. Orang tua dan keluarga penulis yang telah memberikan dukungan dan doa selama menyusun laporan Tugas Akhir.
2. Ibu Dr. Dra. Kartika Fithriasari, M.Si selaku Kepala Departemen Statistika FSAD ITS, Ibu Dr. Santi Wulan Purnami, S.Si., M.Si selaku Sekretaris Departemen I (Bidang Akademik, Kemahasiswaan, Penelitian dan Pengabdian Kepada Masyarakat), dan Ibu Dr. Vita Ratnasari, S.Si., M.Si selaku Sekretaris Departemen II (Bidang Sumber Daya Keuangan, Sumber Daya Manusia, dan Sarana Prasarana) yang telah memfasilitasi sarana dan prasarana untuk kegiatan belajar di Statistika ITS.
3. Bapak Dr. rer. pol. Dedy Dwi Prastyo, S.Si., M.Si selaku dosen pembimbing utama yang telah memberikan kesempatan kepada penulis untuk meneliti kasus pada penelitian ini dan memberikan bimbingan serta saran dalam penyusunan laporan Tugas Akhir.
4. Almarhum Bapak Dr. Suhartono, S.Si., M.Sc selaku dosen pembimbing kedua hingga proposal Tugas Akhir yang telah memberikan bimbingan, saran, dan motivasi dalam penyusunan proposal Tugas Akhir.
5. Ibu Santi Puteri Rahayu, S.Si., M.Si., Ph.D dan Bapak M. Sjahid Akbar, S.Si., M.Si selaku dosen penguji yang telah memberikan masukan dalam penyusunan laporan Tugas Akhir.

6. Bapak Dr. Achmad Choiruddin, S.Si., M.Sc., selaku dosen wali yang telah memberikan motivasi dan doa selama proses perkuliahan di Departemen Statistika.
7. Yama sebagai sahabat terbaik di Departemen Statistika yang selalu memberikan pencerahan dan motivasi selama proses perkuliahan hingga penyusunan laporan Tugas Akhir ini.
8. Indra, Alvian, Rofiq, Angga, dan Hardan sebagai teman ambis dan *gamer* di Departemen Statistika yang selalu membantu dan memberikan pencerahan selama proses perkuliahan.
9. Radha, Nanda, Adilia, Odita, Okka, Reza, Yana, Tata, dan Nadya yang selalu ada ketika kesulitan di Departemen Statistika dan selalu memberikan motivasi dalam penyusunan laporan Tugas Akhir.
10. Teman-teman Gamanta 2017 yang selalu membantu penulis di Departemen Statistika dan memberikan semangat dalam penyusunan laporan Tugas Akhir.
11. Dika, Daffa, Boim, Fajri, Gilang, Raja, Ifan, Bayu, Raji, dan teman-teman SMA lainnya yang telah memberikan semangat dan dukungan kepada penulis dalam penyusunan laporan Tugas Akhir.
12. Serta pihak-pihak lain yang tidak dapat disebutkan satu per satu oleh penulis.

Penulis menyadari bahwa masih terdapat kekurangan di dalam penulisan dan penyusunan laporan Tugas Akhir ini, oleh karena itu penulis mengharapkan saran dan kritik yang membangun dari pembaca sebagai perbaikan laporan Tugas Akhir ini. Harapan penulis semoga laporan Tugas Akhir ini dapat bermanfaat bagi para pembaca dan khususnya bagi penulis.

Surabaya, Januari 2021

Penulis

DAFTAR ISI

HALAMAN JUDUL i
LEMBAR PENGESAHAN vii
KATA PENGANTAR xii
DAFTAR ISI xv
DAFTAR TABEL xix
DAFTAR GAMBAR xxv
DAFTAR LAMPIRAN xxvii
BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 5
1.3 Tujuan 7
1.4 Manfaat 7
1.5 Batasan Masalah 7
BAB II TINJAUAN PUSTAKA 9
2.1 Imputasi *k-Nearest Neighbor* 9
2.2 *Synthetic Features Generation* 10
2.3 Konsep *Drift* 10
2.4 *Generalized Extreme Value Regression* (GEVR) 11
2.5 *Logistic Regression* 14
2.6 *Regularized Logistic Regression* 17
2.6.1 *Lasso Logistic Regression* 19
2.6.2 *Elastic-Net Logistic Regression* 19
2.7 *Support Vector Machine* (SVM) 21
2.7.1 *Support Vector Machine* (SVM) dalam Klasifikasi Linier 21
2.7.2 *Support Vector Machine* (SVM) dalam Klasifikasi Nonlinier 24
2.8 *Regularized Support Vector Machine* 25
2.8.1 *Lasso Support Vector Machine* 25
2.8.2 *Elastic-Net Support Vector Machine* 26
2.9 *Extreme Gradient Boosting* (XGB) 26
2.10 Prosedur *Stepwise* 29
2.11 Uji Signifikansi Parameter 30

2.12 Evaluasi Hasil Klasifikasi 31
2.13 *Financial Distress* 32
a. Likuiditas 33
b. Profitabilitas 33
c. *Leverage* 33
d. Arus Kas Operasi 34
2.14 Rasio Keuangan 34
a. Rasio Aktivitas 34
b. Rasio Profitabilitas 35
c. Rasio Solvabilitas 35
d. Rasio Likuiditas 35
2.15 Indikator Ekonomi Makro 35
BAB III METODOLOGI PENELITIAN 37
3.1 Sumber Data 37
3.2 Langkah Analisis 44
BAB IV ANALISIS DAN PEMBAHASAN 53
4.1 Metode Klasifikasi Tanpa *Synthetic Features Generation* 53
4.1.1 *Extreme Gradient Boosting* (XGB) Secara Serentak 53
4.1.2 *Extreme Gradient Boosting* (XGB) dengan Seleksi Variabel 55
4.2 Metode Klasifikasi dengan *Synthetic Features Generation* 58
4.2.1 *Generalized Extreme Value Regression* dengan *Synthetic Features Generation* Secara Serentak . 59
4.2.2 *Generalized Extreme Value Regression* (GEVR) dengan *Synthetic Features Generation* dan Seleksi Variabel 64
4.2.3 *Logistic Regression* dengan *Synthetic Features Generation* Secara Serentak 67
4.2.4 *Logistic Regression* dengan *Synthetic Features Generation* dan Seleksi Variabel 73
4.2.5 *Lasso Logistic Regression* dengan *Synthetic Features Generation* 77

4.2.6 *Elastic-Net Logistic Regression* dengan *Synthetic Features Generation* 81
4.2.7 *Support Vector Machine* (SVM) dengan *Synthetic Features Generation* 85
4.2.8 *Lasso Support Vector Machine* dengan *Synthetic Features Generation* 87
4.2.9 *Elastic-Net Support Vector Machine* dengan *Synthetic Features Generation*..... 90
4.2.10 *Extreme Gradient Boosting* (XGB) dengan *Synthetic Features Generation* Secara Serentak . 93
4.2.11 *Extreme Gradient Boosting* (XGB) dengan *Synthetic Features Generation* dan Seleksi Variabel 95
4.3 Menentukan Model Terbaik 98
4.3.1 Perbandingan Model Secara Serentak..... 99
4.3.2 Perbandingan Model dengan Seleksi Variabel . 106
4.3.3 Perbandingan Model Secara Serentak dan Model dengan Seleksi Variabel..... 114
BAB V KESIMPULAN DAN SARAN..... 123
5.1 Kesimpulan 123
5.2 Saran..... 126
DAFTAR PUSTAKA 127
LAMPIRAN 131
BIODATA PENULIS 182

(*Halaman ini sengaja dikosongkan*)

DAFTAR TABEL

Tabel 2.1 *Confusion Matrix*. 31
Tabel 3.1 Variabel Penelitian..... 37
Tabel 3.2 Struktur Data Penelitian.....42
Tabel 3.3 Perbandingan Jumlah Unit Observasi Tiap *Size*. 43
Tabel 3.4 Perbandingan Langkah Analisis dengan Penelitian Sebelumnya. 51
Tabel 4.1a *Confusion Matrix* Model XGB Tanpa *Synthetic Features Generation* Untuk *Size* 0 53
Tabel 4.1b *Confusion Matrix* Model XGB Tanpa *Synthetic Features Generation* Untuk *Size* 1 53
Tabel 4.1c *Confusion Matrix* Model XGB Tanpa *Synthetic Features Generation* Untuk *Size* 2 53
Tabel 4.1d *Confusion Matrix* Model XGB Tanpa *Synthetic Features Generation* Untuk *Size* 3 54
Tabel 4.2 Evaluasi Hasil Klasifikasi Model XGB Tanpa *Synthetic Features Generation*. 54
Tabel 4.3 Nilai *Gain* Tiap Variabel Prediktor Model XGB Tanpa *Synthetic Features Generation*.....55
Tabel 4.4 Variabel Prediktor yang Signifikan dalam Model XGB Tanpa *Synthetic Features Generation*.....56
Tabel 4.5a *Confusion Matrix* Model XGB Tanpa *Synthetic Features Generation* dengan Seleksi Variabel Untuk *Size* 0.....57
Tabel 4.5b *Confusion Matrix* Model XGB Tanpa *Synthetic Features Generation* dengan Seleksi Variabel Untuk *Size* 1.....57
Tabel 4.5c *Confusion Matrix* Model XGB Tanpa *Synthetic Features Generation* dengan Seleksi Variabel Untuk *Size* 2.....57
Tabel 4.5d *Confusion Matrix* Model XGB Tanpa *Synthetic Features Generation* dengan Seleksi Variabel Untuk *Size* 3.....57

Tabel 4.6 Evaluasi Hasil Klasifikasi Model XGB Tanpa *Synthetic Features Generation* dengan Seleksi Variabel.....58
Tabel 4.7 Model GEVR dengan *Synthetic Features Generation* Tiap Size. 59
Tabel 4.8 Variabel Signifikan Dalam Model GEVR Secara Serentak.....61
Tabel 4.9a *Confusion Matrix* Model GEVR dengan *Synthetic Features Generation* Untuk *Size* 0.....62
Tabel 4.9b *Confusion Matrix* Model GEVR dengan *Synthetic Features Generation* Untuk *Size* 1.....62
Tabel 4.9c *Confusion Matrix* Model GEVR dengan *Synthetic Features Generation* Untuk *Size* 2.....62
Tabel 4.9d *Confusion Matrix* Model GEVR dengan *Synthetic Features Generation* Untuk *Size* 3.....62
Tabel 4.10 Evaluasi Hasil Klasifikasi Model GEVR dengan *Synthetic Features Generation* Secara Serentak.....63
Tabel 4.11 Variabel Signifikan Dalam Model GEVR dengan *Stepwise*. 64
Tabel 4.12 Model GEVR dengan *Synthetic Features Generation* dan Seleksi Variabel Tiap Size. 65
Tabel 4.13a *Confusion Matrix* Model GEVR dengan *Synthetic Features Generation* dan Seleksi Variabel Untuk *Size* 0 66
Tabel 4.13b *Confusion Matrix* Model GEVR dengan *Synthetic Features Generation* dan Seleksi Variabel Untuk *Size* 1 66
Tabel 4.13c *Confusion Matrix* Model GEVR dengan *Synthetic Features Generation* dan Seleksi Variabel Untuk *Size* 2 66
Tabel 4.13d *Confusion Matrix* Model GEVR dengan *Synthetic Features Generation* dan Seleksi Variabel Untuk *Size* 3 66
Tabel 4.14 Evaluasi Hasil Klasifikasi Model GEVR dengan *Synthetic Features Generation* dan Seleksi Variabel 67

Tabel 4.15 Model *Logistic Regression* dengan *Synthetic Features Generation* Secara Serentak.....68
Tabel 4.16 Variabel Signifikan Dalam Model *Logistic Regression* Secara Serentak.....70
Tabel 4.17a *Confusion Matrix* Model *Logistic Regression* dengan *Synthetic Features Generation* Secara Serentak Untuk *Size* 0. 71
Tabel 4.17b *Confusion Matrix* Model *Logistic Regression* dengan *Synthetic Features Generation* Secara Serentak Untuk *Size* 1 71
Tabel 4.17c *Confusion Matrix* Model *Logistic Regression* dengan *Synthetic Features Generation* Secara Serentak Untuk *Size* 2 71
Tabel 4.17d *Confusion Matrix* Model *Logistic Regression* dengan *Synthetic Features Generation* Secara Serentak Untuk *Size* 3 71
Tabel 4.18 Evaluasi Hasil Klasifikasi Model *Logistic Regression* dengan *Synthetic Features Generation* Secara Serentak 72
Tabel 4.19 Variabel Signifikan Dalam Model *Logistic Regression* dengan *Stepwise*.....73
Tabel 4.20 Model *Logistic Regression* dengan *Synthetic Features Generation* dan Seleksi Variabel. 74
Tabel 4.21a *Confusion Matrix* Model *Logistic Regression* dengan *Synthetic Features Generation* dan Seleksi Variabel Untuk *Size* 0. 75
Tabel 4.21b *Confusion Matrix* Model *Logistic Regression* dengan *Synthetic Features Generation* dan Seleksi Variabel Untuk *Size* 1 75
Tabel 4.21c *Confusion Matrix* Model *Logistic Regression* dengan *Synthetic Features Generation* dan Seleksi Variabel Untuk *Size* 2 75
Tabel 4.21d *Confusion Matrix* Model *Logistic Regression* dengan *Synthetic Features Generation* dan Seleksi Variabel Untuk *Size* 3 76

Tabel 4.22 Evaluasi Hasil Klasifikasi Model *Logistic Regression* dengan *Synthetic Features Generation* dan Seleksi Variabel.....76
Tabel 4.23 Variabel Model *Lasso Logistic Regression* dengan *Synthetic Features Generation*. 77
Tabel 4.24 Model *Lasso Logistic Regression* dengan *Synthetic Features Generation*. 78
Tabel 4.25a *Confusion Matrix* Model *Lasso Logistic Regression* dengan *Synthetic Features Generation* Untuk *Size* 0 79
Tabel 4.25b *Confusion Matrix* Model *Lasso Logistic Regression* dengan *Synthetic Features Generation* Untuk *Size* 1 79
Tabel 4.25c *Confusion Matrix* Model *Lasso Logistic Regression* dengan *Synthetic Features Generation* Untuk *Size* 2 79
Tabel 4.25d *Confusion Matrix* Model *Lasso Logistic Regression* dengan *Synthetic Features Generation* Untuk *Size* 3 79
Tabel 4.26 Evaluasi Hasil Klasifikasi Model *Lasso Logistic Regression* dengan *Synthetic Features Generation.* 80
Tabel 4.27 Variabel Model *Elastic-Net Logistic Regression* dengan *Synthetic Features Generation*. 81
Tabel 4.28 Model *Elastic-Net Logistic Regression* dengan *Synthetic Features Generation*. 82
Tabel 4.29a *Confusion Matrix* Model *Elastic-Net Logistic Regression* dengan *Synthetic Features Generation* Untuk Size 0. 84
Tabel 4.29b *Confusion Matrix* Model *Elastic-Net Logistic Regression* dengan *Synthetic Features Generation* Untuk *Size* 1. 84
Tabel 4.29c *Confusion Matrix* Model *Elastic-Net Logistic Regression* dengan *Synthetic Features Generation* Untuk *Size* 2. 84

Tabel 4.29d *Confusion Matrix* Model *Elastic-Net Logistic Regression* dengan *Synthetic Features Generation* Untuk *Size* 3. 84
Tabel 4.30 Evaluasi Hasil Klasifikasi Model *Elastic-Net Logistic Regression* dengan *Synthetic Features Generation*. 85
Tabel 4.31a *Confusion Matrix* Model SVM dengan *Synthetic Features Generation* Untuk *Size* 0. 86
Tabel 4.31b *Confusion Matrix* Model SVM dengan *Synthetic Features Generation* Untuk *Size* 1 86
Tabel 4.31c *Confusion Matrix* Model SVM dengan *Synthetic Features Generation* Untuk *Size* 2 86
Tabel 4.31d *Confusion Matrix* Model SVM dengan *Synthetic Features Generation* Untuk *Size* 3 86
Tabel 4.32 Evaluasi Hasil Klasifikasi Model SVM dengan *Synthetic Features Generation*. 87
Tabel 4.33 Variabel Model *Lasso* SVM dengan *Synthetic Features Generation*. 88
Tabel 4.34a *Confusion Matrix* Model *Lasso* SVM dengan *Synthetic Features Generation* Untuk *Size* 0. 89
Tabel 4.34b *Confusion Matrix* Model *Lasso* SVM dengan *Synthetic Features Generation* Untuk *Size* 1 89
Tabel 4.34c *Confusion Matrix* Model *Lasso* SVM dengan *Synthetic Features Generation* Untuk *Size* 2 89
Tabel 4.34d *Confusion Matrix* Model *Lasso* SVM dengan *Synthetic Features Generation* Untuk *Size* 3 89
Tabel 4.35 Evaluasi Hasil Klasifikasi Model *Lasso* SVM dengan *Synthetic Features Generation*. 90
Tabel 4.36 Variabel Model *Elastic-Net Support Vector Machine* dengan *Synthetic Features Generation*. 91
Tabel 4.37a *Confusion Matrix* Model *Elastic-Net* SVM dengan *Synthetic Features Generation* Untuk *Size* 0. 91
Tabel 4.37b *Confusion Matrix* Model *Elastic-Net* SVM dengan *Synthetic Features Generation* Untuk *Size* 1 92
Tabel 4.37c *Confusion Matrix* Model *Elastic-Net* SVM dengan *Synthetic Features Generation* Untuk *Size* 2 92

Tabel 4.37d *Confusion Matrix* Model *Elastic-Net* SVM dengan *Synthetic Features Generation* Untuk *Size* 3 92
Tabel 4.38 Evaluasi Hasil Klasifikasi Model *Elastic-Net* SVM dengan *Synthetic Features Generation*. 92
Tabel 4.39a *Confusion Matrix* Model XGB dengan *Synthetic Features Generation* Untuk Size 0.....94
Tabel 4.39b *Confusion Matrix* Model XGB dengan *Synthetic Features Generation* Untuk *Size* 1 94
Tabel 4.39c *Confusion Matrix* Model XGB dengan *Synthetic Features Generation* Untuk *Size* 2 94
Tabel 4.39d *Confusion Matrix* Model XGB dengan *Synthetic Features Generation* Untuk *Size* 3 94
Tabel 4.40 Evaluasi Hasil Klasifikasi Model XGB dengan *Synthetic Features Generation* Secara Serentak. 94
Tabel 4.41 Variabel Model XGB dengan *Synthetic Features Generation* dan Seleksi Variabel. 96
Tabel 4.42a *Confusion Matrix* Model XGB dengan *Synthetic Features Generation* dan Seleksi Variabel Untuk Size 0.....97
Tabel 4.42b *Confusion Matrix* Model XGB dengan *Synthetic Features Generation* dan Seleksi Variabel Untuk *Size* 1 97
Tabel 4.42c *Confusion Matrix* Model XGB dengan *Synthetic Features Generation* dan Seleksi Variabel Untuk *Size* 2 97
Tabel 4.42d *Confusion Matrix* Model XGB dengan *Synthetic Features Generation* dan Seleksi Variabel Untuk *Size* 3 97
Tabel 4.43 Evaluasi Hasil Klasifikasi Model XGB dengan *Synthetic Features Generation* dan Seleksi Variabel. 98
Tabel 4.44 Perbandingan Model Secara Serentak. 99
Tabel 4.45 Perbandingan Model dengan Seleksi Variabel. 106
Tabel 4.46 Perbandingan *Sensitivity* Model Secara Serentak dan Seleksi Variabel 114

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi Konsep *Drift*.....11
Gambar 2.2 Perbandingan $\pi(s)$ GEVR dan LOGIT 12
Gambar 2.3 Perbandingan Penalti *Lasso* dan *Elastic-Net*.....20
Gambar 2.4 Fungsi Klasifikasi untuk Data yang Dapat Dipisahkan Secara Linier 21
Gambar 2.5 Pemisah *Hyperplane* $x^T w + b = 0$ dan Margin yang Dapat Dipisahkan Secara Linier.....22
Gambar 2.6 Ilustrasi *Support Vector Machine* (SVM) dalam Klasifikasi Nonliner 24
Gambar 3.1 Ilustrasi Konsep *Drift* dengan Size 0, Size 1, Size 2, dan Size 3. 44
Gambar 3.2 Ilustrasi *Synthetic Features Generation*. 44
Gambar 3.3 Diagram Alir. 50
Gambar 4.1 Perbandingan *Accuracy*, AUC, dan *F*1-*score* Model Secara Serentak Untuk Size 0. 101
Gambar 4.2 Perbandingan *Accuracy*, AUC, dan *F*1-*score* Model Secara Serentak Untuk Size 1. 102
Gambar 4.3 Perbandingan *Accuracy*, AUC, dan *F*1-*score* Model Secara Serentak Untuk Size 2. 104
Gambar 4.4 Perbandingan *Accuracy*, AUC, dan *F*1-*score* Model Secara Serentak Untuk Size 3. 105
Gambar 4.5 Perbandingan *Accuracy*, AUC, dan *F*1-*score* Model dengan Seleksi Variabel Untuk Size 0. 109
Gambar 4.6 Perbandingan *Accuracy*, AUC, dan *F*1-*score* Model dengan Seleksi Variabel Untuk Size 1. 110
Gambar 4.7 Perbandingan *Accuracy*, AUC, dan *F*1-*score* Model dengan Seleksi Variabel Untuk Size 2. 112
Gambar 4.8 Perbandingan *Accuracy*, AUC, dan *F*1-*score* Model dengan Seleksi Variabel Untuk Size 3. 113
Gambar 4.9 Perbandingan *Sensitivity* Model XGB dengan dan Tanpa *Synthetic Features Generation* Secara Serentak dan Seleksi Variabel Menggunakan Data *Training* dan Data *Testing*.....118

Gambar 4.10 Perbandingan *Sensitivity* Model GEVR Secara Serentak dan Seleksi Variabel Menggunakan Data *Training* dan Data *Testing*. 119
Gambar 4.11 Perbandingan *Sensitivity* Model LOGIT Secara Serentak dan Seleksi Variabel menggunakan Data *Training* dan Data *Testing*. 120
Gambar 4.12 Perbandingan *Sensitivity* Model SVM Secara Serentak dan Seleksi Variabel Menggunakan Data *Training* dan Data *Testing*. 121

DAFTAR LAMPIRAN

Lampiran 1. Variabel Baru Hasil *Synthetic Features Generation*. 131
Lampiran 2. Estimasi Parameter Model GEVR Secara Serentak. 132
Lampiran 3. Estimasi Parameter Model *Stepwise* GEVR.....138
Lampiran 4. Estimasi Parameter Model LOGIT Secara Serentak. 141
Lampiran 5. Estimasi Parameter Model *Stepwise* LOGIT.....147
Lampiran 6. Estimasi Parameter Model *Lasso* LOGIT. 151
Lampiran 7. Estimasi Parameter Model *Elastic-Net* LOGIT...153
Lampiran 8. Estimasi Parameter Model *Lasso* SVM.....154
Lampiran 9. Estimasi Parameter Model *Elastic-Net* SVM. 156
Lampiran 10. Nilai *Gain* Model XGB dengan *Synthetic Features Generation*. 157
Lampiran 11. Hasil Akurasi Model Penelitian (Savera, 2020) 160
Lampiran 12. *Syntax Generalized Extreme Value Regression* 161
Lampiran 13. *Syntax Logistic Regression*.....166
Lampiran 14. *Syntax Lasso Logistic Regression* 167
Lampiran 15. *Syntax Elastic-Net Logistic Regression*.....168
Lampiran 16. *Syntax Support Vector Machine*.....173
Lampiran 17. *Syntax Lasso Support Vector Machine*.....174
Lampiran 18. *Syntax Elastic-Net Support Vector Machine*.....176
Lampiran 19. *Syntax Extreme Gradient Boosting* 178
Lampiran 20. Surat Keterangan Data 181

(*Halaman ini sengaja dikosongkan*)

BAB I
PENDAHULUAN

1.1 Latar Belakang

Perusahaan dalam prosesnya terkadang tidak selalu berjalan sesuai dengan rencana yang sudah ditetapkan. Pada beberapa situasi dan kondisi tertentu, perusahaan mungkin akan menghadapi risiko seperti kesulitan keuangan, baik kesulitan dalam tingkat yang ringan hingga kesulitan yang mengakibatkan perusahaan harus dilikuidasi. Tingkat kesehatan keuangan suatu perusahaan dapat dilihat dari laporan keuangan, seperti neraca. Neraca memuat informasi yang relevan untuk mengidentifikasi aliran dana masuk (*cash inflow*) dan aliran dana keluar (*cash outflow*) bersih yang dihasilkan oleh aset perusahaan. Selain itu, informasi dapat dilihat melalui laporan laba rugi suatu perusahaan yang memuat laporan sejumlah uang masuk dan uang keluar pada kurun waktu tertentu yang dicatat untuk menentukan kondisi perusahaan berada dalam kondisi surplus atau bahkan defisit. Berbekal informasi mengenai laporan arus kas, perusahaan dapat melakukan investigasi mengenai prestasi dan pencapaian perusahaan selama beberapa waktu ke belakang. Perusahaan yang sedang dirundung masalah tentu ditandai dengan berbagai permasalahan, mulai dari tanda-tanda kecil seperti keterlambatan membayar hutang hingga tanda yang paling mencolok yaitu terjadinya pemutusan hubungan kerja (PHK). Kondisi kesulitan keuangan pada perusahaan tersebut dapat dikatakan sebagai *financial distress.*

Financial distress didefinisikan sebagai tahap akhir dari penurunan kondisi perusahaan yang mendahului banyak peristiwa bencana seperti kebangkrutan atau likuidasi (Platt & Platt, 2002). Masalah *financial distress* jika tidak segera ditanggulangi akan berakhir dengan kebangkrutan. Kebangkrutan adalah kondisi disaat perusahaan mengalami ketidakcukupan dana untuk menjalankan usahanya (Kamaludin, Susena, & Usman, 2015). Menurut Undang-Undang Kepailitan No.4 Tahun 1998, debitur yang mempunyai dua atau lebih kreditur dan tidak membayar

sedikitnya satu utang yang telah jatuh tempo dan tidak dapat ditagih, dinyatakan pailit dengan putusan pengadilan yang berwenang, baik atas permohonannya sendiri, maupun atas permintaan seorang atau lebih kreditur. Kesulitan keuangan yang dihadapi oleh perusahaan mengakibatkan manajemen harus berfikir ekstra untuk mengambil tindakan yang dapat menyehatkan keuangan perusahaan. Kesulitan keuangan biasanya dimulai ketika arus kas (*cash flow*) tidak mencukupi lagi untuk mendanai hutang pada saat ini. Beberapa indikasi lain yang muncul dengan ditandai oleh tingginya *loan default*, yaitu peristiwa yang terjadi pada saat perusahaan gagal membayar bunga dan pokok pinjaman (Kamaludin, Manajemen Keuangan-Teori & Penerapan, Edisi Revisi, 2012). Menurut Ross *et al* (1993) dalam Afriyeni (2013), ketidakmampuan perusahaan dalam memenuhi kewajibannya dapat ditunjukkan dengan dua metode, yaitu *stock-based insolvency* dan *flow-based insolvency*. *Stock-based insolvency* adalah kondisi yang menunjukkan suatu kondisi ekuitas negatif dari neraca perusahaan (*negative net worth*), sedangkan *flow-based insolvency* ditunjukkan oleh kondisi arus kas operasi (*operating cash flow*) yang tidak dapat memenuhi kewajiban-kewajiban lancar perusahaan.

Salah satu faktor yang dapat mengakibatkan kondisi *financial distress* ialah dampak krisis keuangan global yang terjadi tahun 2008. Krisis keuangan global di tahun 2008 berawal dari bangkrutnya bank investasi terbesar keempat di Amerika Serikat pada saat itu yaitu Lehman Brothers. Hal tersebut terjadi akibat tidak seimbangnya sektor keuangan dengan sektor produksi karena adanya praktik monopoli sumber daya ekonomi oleh korporasi besar dan negara maju terhadap negara miskin. Dampak yang ditimbulkan oleh krisis keuangan global terhadap perekonomian Indonesia ialah pada triwulan IV tahun 2008 menurun sebesar 3,6% dibandingkan triwulan III tahun 2008 dan mengalami peningkatan 5,2% dibandingkan triwulan IV tahun 2007 yang menunjukkan pertumbuhan ekonomi yang lambat dibandingkan pertumbuhan ekonomi triwulan sebelumnya di tahun 2008 yaitu

6,2% pada triwulan I, 6,4% pada triwulan II, dan 6,4% pada triwulan III. Turunnya pertumbuhan ekonomi Indonesia pada triwulan IV tahun 2008 disebabkan oleh menurunnya pertumbuhan ekspor barang dan jasa sebesar 5,5% dibandingkan triwulan III dan hanya meningkat 1,8% dibandingkan triwulan IV tahun 2007. Menurunnya pertumbuhan ekspor barang dan jasa di Indonesia diakibatkan menurunnya harga minyak dan permintaan komoditas ekspor Indonesia sebagai dampak dari krisis keuangan global. Dalam dinamika pasar modal, tentunya diperlukan sikap kritis dari investor dan pelaku usaha untuk meminimalisir dampak dari turunnya pertumbuhan ekonomi di Indonesia. Untuk menghasilkan keputusan yang tepat, maka perlu dilakukan prediksi. Prediksi *financial distress* merupakan sebuah *early warning system* kepada pihak-pihak yang mendapatkan dampak dari kebangkrutan. Sehingga dapat segera melakukan tindakan pencegahan dan evaluasi terkait laporan keuangan perusahaan. Selain itu, tujuan dari prediksi *financial distress* adalah untuk menilai kondisi keuangan suatu perusahaan dan perspektif masa depannya dalam konteks operasi jangka panjang (Zhang, Wang, & Ji, 2013).

Terdapat beberapa penelitian yang telah dilakukan sebelumnya dalam memprediksi *financial distress* pada suatu perusahaan di Indonesia. Pada penelitian sebelumnya, Savera (2020) meneliti mengenai prediksi *financial distress* pada perusahaan sektor industri di Indonesia menggunakan metode *Support Vector Machine* (SVM), *Generalized Extreme Value Regression* (GEVR), dan *Logistic Regression* dengan seleksi variabel. Hasil kesimpulannya ialah prediksi klasifikasi *financial distress* baik secara serentak maupun seleksi variabel dengan metode GEVR memberikan ketepatan klasifikasi terbaik. Penelitian lain juga dilakukan oleh Pasaribu (2008) yang meneliti mengenai prediksi *financial distress* di Bursa Efek Indonesia dengan studi kasus perusahaan publik industri perdagangan. Penelitian tersebut mendapatkan hasil kesimpulan bahwa model regresi logistik biner dengan indikator rasio lancar dan indikator perputaran aset memiliki tingkat akurasi terbaik. Selain itu, Almilia

(2006) dalam penelitiannya mengenai prediksi kondisi *financial distress* perusahaan *Go-Public* dengan menggunakan analisis multinomial logit menghasilkan kesimpulan bahwa rasio keuangan dari laporan laba rugi, neraca, dan laporan arus kas (CATA, TLTA, NFATA, CFFOCL, CFFOTS, dan CFFOTL) adalah variabel signifikan yang menentukan kondisi keuangan perusahaan yang tertekan. Kemudian terdapat penelitian mengenai prediksi kebangkrutan perusahaan sektor manufaktur di Polandia menggunakan metode *ensemble boosted trees* dengan *synthetic features generation* yang ditulis oleh Zieba *et al* (2016) dan didapatkan kesimpulan bahwa dengan metode tersebut menghasilkan prediksi yang lebih baik dibandingkan hasil yang diperoleh dari semua metode referensi yang diterapkan untuk masalah prediksi kondisi keuangan perusahaan sebelumnya. *Synthetic features generation* merupakan suatu metode yang mengkombinasikan ukuran ekonometrik menggunakan operasi aritmatika seperti penjumlahan, pengurangan, perkalian, dan pembagian (Zieba, Tomczak, & Tomczak, 2016).

Berdasarkan penjelasan yang telah disebutkan sebelumnya, maka pada Tugas Akhir ini akan dilakukan prediksi *financial distress* perusahaan sektor industri di Indonesia dengan metode klasifikasi (*Generalized Extreme Value Regression* (GEVR), *Logistic Regression*, *Support Vector Machine* (SVM), dan *Extreme Gradient Boosting* (XGB)) dan melibatkan *synthetic features generation* secara serentak dan seleksi variabel. *Generalized Extreme Value Regression* (GEVR) dapat menjadi solusi apabila variabel dependen mewakili peristiwa langka, karena pada model regresi logistik menunjukkan kelemahan yang cukup relevan (Calabrese & Osmetti, 2011). Pada penelitian ini ingin dilihat seberapa besar pengaruh metode *Extreme Gradient Boosting* (XGB) dan *synthetic features generation* dalam meningkatkan hasil prediksi *financial distress* perusahaan sektor industri di Indonesia dari hasil penelitian Savera (2020), dikarenakan metode tersebut menghasilkan prediksi terbaik pada penelitian Zieba *et al* (2016) yang memprediksi *financial distress* perusahaan sektor

manufaktur di Polandia. Kemudian pada penelitian ini juga dilakukan seleksi variabel dengan prosedur *stepwise*, serta menggunakan metode *Lasso* dan *Elastic-Net* pada *Logistic Regression* dan *Support Vector Machine* (SVM). Metode *Lasso* secara otomatis memilih variabel yang relevan dan mengecualikan variabel yang tidak relevan dengan mengecilkan koefisiennya menjadi nol (Hardle & Prastyo, 2014) dan metode *Elastic-Net* digunakan untuk mengatasi kelemahan *Lasso* (Zou & Hastie, 2005). Variabel yang digunakan pada penelitian ini merupakan rasio keuangan yang diperoleh dari penelitian Savera (2020) dengan data penelitian yang dikelompokkan menggunakan konsep *drift*. Perusahaan yang mengalami *financial distress* dikategorikan 1 dan perusahaan yang tidak mengalami *financial distress* (sehat) dikategorikan 0. Pada penelitian ini perusahaan yang mengalami kondisi *financial distress* dilihat berdasarkan nilai ICR yang kurang dari 1 atau nilai ROA negatif. Pemilihan metode terbaik dalam memprediksi *financial distress* perusahaan sektor industri di Indonesia dilakukan dengan melihat nilai *accuracy*, AUC, dan *F*1-score terbaik dari hasil evaluasi klasifikasi pada masing-masing metode. Penelitian ini bertujuan sebagai *early warning system* kepada pihak-pihak yang dapat dirugikan dari dampak bangkrutnya perusahaan sektor industri di Indonesia, sehingga pihak-pihak tersebut dapat melakukan tindakan pencegahan dan evaluasi pada kondisi keuangan perusahaan.

1.2 Rumusan Masalah

Masalah *financial distress* dapat berakhir dengan kebangkrutan apabila jika tidak segera ditanggulangi. Untuk mengantisipasi dan meminimalisir dampak dari bangkrutnya suatu perusahaan, maka dilakukan prediksi *financial distress* untuk menilai kondisi keuangan suatu perusahaan dan perspektif masa depannya dalam konteks operasi jangka panjang. Hasil prediksi *financial distress* dapat menjadi sebuah *early warning system* kepada pihak-pihak yang mendapatkan dampak dari bangkrutnya suatu perusahaan, sehingga dapat segera melakukan tindakan pencegahan dan evaluasi terkait laporan keuangan perusahaan.

Untuk mendapatkan keputusan yang tepat dalam memprediksi kondisi *financial distress*, maka diperlukan penentuan model klasifikasi terbaik yang memiliki tingkat akurasi yang tinggi. Pada penelitian sebelumnya, Savera (2020) meneliti mengenai prediksi *financial distress* pada perusahaan sektor industri di Indonesia menggunakan metode *Support Vector Machine* (SVM), *Generalized Extreme Value Regression* (GEVR), dan *Logistic Regression* dengan seleksi variabel. Hasil kesimpulannya ialah prediksi klasifikasi *financial distress* baik secara serentak maupun seleksi variabel dengan metode GEVR memberikan ketepatan klasifikasi terbaik. Pada penelitian ini ingin dilihat seberapa besar pengaruh metode *Extreme Gradient Boosting* (XGB) dan *synthetic features generation* dalam meningkatkan hasil prediksi *financial distress* perusahaan sektor industri di Indonesia dari hasil penelitian Savera (2020), dikarenakan metode tersebut menghasilkan prediksi terbaik pada penelitian Zieba *et al* (2016) yang memprediksi *financial distress* perusahaan sektor manufaktur di Polandia. Sehingga pada penelitian ini akan dilakukan perbandingan beberapa model klasifikasi yaitu *Generalized Extreme Value Regression* (GEVR), *Logistic Regression, Support Vector Machine* (SVM), dan *Extreme Gradient Boosting* (XGB) dengan melibatkan *synthetic features generation* secara serentak dan seleksi variabel. Secara spesifik rumusan masalah yang dianalisis pada penelitian ini ialah sebagai berikut:

1. Bagaimana prediksi *financial distress* perusahaan sektor industri di Indonesia dengan metode klasifikasi (*Extreme Gradient Boosting* (XGB)) secara serentak dan seleksi variabel?
2. Bagaimana prediksi *financial distress* perusahaan sektor industri di Indonesia dengan metode klasifikasi (*Generalized Extreme Value Regression* (GEVR), *Logistic Regression, Support Vector Machine* (SVM), dan *Extreme Gradient Boosting* (XGB)) dan melibatkan *synthetic features generation* secara serentak dan seleksi variabel?

3. Bagaimana perbandingan performansi metode prediksi *financial distress* perusahaan sektor industri di Indonesia?

1.3 Tujuan

Berdasarkan rumusan masalah tersebut, maka tujuan dari pembuatan penelitian ini adalah sebagai berikut:

1. Memprediksi *financial distress* perusahaan sektor industri di Indonesia dengan metode klasifikasi (*Extreme Gradient Boosting* (XGB)) secara serentak dan seleksi variabel.
2. Memprediksi *financial distress* perusahaan sektor industri di Indonesia dengan metode klasifikasi (*Generalized Extreme Value Regression* (GEVR), *Logistic Regression*, *Support Vector Machine* (SVM), dan *Extreme Gradient Boosting* (XGB)) dan melibatkan *synthetic features generation* secara serentak dan seleksi variabel.
3. Menentukan metode terbaik dalam memprediksi *financial distress* perusahaan sektor industri di Indonesia.

1.4 Manfaat

Adapun manfaat yang ingin dicapai pada penelitian ini adalah sebagai berikut:

1. Menjadi *early warning system* kepada pihak-pihak yang dapat dirugikan dari dampak bangkrutnya perusahaan sektor industri di Indonesia.
2. Mengetahui penerapan metode klasifikasi (*Generalized Extreme Value Regression* (GEVR), *Logistic Regression*, *Support Vector Machine* (SVM), dan *Extreme Gradient Boosting* (XGB)) dan *synthetic features generation* dalam memprediksi *financial distress* perusahaan sektor industri di Indonesia.

1.5 Batasan Masalah

Batasan masalah yang digunakan pada penelitian ini diantaranya adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini merupakan data laporan keuangan perusahaan sektor industri di Indonesia dari tahun 2005 hingga 2018 pada kuartal empat.

2. Sangat banyak variabel yang mempengaruhi terjadinya *financial distress* pada perusahaan sektor industri di Indonesia, tetapi pada penelitian ini penulis menggunakan 18 variabel prediktor.
3. Model yang dibentuk dalam penelitian ini merupakan model prediksi *financial distress* yang tidak dipengaruhi oleh kondisi pandemi.

BAB II
TINJAUAN PUSTAKA

Pada penelitian ini digunakan beberapa metode analisis untuk mencapai tujuan yang telah dijelaskan pada bab sebelumnya. Metode klasifikasi yang digunakan pada penelitian ini ialah *Generalized Extreme Value Regression* (GEVR), *Logistic Regression*, *Support Vector Machine* (SVM), dan *Extreme Gradient Boosting* (XGB). Selain itu, metode yang mendukung seperti konsep *drift*, *synthetic features generation*, evaluasi hasil prediksi, serta teori mengenai variabel yang digunakan dalam penelitian ini dijelaskan dalam bab ini.

2.1 Imputasi *k-Nearest Neighbor*

Metode imputasi melibatkan penggantian nilai yang hilang dengan yang diperkirakan berdasarkan beberapa informasi yang tersedia dalam kumpulan data (Batista & Monard, 2002). Salah satu metode imputasi *missing value* ialah dengan menggunakan imputasi *k-Nearest Neighbor*. Jarak antara dua pengamatan adalah rata-rata terbobot dari kontribusi masing-masing, dimana bobot harus mewakili kepentingan variabel (Kowarik & Templ, 2016). Jarak antara pengamatan ke-*i* dan ke-*j* didefinisikan sebagai berikut.

$$d_{ij} = \frac{\sum_{k=1}^{p} w_k \delta_{i,j,k}}{\sum_{k=1}^{p} w_k}$$

Dimana w_k adalah bobot dan $\delta_{i,j,k}$ adalah kontribusi variabel ke-*k*. Untuk variabel kontinu, jarak absolut dibagi dengan kisaran total yang digunakan sebagai berikut.

$$\delta_{i,j,k} = \frac{\left| x_{i,k} - x_{j,k} \right|}{r_k}$$

Dimana $x_{i,k}$ adalah nilai dari variabel ke-*k* dari observasi ke-*i*, $x_{j,k}$ adalah nilai dari variabel ke-*k* dari observasi ke-*j* dan r_k

adalah kisaran/range dari variabel ke-*k*. Untuk variabel nominal dan biner jarak 0/1 yang digunakan sebagai berikut.

$$\delta_{i,j,k} = \begin{cases} 0 \text{ jika } x_{i,k} = x_{j,k} \\ 1 \text{ jika } x_{i,k} \neq x_{j,k} \end{cases}$$

2.2 *Synthetic Features Generation*

Synthetic features generation merupakan suatu metode yang mengkombinasikan ukuran ekonometrik menggunakan operasi aritmatika seperti penjumlahan, pengurangan, perkalian, dan pembagian (Zieba, Tomczak, & Tomczak, 2016). *Synthetic features* dihasilkan oleh pemilihan acak dari dua *feature* yang ada dan pemilihan acak operasi aritmatika. *Synthetic features generation* dihasilkan melalui dua *feature* yaitu f_1 dan f_2. Nilai *feature* baru ditunjukkan pada formula berikut.

$$f_{new} = f_1 \circ f_2$$

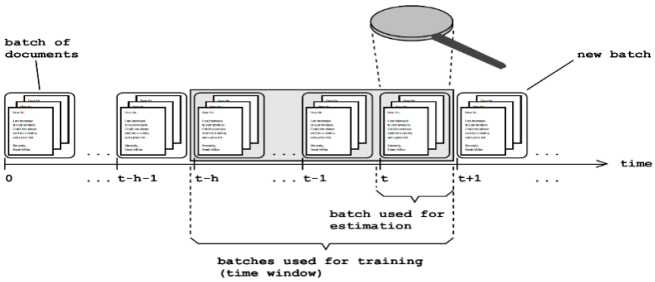
Operasi $\circ$ diambil sampelnya secara *uniform* dari himpunan $\{+, -, \times, /\}$. Nilai *feature* baru f_{new} dihitung untuk semua *feature* dalam dataset D.

2.3 Konsep *Drift*

Konsep *drift* merupakan pengelompokkan data dari waktu ke waktu tanpa kehilangan keumuman *batch* yang diasumsikan memiliki ukuran yang sama dengan masing-masing berisi *m* sampel (Scholz & Klinkenberg, 2007).

$$\underbrace{z_{(1,1)}, \ldots, z_{(1,m)}}_{batch\ 1}, \underbrace{z_{(2,1)}, \ldots, z_{(2,m)}}_{batch\ 2}, \ldots, \underbrace{z_{(t,1)}, \ldots, z_{(t,m)}}_{batch\ t}, \underbrace{z_{(t+1,1)}, \ldots, z_{(t+1,m)}}_{batch\ t+1}$$

$z_{(i,j)}$ menunjukkan *sampel* ke-*j* dari *batch i* dimana data pada setiap *batch* didistribusikan secara independen sehubungan dengan distribusi $P_i(x, y)$. Tujuan dari pengelompokkan data berdasarkan *time window size* dengan konsep *drift* ialah untuk memprediksi kejadian pada suatu waktu dengan mempertimbangkan kondisi pada waktu sebelumnya sesuai dengan *time window size.* Ilustrasi konsep *drift* ditampilkan pada Gambar 2.1.



Gambar 2.1 Ilustrasi Konsep *Drift* (Scholz & Klinkenberg, 2007).

Gambar 2.1 mengilustrasikan untuk contoh aplikasi pemfilteran informasi, apakah dokumen yang masuk dari aliran dokumen harus diklasifikasikan, misalnya dikategorikan menarik atau tidak menarik. Dokumen diasumsikan tiba dalam *batch* dan label klasifikasi yang benar untuk dokumen dalam *batch* 1 hingga t diketahui sistem. Disini *time window* terdiri dari *batch* $t-h$ ke t yang digunakan sebagai data *training*, sehingga kemudian diterapkan untuk mengklasifikasikan dokumen dalam *batch* yang baru tiba yaitu $t+1$.

2.4 *Generalized Extreme Value Regression* (GEVR)

Generalized Extreme Value Regression (GEVR) merupakan suatu model regresi dengan variabel dependen biner. *Generalized Extreme Value Regression* (GEVR) dapat menjadi solusi apabila variabel dependen mewakili peristiwa langka, karena pada model regresi logistik menunjukkan kelemahan yang cukup relevan (Calabrese & Osmetti, 2011). Fungsi distribusi *Generalized Extreme Value Regression* (GEVR) ditunjukkan pada formula (2.1).

$$F_X\left(x\right) = \exp\left\{-\left[1-\tau\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\tau}}\right\} \tag{2.1}$$

$$-\infty < \tau < \infty,\ \ -\infty < \mu < \infty,\ \ \sigma > 0$$

Parameter τ disebut sebagai parameter bentuk, sedangkan μ dan σ masing-masing adalah parameter lokasi dan skala. Untuk variabel respon biner Y dan variabel penjelas X misalkan nilai

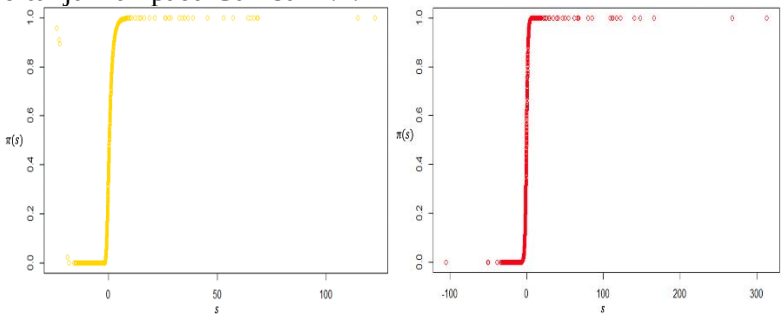
$\pi(x) = P\{Y = 1|X = x\}$, sehingga kurva respon didefinisikan oleh fungsi (2.2) berikut.

$$\pi\left(x\right) = \exp\left\{-\left[1+\tau\left(\vec{\beta}'\vec{x}\right)\right]^{-1/\tau}\right\} \tag{2.2}$$

dengan

$$\vec{\beta}' = \left[\beta_0,\ \beta_1,\ \ldots,\ \beta_k\right] \qquad \vec{x}' = \left[1,\ x_1,\ \ldots,\ x_k\right]$$

Dimana ketika $\tau \to 0$, fungsi (2.2) menjadi kurva respon dari model log log dan ketika nilai $\tau > 0$ maka menjadi kurva respon dari distribusi Weibull (Calabrese & Osmetti, 2011). Misalkan diberikan variabel prediktor X dengan nilai skor $s = \beta' X$ dan kurva respon $\pi(s)$, maka ilustrasi perbandingan model *Generalized Extreme Value Regression* (GEVR) dan model *Logistic Regression* ditunjukkan pada Gambar 2.2.



(a) (b)

Gambar 2.2 Perbandingan $\pi(s)$ GEVR (a) dan LOGIT (b).

Gambar 2.2 menunjukkan ilustrasi perbedaan nilai $\pi(s)$ untuk model *Generalized Extreme Value Regression* (GEVR) dan model *Logistic Regression*. Model *Generalized Extreme Value Regression* (GEVR) cenderung lebih fleksibel dibandingkan model *Logistic Regression*, karena nilai $\pi(s)$ pada model *Generalized Extreme Value Regression* (GEVR) yang ditunjukkan pada persamaan (2.2) dipengaruhi nilai τ. Estimasi koefisien untuk variabel independen menyatakan laju perubahan fungsi variabel dependen. Jika pada model *Generalized Extreme Value Regression* (GEVR) hanya terdapat satu variabel independen, maka fungsi *link* ditampilkan pada persamaan (2.3).

$$g\left(\pi\left(x\right)\right)=\frac{\left[-\ln\pi\left(x\right)\right]^{-\tau}-1}{\tau}=\beta_0+\beta_1 x \tag{2.3}$$

dengan

$$\beta_1=g\left(\pi\left(x+1\right)\right)-g\left(\pi\left(x\right)\right)$$

Dimana untuk semua nilai τ dan β_0 tetap, jika parameter β_1 positif maka peningkatan satu satuan pada variabel independen dikaitkan dengan pengurangan nilai $\pi(x)$. Sebaliknya jika nilai β_1 negatif, maka peningkatan satu satuan dalam variabel independen dikaitkan dengan peningkatan nilai $\pi(x)$ (Calabrese & Osmetti, 2011). Estimasi parameter dalam *Generalized Extreme Value Regression* (GEVR) dilakukan dengan metode *Maximum Likelihood* berikut.

$$l\left(\beta,\tau\right)=\prod_{i=1}^{n}f\left(x_i\right)=\prod_{i=1}^{n}\pi\left(x_i\right)^{y_i}\left(1-\pi\left(x_i\right)\right)^{1-y_i}$$

$$=\prod_{i=1}^{n}\exp\left\{-\left[1+\tau\left(\beta' x_i\right)\right]^{-\frac{1}{\tau}}\right\}^{y_i}\left(1-\exp\left\{-\left[1+\tau\left(\beta' x_i\right)\right]^{-\frac{1}{\tau}}\right\}\right)^{1-y_i}$$

Fungsi likelihood tersebut lebih mudah dimaksimumkan dalam bentuk $\ln l(\beta,\tau)$ dan dinyatakan dengan $L(\beta,\tau)$.

$$L\left(\beta,\tau\right)=\sum_{i=1}^{n}\left\{-y_i\left[1+\tau\left(\beta' x_i\right)\right]^{-\frac{1}{\tau}}+\left(1-y_i\right)\ln\left[1-\exp\left\{-\left[1+\tau\left(\beta' x_i\right)\right]^{-\frac{1}{\tau}}\right\}\right]\right\}$$

Nilai β maksimum didapatkan melalui turunan $L(\beta,\tau)$ terhadap β dan hasilnya adalah sama dengan nol.

$$\frac{\partial L\left(\beta,\tau\right)}{\partial\beta_j}=-\sum_{i=1}^{n}x_{ij}\frac{y_i\ln\left[\pi\left(x_i\right)\right]-\pi\left(x_i\right)}{1+\tau\beta' x_i\left(1-\pi\left(x_i\right)\right)}$$

Nilai τ maksimum didapatkan melalui turunan $L(\beta,\tau)$ terhadap τ dan hasilnya adalah sama dengan nol.

$$\frac{\partial L\left(\beta,\tau\right)}{\partial\tau}=\sum_{i=1}^{n}\left[\frac{1}{\tau^2}\ln\left(1+\tau\beta' x_i\right)-\frac{\beta' x_i}{\tau\left(1+\tau\beta' x_i\right)}\right]\frac{y_i-\pi\left(x_i\right)}{1-\pi\left(x_i\right)}\ln\left[\pi\left(x_i\right)\right]$$

Untuk mendapatkan nilai taksiran β dan τ dari turunan pertama fungsi $L(\beta, \tau)$ yang non linier ialah dengan menggunakan metode iterasi Newton Raphson.

2.5 *Logistic Regression*

Logistic Regression merupakan salah satu metode yang dapat digunakan untuk menentukan hubungan antara variabel respon yang dikotomus (skala nominal atau ordinal dengan dua kategori) atau polikotomus (skala nominal atau ordinal dengan lebih dari dua kategori) dengan satu atau lebih variabel prediktor (Agresti, 2013). Hasil dari variabel respon terdiri dari 2 kategori, yaitu "sukses" dan "gagal" yang disimbolkan dengan $y = 1$ (sukses) dan $y = 0$ (gagal). Dalam keadaan seperti itu, variabel respon mengikuti distribusi Bernoulli untuk setiap pengamatan (Hosmer & Lemeshow, 2000). *Logistic Regression* merupakan metode klasifikasi yang paling sederhana dan sebaiknya digunakan untuk data yang dapat dipisahkan secara linier. Fungsi probabilitas untuk setiap pengamatan diberikan pada persamaan (2.4).

$$f(y) = \pi^{y}(1-\pi)^{1-y} \tag{2.4}$$

Dimana jika $y = 0$ maka $f(y) = 1 - \pi$ dan jika $y = 1$ maka $f(y) = \pi$. Fungsi *Logistik Regression* dapat dituliskan pada persamaan (2.5).

$$\pi(x) = \frac{e^{(\beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p)}}{1 + e^{(\beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p)}} \tag{2.5}$$

Penentuan kategori berdasarkan nilai $\pi(x)$ dimana ketika $\pi(x)$ besar dari 0,5, maka observasi tersebut memiliki kategori 1. Kemudian dilakuan transformasi pada $\pi(x)$ untuk mendapatkan fungsi *link* dan transformasinya didefinisikan dalam istilah $\mathrm{g}(x)$ yang ditunjukkan pada persamaan (2.6).

$$\mathrm{g}(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p \tag{2.6}$$

Estimasi parameter dalam regresi logistik dilakukan dengan metode *Maximum Likelihood*. Setiap pasangan pengamatan

diasumsikan independen sehingga fungsi likelihoodnya merupakan gabungan dari fungsi distribusi masing-masing pasangan yaitu sebagai berikut.

$$l(\beta)=\prod_{i=1}^{n} f(x_i)=\prod_{i=1}^{n}\pi(x_i)^{y_i}\left(1-\pi(x_i)\right)^{1-y_i}$$

$$=\left\{\prod_{i=1}^{n}\left(1-\pi(x_i)\right)\right\}\left\{\prod_{i=1}^{n} e^{\left(\ln\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right)^{y_i}\right)}\right\}$$

$$=\left\{\prod_{i=1}^{n}\left(1-\pi(x_i)\right)\right\} e^{\left(\sum_{i=1}^{n}\ln\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right)^{y_i}\right)}$$

$$=\left\{\prod_{i=1}^{n}\frac{1}{1+e^{\sum_{j=0}^{p}\beta_j x_{ij}}}\right\} e^{\left(\sum_{i=1}^{n} y_i \ln\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right)\right)}$$

$$=\left\{\prod_{i=1}^{n}\left(1+e^{\sum_{j=0}^{p}\beta_j x_{ij}}\right)^{-1}\right\} e^{\left(\sum_{j=0}^{p}\left(\sum_{i=1}^{n} y_i x_{ij}\right)\beta_j\right)}$$

Fungsi likelihood tersebut lebih mudah dimaksimumkan dalam bentuk $\ln l(\beta)$ dan dinyatakan dengan $L(\beta)$.

$$L(\beta)=\sum_{j=0}^{p}\left(\sum_{i=1}^{n} x_{ij} y_i\right)\beta_j-\sum_{i=1}^{n}\ln\left(1+e^{\sum_{j=0}^{p}\beta_j x_{ij}}\right)$$

Nilai β maksimum didapatkan melalui turunan $L(\beta)$ terhadap β dan hasilnya adalah sama dengan nol.

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

sehingga

$$\sum_{i=1}^{n} x_{ij} y_i - \sum_{i=1}^{n} x_{ij} \hat{\pi}\left(x_i\right) = 0 \ ; \ j = 0,1,\ldots, p$$

Estimasi varians dan kovarians dikembangkan melalui teori MLE (*Maximum Likelihood Estimation*) dari koefisien parameternya (Hosmer & Lemeshow, 2000). Teori tersebut menyatakan bahwa estimasi varians kovarians didapatkan melalui turunan kedua $L(\beta)$.

$$\frac{\partial L\left(\beta\right)}{\partial \beta_j \beta_u} = \sum_{i=1}^{n} x_{ij} x_{iu} \pi\left(x_i\right)\left(1 - \pi\left(x_i\right)\right) \ ; \ j,u = 0,1,\ldots, p$$

Matriks varians kovarians berdasarkan estimasi parameter diperoleh melalui invers matriks dan diberikan sebagai berikut.

$$\operatorname{cov}\left(\beta\right) = \left\{x^T \operatorname{Diag}\left[\hat{\pi}\left(x_i\right)\left(1 - \hat{\pi}\left(x_i\right)\right)\right]\right\}^{-1}$$

dimana

$$x^T = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{n1} \\ \vdots & \vdots & \cdots & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{nk} \end{bmatrix}$$

Untuk mendapatkan nilai taksiran β dari turunan pertama fungsi $L(\beta)$ yang non linier maka digunakan metode iterasi Newton Raphson. Persamaan yang digunakan ditunjukkan pada persamaan (2.7).

$$\beta^{(t+1)} = \beta^{(t)} - \left(H^{(t)}\right)^{-1} q^{(t)} \ ; \ t = 1,2,\ldots, \text{konvergen} \qquad (2.7)$$

dimana

$$q^{(t)} = \left(\frac{\partial L\left(\beta\right)^{(t)}}{\partial \beta_0}, \frac{\partial L\left(\beta\right)^{(t)}}{\partial \beta_1}, \ldots, \frac{\partial L\left(\beta\right)^{(t)}}{\partial \beta_k}\right)$$

dan $H^{(t)}$ merupakan matriks Hessian dengan elemen-elemennya ialah $h_{ju} = \dfrac{\partial^2 L(\beta)^{(t)}}{\partial\beta_j \partial\beta_u}$

$$H^{(t)} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1k} \\ h_{21} & h_{22} & \cdots & h_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ h_{k1} & hk2 & \cdots & h_{kk} \end{bmatrix}$$

Langkah-langkah iterasi Newton Raphson ialah sebagai berikut.

1. Menentukan nilai dugaan awal $\beta^{(0)}$ kemudian dengan menggunakan persamaan (2.5) didapatkan $\pi(x_i)^{(0)}$.
2. Dari langkah 1 diperoleh matriks Hessian $H^{(0)}$ dan vektor $q^{(0)}$.
3. Mengulangi langkah diatas untuk $t > 0$ hingga $\pi(x_i)^{(t)}$ dan $\beta^{(t)}$ konvergen.

2.6 *Regularized Logistic Regression*

Probabilitas *default* dan non *default* masing-masing untuk perusahaan ke-*i* diberikan oleh prediktor *default* yang diformulakan sebagai berikut.

$$P(\gamma_i = 1 \mid x_i) = \frac{e^{\beta_0 + x_i^T \beta}}{1 + e^{\beta_0 + x_i^T \beta}}$$

$$= \frac{1}{1 + e^{-(\beta_0 + x_i^T \beta)}}$$

dan

$$P(\gamma_i = 0 \mid x_i) = 1 - P(\gamma_i = 1 \mid x_i)$$

$$= \frac{1}{1 + e^{\beta_0 + x_i^T \beta}}$$

Log *odds ratio* merupakan model regresi linier dengan persamaan sebagai berikut.

$$\log\left\{\frac{P\left(\gamma_i=1\mid x_i\right)}{P\left(\gamma_i=0\mid x_i\right)}\right\}=\beta_0+x_i^T\beta$$

Kemudian memaksimumkan fungsi *regularized* log-likelihood dengan persamaan sebagai berikut.

$$\max_{\beta_0,\beta}\left\{l\left(\beta_0,\beta\right)-\lambda R_\gamma\left(\beta\right)\right\}$$

Dimana $R_\gamma(\beta)$ merupakan fungsi regularisasi dari fungsi log-likelihood yang diberikan pada persamaan berikut.

$$l\left(\beta_0,\beta\right)=n^{-1}\sum_{i=1}^{n}\left[\gamma_i\log P\left(\gamma_i=1\mid x_i\right)+\left(1-\gamma_i\right)\log\left(1-P\left(\gamma_i=1\mid x_i\right)\right)\right]$$

$$=n^{-1}\sum_{i=1}^{n}\left\{I\left(\gamma_i=1\right)\log P\left(\gamma_i=1\mid x_i\right)\right\}+I\left(\gamma_i=0\right)\log P\left(\gamma_i=0\mid x_i\right)$$

$$=n^{-1}\sum_{i=1}^{n}I\left(\gamma_i=1\right)\left[\left(\beta_0+x_i^T\beta\right)-\log\left\{1+e^{\beta_0+x_i^T\beta}\right\}\right]$$

Persamaan tersebut dimaksimalkan menggunakan algoritma Newton Raphson yang menghasilkan skema *Iterative Reweighted Least Square* (IRLS). Hardle dan Prastyo (2014) mengikuti algoritma Friedman *et al.* (2010) yang menggunakan pendekatan serupa dengan IRLS yaitu menggunakan penurunan koordinat *cyclical* yang dihitung di sepanjang jalur regularisasi. Terdapat tiga langkah dalam algoritma penurunan koordinat *cyclical* yaitu *loop* luar, tengah, dan dalam. Di *loop* luar, nilai λ diatur dan menghasilkan $\left(\tilde{\beta}_0,\tilde{\beta}\right)$. Di *loop* tengah, estimasi kuadratik (Taylor *expansion*) dari fungsi log-likelihood diperbarui sehingga didapatkan persamaan (2.8).

$$l_Q\left(\beta_0,\beta\right)=-\frac{1}{2n}\sum_{i=1}^{n}\omega_i\left(z_i-\beta_0-x_i^T\beta\right)^2+C\left(\tilde{\beta}_0,\beta\right) \qquad (2.8)$$

dengan respon

$$z_i=\tilde{\beta}_0+x_i^T\tilde{\beta}+\frac{\gamma_i-\tilde{P}\left(\gamma_i=1\mid x_i\right)}{\tilde{P}\left(\gamma_i=1\mid x_i\right)\tilde{P}\left(\gamma_i=0\mid x_i\right)}$$

dan pembobot

$$\omega_i = \tilde{P}\left(\gamma_i = 1 \mid x_i\right)\tilde{P}\left(\gamma_i = 0 \mid x_i\right)$$

Dimana $\tilde{P}(\gamma_i = 1|x_i)$ dan $\tilde{P}(\gamma_i = 0|x_i)$ dievaluasi pada estimasi saat ini dan $C\left(\tilde{\beta}_0, \tilde{\beta}\right)$ merupakan konstanta. Istilah pertama dalam fungsi $l_Q(\beta_0, \beta)$ adalah *Weighted Least Square* (WLS). Di *loop* dalam, algoritma penurunan koordinat digunakan untuk menyelesaikan masalah *Penalized Weighted Least Square* (PWLS) berikut.

$$\min_{\beta_0, \beta}\left\{-l_Q\left(\beta_0, \beta\right) + \lambda R_\gamma\left(\beta\right)\right\} \tag{2.9}$$

Setiap *loop* penurunan koordinat dalam berlanjut hingga perubahan maksimum pada persamaan (2.9) kurang dari *treshold* yang sangat kecil. Langkah selanjutnya adalah menurunkan nilai λ dan mengulangi ketiga *loop* tersebut hingga estimasi konvergen. Setelah estimasi akhir diperoleh, representasi dengan prediktor relevan yang dipilih digunakan untuk memprediksi *default*.

2.6.1 *Lasso Logistic Regression*

Pendekatan regularisasi *Lasso* secara otomatis memilih variabel yang relevan dan mengecualikan variabel yang tidak relevan dengan mengecilkan koefisiennya menjadi nol (Hardle & Prastyo, 2014). Zou dan Hastie (2005) dalam Hardle dan Prastyo (2014) menyatakan dua kekuarangan metode *Lasso* yaitu jumlah prediktor yang dipilih dibatasi oleh jumlah ukuran sampel dan metode *Lasso* cenderung memilih hanya satu (atau beberapa) prediktor dari subset prediktor yang berkorelasi dan menyusut sisanya menjadi nol. Regularisasi untuk *Logistic Regression* ditunjukkan pada persamaan (2.10).

$$\min_{\beta_0, \beta}\sum_{i=1}^{n} L\left\{y_i, f\left(x_i\right) + \lambda\left(\gamma\left\|\beta\right\|_1\right)\right\} \tag{2.10}$$

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

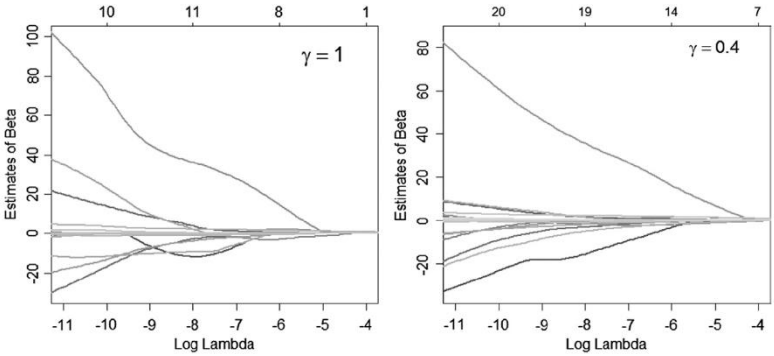
2.6.2 *Elastic-Net Logistic Regression*

Zou dan Hastie (2005) memperkenalkan penalti *Elastic-Net* untuk mengatasi kelemahan *Lasso*. Penalti *Elastic-Net* adalah

campuran dari penalti L_1-norm dan L_2-norm. Dua keuntungan dari penalti *Elastic-Net* ialah jumlah prediktor yang dipilih tidak dibatasi oleh ukuran sampel dan kelompok prediktor yang berkorelasi dapat dipilih bersama-sama. Fungsi persamaan *Elastic-Net Logistic Regression* ditunjukkan pada persamaan (2.11).

$$\min_{\beta_0,\beta} \sum_{i=1}^{n} L\left\{ y_i, f\left(x_i\right) + \lambda\left[\left(1-\gamma\right)\frac{1}{2}\left\|\beta\right\|_2^2 + \gamma\left\|\beta\right\|_1\right]\right\} \tag{2.11}$$

dengan bobot $0 < \gamma < 1$ yang harus dioptimalkan bersamaan dengan $\lambda > 0$ yang merupakan *tunning parameter*. Ilustrasi perbedaan penalti dengan *Lasso* dan *Elastic-Net* ditunjukkan pada Gambar berikut.



(a) (b)

Gambar 2.3 Perbandingan Penalti *Lasso* (a) dan *Elastic-Net* (b) (Hardle & Prastyo, 2014).

Gambar 2.3 menunjukkan perbedaan penalti variabel dengan metode *Lasso* dan *Elastic-Net*. Berdasarkan Gambar 2.3 tersebut dapat diketahui bahwa ada hubungan negatif antara nilai λ dan koefisien β. Semakin kecil nilai λ, maka nilai koefisien β cenderung semakin besar. Gambar 2.3 (b) menunjukkan hasil penalti dengan metode *Elastic-Net* cenderung mengakibatkan nilai koefisien β lebih kecil dan cenderung mendekati nol jika dibandingkan dengan penalti variabel menggunakan metode *Lasso* yang ditunjukkan pada Gambar 2.3 (a).

2.7 *Support Vector Machine* (SVM)

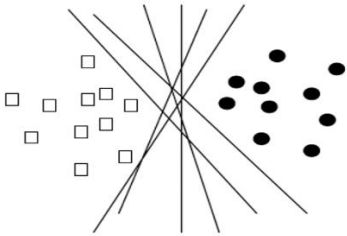
Support Vector Machine (SVM) adalah metode *machine learning* yang bekerja atas prinsip *Structural Risk Minimization* (SRM) dengan tujuan menemukan *hyperplane* terbaik yang memisahkan dua buah *class* pada *input space* (Afriansyah, 2017). *Support Vector Machine* (SVM) berfungsi dengan baik pada data yang tidak terstruktur dan semi terstruktur seperti teks dan gambar. Teknik *Support Vector Machine* (SVM) juga disebut sebagai sebuah metode klasifikasi yang menemukan *hyperplane* atau fungsi $g(x) = w^T x + b$ yang memisahkan dua kelas dengan margin maksimum (Awad & Khanna, 2015). Metode *Support Vector Machine* (SVM) terdiri dari dua jenis, yaitu untuk kasus yang dapat dipisahkan secara linier dan untuk kasus yang dipisahkan secara nonlinier dengan penjelasan sebagai berikut.

2.7.1 *Support Vector Machine* (SVM) dalam Klasifikasi Linier

Setiap observasi terdiri dari sepasang p prediktor $x_i^T = (x_i^1, \ldots, x_i^p) \in \mathbb{R}^p, i = 1, \ldots, n$ dan $y_i \in \mathcal{Y} = \{-1,1\}$ yang terkait dengan urutan D_n sebagai berikut.

$$D_n = \{(x_1, y_1), \ldots, (x_n, y_n)\} \in Q \times \{-1,1\}$$

dari pasangan i.i.d yang diambil dari distribusi probabilitas $F(x, y)$. Domain Q adalah beberapa himpunan set tidak kosong dimana x_i merupakan prediktor dan y_i adalah target atau label. Konsep kunci untuk mendefinisikan pengklasifikasi linier ialah hasil dot (titik) (Hardle, Prastyo, & Hafner, 2014). Visualisasi fungsi klasifikasi untuk data yang dapat dipisahkan secara linier ditampilkan pada Gambar 2.4.



Gambar 2.4 Fungsi Klasifikasi untuk Data yang Dapat Dipisahkan Secara Linier (Hardle, Prastyo, & Hafner, 2014).

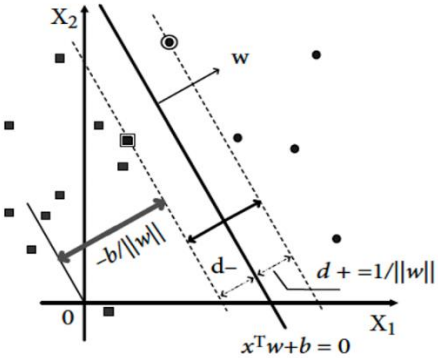
Ruang data yang ditampilkan Gambar 2.4 ditunjukkan pada persamaan berikut.

$$F = \left\{ x^T w + b, w \in \mathbb{R}^p, b \in \mathbb{R} \right\}$$

dimana *w* sebagai vektor bobot dan *b* adalah deviasi dari titik awal. Berikut persamaan untuk batas keputusan (pemisah *hyperplane*).

$$f(x) = x^T w + b = 0 \tag{2.12}$$

Persamaan (2.12) membagi ruang menjadi dua wilayah seperti pada Gambar 2.5. Himpunan titik *x* pada $f(x) = x^T w = 0$ adalah semua titik yang tegak lurus terhadap *w* dan melewati titik asal. Konstanta *b* menerjemahkan *hyperplane* dari asalnya.



Gambar 2.5 Pemisah *Hyperplane* $x^T w + b = 0$ dan Margin yang Dapat Dipisahkan Secara Linier (Hardle, Prastyo, & Hafner, 2014).

Panjang vektor *w* dilambangkan dengan $norm\|w\| = \sqrt{w^T w}$. Unit vektor *w*, dimana $\|w\| = 1$ dalam arah *w* diberikan oleh $w/\|w\|$. Selanjutnya margin dari *hyperplane* $f(x)$ sehubungan dengan kumpulan data D_n berikut.

$$d_D(f) = \frac{1}{2} w^T (x_+ - x_-)$$

dimana unit vektor *w* searah dengan *w*. Diasumsikan bahwa x_+ dan x_- berjarak sama dari *hyperplane* pemisah dengan persamaan sebagai berikut.

$$f\left(x_{+}\right)=w^{T}x_{+}+b=a$$

$$f\left(x_{-}\right)=w^{T}x_{-}+b=-a$$

dengan konstanta $a>0$. Misalkan ditetapkan $a=1$, maka $d_D(f)=1$. Untuk membuat margin geometrik bermakna, bagi $d_D(f)\|w\|$ dengan *norm* vektor *w* untuk mendapatkan persamaan (2.13).

$$\frac{d_D\left(f\right)}{\|w\|}=\frac{1}{\|w\|} \tag{2.13}$$

Misalkan $x^T w+b=0$ adalah *hyperplane* pemisah dan misalkan $y_i \in \{-1,+1\}$ mengkodekan respon biner untuk observasi ke-*i*, maka (d_+) dan (d_-) akan menjadi jarak terpendek antara *hyperplane* pemisah dan objek terdekat dari kelas $(+1)$ dan (-1). Karena pemisahan dapat dilakukan tanpa kesalahan (*error*), semua pengamatan $i=1,2,\ldots,n$ harus memenuhi batasan berikut.

$$x_i^T w+b\geq +1 \quad \text{untuk } y_i=+1$$

$$x_i^T w+b\leq -1 \quad \text{untuk } y_i=-1$$

Kemudian dapat digabungkan kedua batasan tersebut menjadi satu sebagai berikut.

$$y_i\left(x_i^T w+b\right)-1\geq 0, \quad i=1,2,\ldots,n \tag{2.14}$$

Oleh karena itu fungsi tujuan dari kasus yang dapat dipisahkan secara linier akan menjadi margin yang memaksimalkan persamaan (2.13) atau ekuivalen dengan persamaan (2.15).

$$\min_{w}\frac{1}{2}\|w\|^2 \tag{2.15}$$

Berdasarkan persamaan (2.14), *Lagrangian* untuk masalah tunggal dalam kasus ini ditunjukkan pada persamaan (2.16).

$$\min_{w,b} L_p\left(w,b\right)=\frac{1}{2}\|w\|^2-\sum_{i=1}^{n}a_i\left(y_i\left(x_i^T w+b\right)-1\right) \tag{2.16}$$

Sedangkan *Lagrangian* untuk masalah ganda ditunjukkan pada persamaan (2.17).

$$\max_{a} L_D(a) = \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j y_i y_j x_i^T x_j \tag{2.17}$$

Setelah menyelesaikan masalah ganda, kemudian dapat mengklasifikasikan objek dengan menggunakan aturan klasifikasi berikut.

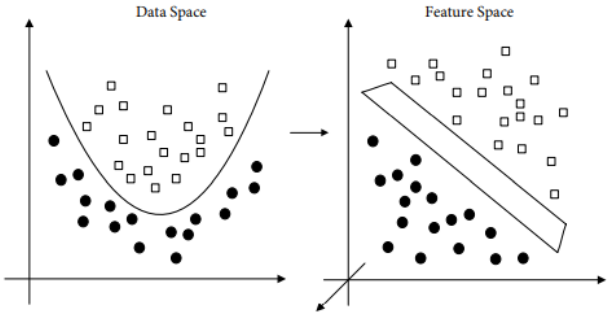
$$\hat{g}(x) = \text{sign}\left(x^T \hat{w} + \hat{b}\right)$$

dengan

$$\hat{w} = \sum_{i=1}^{n} \hat{a}_i y_i x_i$$

2.7.2 *Support Vector Machine* (SVM) dalam Klasifikasi Nonlinier

Metode *Support Vector Machine* (SVM) dalam klasifikasi nonliner dilakukan apabila data penelitian dipisahkan secara nonlinier dengan ilustrasi ditunjukkan pada Gambar 2.6.



Gambar 2.6 Ilustrasi *Support Vector Machine* (SVM) dalam Klasifikasi Nonliner (Hardle, Prastyo, & Hafner, 2014).

Klasifikasi nonlinier seperti pada Gambar 2.6 memetakan data dengan struktur nonliner melalui fungsi $\psi : \mathbb{R}^p \rightarrow$ H ke dalam ruang H berdimensi tinggi dimana struktur klasifikasi hampir linier. Untuk mendapatkan fungsi klasifikasi nonlinier dalam ruang data (*data space*), bentuk umum diperoleh dengan menerapkan

kernel trick ke persamaan (2.17) yang ditunjukkan pada persamaan (2.18) (Hardle, Prastyo, & Hafner, 2014).

$$\max_{a} L_D\left(a\right) = \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j y_i y_j K\left(x_i, x_j\right) \tag{2.18}$$

Salah satu *kernel* paling populer yang digunakan dalam *Support Vector Machine* (SVM) adalah *Radial Basis Function* (RBF) dengan persamaan (2.19) berikut.

$$K\left(x_i, x_j\right) = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}\right) \tag{2.19}$$

2.8 *Regularized Support Vector Machine*

Prastyo (2015) dalam Savera (2020) menyatakan bahwa standar L_2-*norm* SVM dapat ditulis dalam bentuk *regularization* pada persamaan (2.20).

$$\min_{w,b} \sum_{i=1}^{n} \left\{1 - y_i f\left(x_i\right)\right\}_+ + \lambda \left\|w\right\|_2^2 \tag{2.20}$$

dengan $\lambda > 0$ merupakan *tunning parameter* dan *Hinge-loss function* ditunjukkan pada persamaan (2.21).

$$L\left\{y_i, f\left(x_i\right)\right\} = \left\{1 - y_i f\left(x_i\right)\right\}_+ + \max\left\{0, 1 - y_i f\left(x_i\right)\right\} \tag{2.21}$$

Fungsi penalti pada persamaan (2.17) yaitu $P(w) = \|w\|_2^2$ menggunakan seluruh input untuk mendapatkan *hyperplane* pemisah, sehingga tidak dapat memilih atau menentukan prediktor yang relevan (Savera, 2020).

2.8.1 *Lasso Support Vector Machine*

Lasso Support Vector Machine merupakan metode dalam mengeleminasi variabel yang tidak relevan dalam model SVM sehingga mendapatkan model dengan variabel yang relevan. Menggunakan metode *Lasso Support Vector Machine*, koefisien pada variabel yang tidak relevan akan mendekati nol. Fungsi persamaan *Lasso Support Vector Machine* ditunjukkan pada persamaan (2.22).

$$\min_{w,b} \sum_{i=1}^{n} \left\{1 - y_i f\left(x_i\right)\right\}_+ + \lambda \left\|w\right\|_1 \tag{2.22}$$

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

2.8.2 *Elastic-Net Support Vector Machine*

Metode *Elastic-Net Support Vector Machine* digunakan untuk mengatasi kelemahan *Lasso Support Vector Machine*. Metode *Elastic-net Support Vector Machine* merupakan campuran dari penalti L_1-norm dan L_2-norm, dimana penalti L_1-norm berperan untuk memilih variabel prediktor dan penalti L_2-norm berperan untuk memilih variabel respon. Fungsi persamaan *Elastic-Net Support Vector Machine* ditunjukkan pada persamaan (2.23).

$$\min_{w,b} \sum_{i=1}^{n} \left\{1 - y_i f\left(x_i\right)\right\}_+ + \lambda_1 \left\|w\right\|_1 + \lambda_2 \left\|w\right\|_2^2 \tag{2.23}$$

dengan λ_1 dan λ_2 adalah *tunning parameter*.

2.9 *Extreme Gradient Boosting* (XGB)

Gradient Boosting merupakan algoritma regresi dan klasifikasi yang menerapkan konsep ensembel dari prediktor lemah dan umumnya menggunakan pohon keputusan (Syahrani, 2019). Menurut Chen dan Guestrin (2016), *Extreme Gradient Boosting* (XGB) adalah ansambel dari *K Classification and Regression Trees* (CART) dimana x_i merupakan data *training* yang diberikan untuk memprediksi label kelas y_i. Metode klasifikasi dengan *Extreme Gradient Boosting* (XGB) cukup baik dalam mengklasifikasikan data yang cukup besar serta mengklasifikasikan data yang memiliki *feature* campuran atau hanya *feature* numerik. Mengingat bahwa CART memberikan *real score* untuk setiap *leaf* (hasil atau target), skor prediksi untuk CART individu dijumlahkan untuk mendapatkan skor akhir dan dievaluasi melalui *K additive functions* (Mustapha & Saeed, 2016), seperti yang ditunjukkan pada persamaan (2.24).

$$\hat{y}_i = \sum_{k=1}^{K} f_k\left(x_i\right) \quad , \quad f_k \in F \tag{2.24}$$

Dimana f_k mewakili *independent tree structure* dengan *leaf scores* dan *F* adalah daerah dari semua CART. *Regularized objective* untuk mengoptimalkan fungsi tersebut diberikan pada persamaan (2.25).

$$obj\left(\Theta\right) = \sum_{i}^{n} l\left(y_i, \hat{y}_i\right) + \sum_{k}^{K} \Omega\left(f_k\right) \tag{2.25}$$

Loss function l menyatakan ukuran perbedaan antara prediksi $\hat{y}_i$ dan target y_i. Kemudian regularisasi Ω yang mengatur kompleksitas model untuk menghindari *overfitting*, dimana $\Omega(f) = \gamma T + (1/2)\lambda \sum_{j-1}^{T} w_j^2$ dengan *T* dan *w* adalah jumlah *leaf* dan skor pada setiap leaf, sedangkan γ dan λ adalah konstanta untuk mengontrol derajat *regularization*. Selain penggunaan *regularization*, *shrinkage* dan *descriptor subsampling* adalah dua teknik tambahan yang digunakan untuk mencegah *overfitting*. Prosedur *Extreme Gradient Boosting* (XGB) pada data *training* diringkas sebagai berikut (Mustapha & Saeed, 2016).

1. Pada setiap deskriptor, urutkan datanya dan *scan* titik pemisah terbaik.
2. Pilih deskriptor dengan titik pemisah terbaik.
3. Lanjutkan pemisahan (seperti pada (1) dan (2)) hingga *tree* maksimum yang ditentukan tercapai.
4. Tetapkan skor prediksi ke *leaf* dan eliminasi semua *node* negatif dalam urutan bawah ke atas.
5. Ulangi langkah diatas dengan cara *additive* hingga *K trees* tercapai.

Prediksi $\hat{y}_i$ pada langkah ke-*t* dinyatakan pada persamaan (2.26) berikut.

$$\hat{y}_i^{(t)} = \sum_{k=1}^{K} f_k\left(x_i\right) = \hat{y}_i^{(t-1)} + f_t\left(x_i\right) \tag{2.26}$$

Berdasarkan persamaan (2.26), maka persamaan (2.25) dapat dituliskan seperti persamaan (2.27) berikut.

$$obj\left(\Theta\right)^{(t)} = \sum_{i}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t\left(x_i\right)\right) + \Omega\left(f_t\right) \tag{2.27}$$

Secara umum dengan *Taylors expansion* untuk orde kedua didapatkan persamaan (2.28) berikut.

$$obj\left(\Theta\right)^{(t)} = \sum_{i}^{n} l\left(y_i, \hat{y}_i^{(t-1)}\right) + g_i f_t\left(x_i\right) + \frac{1}{2} h_i f_t^2\left(x_i\right) + \Omega\left(f_t\right) \tag{2.28}$$

dimana

$$g_i = \partial_{\hat{y}_i^{(t-1)}} l\left(y_i, \hat{y}_i^{(t-1)}\right) \quad \text{dan} \quad h_i = \partial^2_{\hat{y}_i^{(t-1)}} l\left(y_i, \hat{y}_i^{(t-1)}\right)$$

Fungsi objektif yang disederhanakan tanpa konstanta pada langkah ke-*t* adalah sebagai berikut.

$$obj\left(\Theta\right)^{(t)} = \sum_{i=1}^{n}\left[g_i f_t\left(x_i\right) + \frac{1}{2} h_i f_t^2\left(x_i\right)\right] + \Omega\left(f_t\right)$$

Fungsi objektif dapat ditulis dengan menjabarkan *regularization term* sebagai berikut.

$$\begin{aligned} obj\left(\Theta\right)^{(t)} &= \sum_{i=1}^{n}\left[g_i f_t\left(x_i\right) + \frac{1}{2} h_i f_t^2\left(x_i\right)\right] + \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \\ &= \sum_{j=1}^{T}\left[\left(\sum_{i \in I_j} g_i\right) w_j + \frac{1}{2}\left(\sum_{i \in I_j} h_i + \lambda\right) w_j^2\right] + \gamma T \end{aligned}$$

Dimana $I_j = \{i | q(x_i) = j\}$ adalah himpunan *j leaf* dan *leaf weight* optimal w_j^*, misalkan diberikan struktur $q(x)$, maka fungsi objektif optimal yang mengukur seberapa baik struktur tersebut ditunjukkan oleh persamaan (2.29) dan (2.30).

$$w_j^* = -\frac{G_j}{H_j + \lambda} \tag{2.29}$$

$$obj^* = -\frac{1}{2}\sum_{j=1}^{T}\frac{G_j^2}{H_j + \lambda} + \gamma T \tag{2.30}$$

dimana

$$G_j = \sum_{i \in I_j} g_i \quad \text{dan} \quad H_j = \sum_{i \in I_j} h_i$$

Persamaan (2.31) digunakan untuk menilai *node leaf* selama pemisahan. *Term* pertama, kedua, dan ketiga dari persamaan tersebut mewakili skor di kiri, kanan, dan *original leaf*. Selain itu, *term* terakhir γ merupakan *regularization* pada *additional leaf* (Mustapha & Saeed, 2016).

$$Gain = \frac{1}{2}\left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{\left(G_L + G_R\right)^2}{H_L + H_R + \lambda}\right] - \gamma \qquad (2.31)$$

Seleksi variabel pada *Extreme Gradient Boosting* (XGB) dilakukan dengan melihat nilai *gain* yang didapatkan dari persamaan (2.31). Variabel yang memiliki nilai *gain* sangat kecil (mendekati nol) akan dianggap sebagai variabel yang tidak berpengaruh signifikan dalam model.

2.10 Prosedur *Stepwise*

Goodman (1971) dalam Agresti (2013) mengusulkan metode *forward selection* dan *backward elimination* dalam suatu regresi. *Forward elimination* menambahkan *term* secara berurutan. Setiap tahap akan memilih *term* yang memberikan peningkatan *fit* terbesar. *P-value* minimum digunakan untuk menguji *term* dalam model, karena pada *term* yang berbeda mungkin memiliki nilai *df* yang berbeda. Proses berhenti ketika penambahan lebih lanjut tidak meningkatkan *fit*. Ulangi prosedur ini dan pada setiap tahap, *term* ditambahkan pada tahap sebelumnya untuk melihat apakah *term* tersebut masih signifikan.

Backward elimination dimulai dengan model yang kompleks dan secara berurutan menghapus *term*. Pada setiap tahap, akan dipilih *term* yang penghapusannya memiliki efek seperti *P-value* paling besar. Proses ini berhenti jika penghapusan lebih lanjut menyebabkan *fit* menjadi lebih buruk. *Backward elimination* lebih direkomendasikan daripada *forward elimination*, karena dirasa lebih aman untuk menghapus *term* dari model yang terlalu rumit

daripada menambahkan *term* ke model yang sederhana (Agresti, 2013).

2.11 Uji Signifikansi Parameter

Setelah parameter hasil estimasi diperoleh, maka kemudian dilakukan pengujian signifikansi parameter secara serentak dengan hipotesis sebagai berikut.

H_0 : $\beta_1 = \beta_2 = \ldots = \beta_p = 0$

H_1 : Paling tidak terdapat satu $\beta_j \neq 0$, $j = 1,2,\ldots,p$

Statistik uji untuk menguji signifikansi parameter secara serentak ditunjukkan pada persamaan (2.32).

$$G = -2\ln \frac{\left(\frac{n_1}{n}\right)^{n_1}\left(\frac{n_0}{n}\right)^{n_0}}{\sum_{i=1}^{n} \hat{\pi}_i^{\,y_i}\left(1-\hat{\pi}_i\right)^{(1-y_i)}} \tag{2.32}$$

dimana

$$n_1 = \sum_{i=1}^{n} y_i \qquad n_0 = \sum_{i=1}^{n}\left(1-y_i\right) \qquad n = n_1 + n_0$$

Statistik uji G merupakan *Likelihood Ratio Test* dimana nilai G mengikuti distribusi *Chi-Squared*, sehingga H_0 ditolak jika $G > \chi^2_{(\alpha/2,v)}$ dengan v derajat bebas adalah banyaknya parameter dalam model tanpa β_0.

Ketika dari uji serentak tidak mengeluarkan hasil yang signifikan maka kemudian dilakukan pengujian signifikansi terhadap koefisien β secara parsial terhadap variabel respon yaitu dengan membandingkan parameter hasil maksimum likelihood, dugaan β dengan *standard error* parameter tersebut. Hipotesis pengujian parsial ialah sebagai berikut.

H_0 : $\beta_j = 0$

H_1 : $\beta_j \neq 0$, $j = 1,2,\ldots,p$

Statistik uji untuk melakukan pengujian signifikansi parameter secara parsial ditunjukkan pada persamaan (2.33).

$$W = \frac{\hat{\beta}_j}{SE\left(\hat{\beta}_j\right)} \tag{2.33}$$

Statistik uji W tersebut mengikuti distribusi Normal sehingga H_0 ditolak jika $W > Z_{\alpha/2}$ atau *p-value* $< \alpha$ (Hosmer & Lemeshow, 2000).

2.12 Evaluasi Hasil Klasifikasi

Menurut Johnson dan Wichern (2007), evaluasi prosedur klasifikasi merupakan sebuah evaluasi yang melihat kemungkinan kesalahan klasifikasi yang dilakukan oleh suatu model klasifikasi. Dalam mengukur ketepatan klasifikasi, perlu diketahui jumlah data pada setiap kelas prediksi dan kelas aktual yang terdiri dari TP (*True Positive*), TN (*True Negative*), FP (*False Positive*), dan FN (*False Negative*). Tabel 2.1 merupakan *confusion matrix* yang memuat keempat nilai tersebut.

Tabel 2.1 *Confusion Matrix*.

Kelas Aktual	Kelas Prediksi	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

Pengukuran yang sering digunakan untuk menghitung ketepatan klasifikasi adalah akurasi, *specificity*, dan *sensitivity (Hotho, Nurnberger, & Paass, 2005)*. Berikut merupakan formula dalam menghitung *accuracy*, *specificity*, dan *sensitivity* yang ditampilkan pada persamaan (2.34), (2.35), (2.36).

$$Accuracy = \frac{\mathrm{TN} + \mathrm{TP}}{\mathrm{TN} + \mathrm{TP} + \mathrm{FN} + \mathrm{FP}} \tag{2.34}$$

$$Sensitivity = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2.35}$$

$$Specificity = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{2.36}$$

Sedangkan untuk data yang *imbalanced*, pengukuran ketepatan klasifikasi yang digunakan adalah *Area Under the Curve*

(AUC). AUC merupakan indikator performansi kurva *Receiver Operating Characteristic* (ROC) yang dapat meringkas kinerja sebuah *classifier* menjadi satu nilai. Formula untuk menghitung AUC ditunjukkan pada persamaan (2.37) berikut.

$$\text{AUC} = \frac{1}{2}\left(Sensitivity + Specificity\right) \tag{2.37}$$

Selain menggunakan nilai *accuracy* dan AUC, pada penelitian ini juga menggunakan *F*1-*score* untuk menunjukkan kebaikan suatu model prediksi. *F*1-*score* merupakan *harmonic average* dari *precision* dan *recall* dari hasil model prediksi. Formula untuk menghitung *F*1-*score* ditunjukkan pada persamaan (2.38) berikut.

$$F1\text{-}score = 2 \times \frac{Precision \times Recall}{\left(Precision + Recall\right)} \tag{2.38}$$

dimana

$$Precision = \frac{TP}{TP + FP}$$

dan

$$Recall = \frac{TP}{TP + FN}$$

2.13 *Financial Distress*

Financial distress merupakan kondisi kesulitan keuangan yang terjadi sebelum kebangkrutan (Platt & Platt, 2006). Sedangkan menurut Beaver *et al* (2010) dalam Dwijayanti (2010), *financial distress* didefinisikan sebagai ketidakmampuan perusahaan untuk membayar kewajiban finansial yang telah jatuh tempo. Menurut Asquith *et al* (1994) dalam Savera (2020), perusahaan yang mengalami *financial distress* merupakan perusahaan yang memiliki ICR (*Interest Coverage Ratio*) kurang dari 1. ICR (*Interest Coverage Ratio*) adalah rasio yang menjelaskan kemampuan perusahaan dalam membayar bunga

hutang dan menghindari kebangkrutan. Formula ICR (*Interest Coverage Ratio*) ialah sebagai berikut.

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

Sedangkan menurut Carolina *et al* (2017) terdapat beberapa rasio keuangan yang dapat digunakan untuk mengetahui apakah sebuah perusahaan sedang mengalami *financial distress* diantaranya sebagai berikut.

a. Likuiditas

Likuiditas didefinisikan sebagai rasio yang menunjukkan kemampuan perusahaan dalam memenuhi kewajiban jangka pendek (Kasmir, 2010). Likuiditas perusahaan dapat diketahui dari neraca dengan menghitung rasio lancar perusahaan. Formula rasio lancar (*current ratio*) ialah sebagai berikut.

$$\textit{Current Ratio} = \frac{\textit{Current Asset}}{\textit{Current Liabilities}}$$

b. Profitabilitas

Profitabilitas merupakan rasio yang menggambarkan kemampuan perusahaan dalam mendapatkan keuntungan. Selain itu, rasio ini juga memberikan ukuran tingkat efektivitas manajamen dalam suatu perusahaan (Kasmir, 2010). Profitabilitas perusahaan dapat diketahui dengan menghitung ROA (*Return on Asset*). Menurut Lestari dan Sugiharto (2007) dalam Carolina *et al* (2017), ROA merupakan rasio keuntungan bersih yang diperoleh dari penggunaan aktiva. Apabila ROA negatif, maka dapat diidentifikasikan bahwa perusahaan tersebut mengalami *financial distress*. Formula ROA (*Return on Asset*) ialah sebagai berikut.

$$\text{ROA} = \frac{\text{Laba bersih}}{\text{Total Aktiva}}$$

c. *Leverage*

Leverage didefinisikan sebagai kemampuan perusahaan dalam memenuhi kewajiban jangka pendek maupun jangka panjang (Sartono, 2001). Semakin tinggi rasio *leverage* perusahaan mengindikasikan bahwa perusahaan tersebut semakin banyak

menggunakan utang untuk membiayai kegiatan operasional. *Leverage* perusahaan dapat diketahui dengan menghitung DER (*Dept to Equity Ratio*). Formula DER (*Dept to Equity Ratio*) ialah sebagai berikut.

$$\text{DER} = \frac{\text{Total Hutang}}{\text{Total Modal}}$$

d. Arus Kas Operasi

Arus kas operasi merupakan indikator yang menentukan apakah dari operasi perusahaan menghasilkan kas yang dapat digunakan untuk melunasi pinjaman, memelihara kemampuan operasi, membayar dividen, dan melakukan investasi baru tanpa mengandalkan sumber pendanaan dari luar (Carolina, Marpaung, & Pratama, 2017).

2.14 Rasio Keuangan

Rasio keuangan merupakan ukuran yang menggambarkan kondisi keuangan suatu perusahaan. Rasio keuangan dapat diperoleh dari laporan keuangan seperti neraca. Neraca memuat informasi yang relevan untuk mengidentifikasi aliran dana masuk (*cash inflow*) dan aliran dana keluar (*cash outflow*) bersih yang dihasilkan oleh aset perusahaan. Selain itu, informasi juga dapat diperoleh melalui laporan laba rugi suatu perusahaan yang memuat laporan sejumlah uang masuk dan uang keluar pada kurun waktu tertentu yang dicatat untuk menentukan kondisi perusahaan berada dalam kondisi surplus atau bahkan defisit. Terdapat 4 jenis rasio keuangan yang digunakan dalam penelitian ini, yaitu rasio aktivitas, rasio profitabilitas, dan rasio solvabilitas, dan rasio likuiditas.

a. Rasio Aktivitas

Rasio aktivitas (*activity ratio*) merupakan rasio yang digunakan untuk mengukur efektivitas perusahaan dalam menggunakan aktiva yang dimilikinya atau dapat dikatakan sebagai rasio yang digunakan untuk mengukur tingkat efisiensi (efektivitas) pemanfaatan sumber daya perusahaan (Kasmir, 2019). Rasio aktivitas yang digunakan sebagai variabel dalam penelitian

ini ialah *Earning Before Income Tax to Asset*, *Sales to Total Asset*, *Inventory Turnover Ratio*, dan *Day's Sales in Inventory Ratio*.

b. Rasio Profitabilitas

Rasio profitabilitas merupakan rasio yang digunakan untuk menilai kemampuan perusahaan dalam mencari keuntungan. Selain itu, rasio ini juga memberikan ukuran tingkat efektivitas manajemen suatu perusahaan (Kasmir, 2019). Rasio profitabilitas yang digunakan sebagai variabel dalam penelitian ini ialah *Return on Equity*, *Net Profit Margin*, dan *Operating Profit Margin*.

c. Rasio Solvabilitas

Rasio solvabilitas atau *leverage ratio* merupakan rasio yang digunakan untuk mengukur sejauh mana aktiva perusahaan dibiayai dengan hutang atau dapat dikatakan sebagai rasio yang mengukur kemampuan perusahaan untuk membayar seluruh kewajibannya baik jangka pendek maupun jangka panjang apabila perusahaan dilikuidasi (Kasmir, 2019). Rasio solvabilitas yang digunakan sebagai variabel dalam penelitian ini ialah *Debt to Equity Ratio* dan *Debt to Asset Ratio*.

d. Rasio Likuiditas

Rasio likuiditas merupakan rasio yang berfungsi untuk menunjukkan atau mengukur kemampuan perusahaan dalam memenuhi kewajibannya yang sudah jatuh tempo, baik kewajiban kepada pihak luar perusahaan maupun di dalam perusahaan (Kasmir, 2019). Rasio likuiditas yang digunakan sebagai variabel dalam penelitian ini ialah *Current Ratio*, *Quick Ratio*, *Earning to Debt*, *Working Capital to Total Asset*, dan *Working Capital to Sales*.

2.15 Indikator Ekonomi Makro

Indikator ekonomi makro merupakan indikator yang digunakan untuk melihat stabilitas perekonomian suatu negara. Setiap perubahan dalam indikator ekonomi makro akan berdampak pada dinamika pertumbuhan ekonomi. Indikator ekonomi makro yang digunakan dalam penelitian ini ialah *Real GDP Growth*, BI 7-*day Repo Rate*, USD/IDR, dan Inflasi. *Real GDP Growth* merupakan ukuran pertumbuhan ekonomi berdasarkan

pertumbuhan Produk Domestik Bruto (PDB) dari suatu tahun ke tahun lainnya. Kemudian BI 7-day *Repo Rate* merupakan suku bunga acuan yang ditetapkan oleh Bank Indonesia untuk menjaga perekonomian tetap stabil. Sedangkan USD/IDR merupakan nilai tukar rupiah terhadap Dolar US dan Inflasi merupakan suatu keadaan naiknya harga barang dan jasa secara terus menerus yang berarti terjadi penurunan terhadap nilai mata uang.

BAB III
METODOLOGI PENELITIAN

3.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari penelitian Savera (2020) mengenai prediksi *financial distress* pada perusahaan sektor industri di Indonesia. Data yang digunakan dalam penelitian ini terdiri dari 80% data *training* dan 20% data *testing*. Variabel dalam penelitian ini terdiri dari 1 variabel respon yaitu kondisi keuangan perusahaan yang dilihat dari nilai ICR dan ROA, serta 18 variabel prediktor yang terdiri dari 14 rasio keuangan dan 4 indikator ekonomi makro yang dapat mempengaruhi kondisi keuangan perusahaan yang mengalami *financial distress*. Variabel yang digunakan dalam penelitian ini ditunjukkan pada Tabel 3.1.

Tabel 3.1 Variabel Penelitian.

Notasi	Variabel	Keterangan
Y	ICR atau ROA	Kondisi Keuangan Perusahaan: Jika ICR < 1 atau ROA < 0 maka Y = 1 (*Financial Distress*) Lainnya Y = 0 (Sehat Secara Finansial)
X_1	EBITA	*Earning Before Income Tax to Asset*
X_2	STA	*Sales to Total Asset*
X_3	ITR	*Inventory Turnover Ratio*
X_4	DSIR	*Day's Sales in Inventory Ratio*
X_5	ROE	*Return on Equity*
X_6	NPM	*Net Profit Margin*
X_7	OPM	*Operating Profit Margin*
X_8	DER	*Dept to Equity Ratio*
X_9	DAR	*Dept to Asset Ratio*
X_{10}	CR	*Current Ratio*
X_{11}	QR	*Quick Ratio*
X_{12}	ETD	*Earning to Dept*
X_{13}	WCA	*Working Capital to Total Asset*
X_{14}	WCS	*Working Capital to Sales Ratio*
X_{15}	rgdpg	*Real* GDP *Growth* (%yoy)
X_{16}	bi7drr	BI 7-*Day Repo Rate*

Tabel 3.1 Variabel Penelitian (Lanjutan).

Notasi	Variabel	Keterangan
X_{17}	USD/IDR	USD/IDR
X_{18}	Inflasi	Inflasi

Berikut merupakan definisi dari masing-masing variabel yang digunakan dalam penelitian ini.

1. *Earning Before Income Tax to Asset* (EBITA)

EBITA dapat diartikan sebagai uang yang didapatkan perusahaan sebelum dikurangi uang yang akan dibayarkan sebagai pajak. Penghasilan sebelum pajak menghitung laba operasi dan non operasi perusahaan sebelum pajak dipertimbangkan. Semakin besar nilai EBITA menunjukkan meningkatnya pendapatan perusahaan dan memperkecil peluang perusahaan mengalami *financial distress*.

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

2. *Sales to Total Asset* (STA)

STA digunakan untuk mengukur kemampuan perusahaan untuk menghasilkan penjualan berdasarkan aset. Ketika STA cukup tinggi, ini menjelaskan bahwa manajemen perusahaan mampu mendapatkan sebanyak mungkin penggunaan dari investasi dalam aset dan memperkecil peluang perusahaan mengalami *financial distress*.

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

3. *Inventory Turnover Ratio* (ITR)

ITR merupakan rasio yang menunjukkan kemampuan perusahaan dalam menjual dan mengganti persediaan selama periode tertentu. ITR berguna untuk membantu perusahaan dalam membuat keputusan yang lebih baik mengenai harga, pemasaran, manufaktur, dan pembelian *inventory* baru. Semakin tinggi nilai ITR menunjukkan perusahaan dapat menjual produk dengan cepat dan ada banyak permintaan, sehingga memperkecil peluang perusahaan mengalami *financial distress*.

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

4. *Day's Sales in Inventory Ratio* (DSIR)

DSIR merupakan rasio yang menunjukkan waktu rata-rata hari yang dibutuhkan perusahaan untuk mengubah persediaannya menjadi penjualan. Semakin kecil DSIR menunjukkan waktu yang lebih singkat untuk membersihkan *inventory*. Semakin besar DSIR menunjukkan bahwa perusahaan tidak mengelola persediaannya dengan baik atau memiliki persediaan yang sulit untuk dijual, sehingga apabila tidak segera diatasi dapat mengakibatkan perusahaan mengalami *financial distress*.

$$\text{DSIR} = \frac{Inventory}{\text{COGS}} \times 365$$

5. *Return on Equity* (ROE)

ROE berguna sebagai informasi kepada investor mengenai seberapa efisien kinerja perusahaan. Semakin tinggi ROE, maka semakin efisien perusahaan dalam menghasilkan pendapatan dan pertumbuhan dari pembiayaan ekuitasnya, sehingga memperkecil peluang perusahaan mengalami *financial distress*.

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

6. *Net Profit Margin* (NPM)

NPM merupakan rasio yang menunjukkan banyaknya pendapatan atau laba bersih yang dihasilkan dari persentase pendapatan. Perusahaan dapat menilai apakah praktik saat ini berhasil dan memperkirakan laba berdasarkan pendapatan dengan melacak kenaikan dan penurunan NPM. Tingginya nilai NPM menunjukkan perusahaan dapat secara efektif mengendalikan biaya atau menyediakan barang atau jasa dengan harga yang lebih tinggi daripada biayanya, sehingga memperkecil peluang perusahaan mengalami *financial distress*.

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

7. *Operating Profit Margin* (OPM)

OPM merupakan rasio yang menggambarkan persentase keuntungan yang dihasilkan perusahaan dari operasinya sebelum dikurangi pajak dan bunga. OPM akan semakin meningkat melalui manajemen yang semakin baik, penggunaan sumber daya yang lebih efisien, harga yang lebih baik, dan pemasaran yang lebih efektif. Semakin tingginya nilai OPM menunjukkan peluang perusahaan mengalami *financial distress* semakin kecil.

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

8. *Dept to Equity Ratio* (DER)

DER merupakan rasio yang menggambarkan sejauh mana perusahaan dapat membiayai operasinya melalui hutang dengan dana yang dimiliki sepenuhnya. DER mencerminkan kemampuan ekuitas pemegang saham untuk menutupi semua hutang yang belum dibayar apabila terjadi penurunan kondisi keuangan perusahaan. Tingginya DER dikaitkan dengan risiko yang tinggi, artinya sebuah perusahaan cukup agresif dalam mendanai pertumbuhannya dengan hutang.

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

9. *Dept to Asset Ratio* (DAR)

DAR digunakan untuk mengukur pertumbuhan perusahaan melalui aset yang diperoleh dari waktu ke waktu. Investor dapat menggunakan DAR untuk mengevaluasi apakah suatu perusahaan memiliki cukup dana untuk memenuhi kewajiban utangnya dan untuk menilai apakah suatu perusahaan dapat membayar pengembalian investasi. Semakin tinggi nilai DAR menunjukkan kecilnya peluang perusahaan mengalami *financial distress*.

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

10. *Current Ratio* (CR)

CR merupakan rasio likuiditas yang mengukur kemampuan perusahaan untuk membayar kewajiban jangka pendek. Investor

dapat mengetahui bagaimana perusahaan dapat memaksimalkan aset lancar di neracanya untuk memenuhi utang saat ini dan hutang lainnya. Semakin kecil CR di suatu perusahaan dapat menunjukkan risiko kesulitan atau gagal bayar yang tinggi dan apabila tidak segera diatasi dapat mengakibatkan perusahaan mengalami *financial distress*.

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

11. *Quick Ratio* (QR)

QR merupakan rasio likuiditas jangka pendek perusahaan untuk mengukur kemampuan perusahaan dalam memenuhi kewajiban jangka pendeknya dengan aset likuidnya. Aset likuid adalah aset yang dapat dengan cepat diubah menjadi uang tunai dengan dampak minimal terhadap harga yang diterima di pasar. Semakin tinggi QR menunjukkan semakin baik kesehatan keuangan perusahaan dan memperkecil peluang perusahaan mengalami *financial distress*.

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

12. *Earning to Dept* (ETD)

ETD merupakan rasio yang mengukur tingkat keuangan suatu perusahaan yang membandingkan pendapatannya dengan utang atau kewajiban yang harus dibayar perusahaan. Semakin tinggi ETD menunjukkan kondisi keuangan perusahaan semakin baik dan memperkecil peluang perusahaan mengalami *financial distress*.

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

13. *Working Capital to Total Asset* (WCA)

WCA merupakan rasio yang mengukur kemampuan perusahaan untuk menghasilkan modal kerja dari total aset yang dimiliki. Modal kerja yang dimaksud merupakan sebagian dari aktiva lancar yanag dapat digunakan untuk membiayai operasional perusahaan tanpa mempengaruhi likuiditasnya. Semakin tinggi

WCA menunjukkan likuiditas perusahaan meningkat dari waktu ke waktu.

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

14. *Working Capital to Sales Ratio* (WCS)

WCS merupakan rasio yang mengukur seberapa efisien perusahaan menggunakan modal kerjanya untuk mendukung tingkat penjualannya. WCS bersifat jangka pendek, sehingga dalam menentukan WCS harus membandingkannya dengan angka-angka sebelumnya untuk melihat apakah hal ini biasa terjadi pada perusahaan atau merupakan suatu pengecualian.

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

Struktur data yang digunakan pada penelitian ini ditampilkan pada Tabel 3.2.

Tabel 3.2 Struktur Data Penelitian.

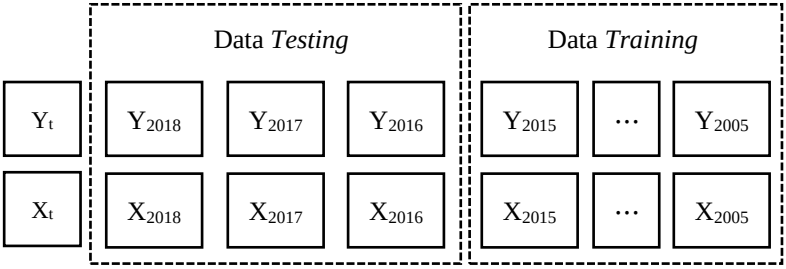
Perusahaan	**Periode**	**t**	$\mathbf{Y_{it}}$	$\mathbf{X_{it,1}}$	$\mathbf{X_{it,2}}$	$\cdots$	$\mathbf{X_{it,18}}$
1	$2005Q_4$	1	Y_{11}	$X_{11,1}$	$X_{11,2}$		$X_{11,18}$
	$2006Q_4$	2	Y_{12}	$X_{12,1}$	$X_{12,2}$	$\ddots$	$X_{12,18}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
	$2018Q_4$	t	Y_{1t}	$X_{1t,1}$	$X_{1t,2}$		$X_{1t,18}$
2	$2005Q_4$	1	Y_{21}	$X_{21,1}$	$X_{21,2}$		$X_{21,18}$
	$2006Q_4$	2	Y_{22}	$X_{22,1}$	$X_{22,2}$	$\ddots$	$X_{22,18}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
	$2018Q_4$	t	Y_{2t}	$X_{2t,1}$	$X_{2t,2}$		$X_{2t,18}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	$2005Q_4$	1	Y_{i1}	$X_{i1,1}$	$X_{i1,2}$		$X_{i1,18}$
	$2006Q_4$	2	Y_{i2}	$X_{i2,1}$	$X_{i2,2}$	$\ddots$	$X_{i2,18}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
	$2018Q_4$	t	Y_{it}	$X_{it,1}$	$X_{it,2}$		$X_{it,18}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
m	$2005Q_4$	1	Y_{m1}	$X_{m1,1}$	$X_{m1,2}$		$X_{m1,18}$
	$2006Q_4$	2	Y_{m2}	$X_{m2,1}$	$X_{m2,2}$	$\ddots$	$X_{m2,18}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
	$2018Q_4$	t	Y_{mt}	$X_{mt,1}$	$X_{mt,2}$		$X_{mt,18}$

Data pada penelitian ini menggunakan konsep *drift* yaitu menganalisis pada waktu ke-*t* dan waktu-waktu sebelumnya. Data penelitian akan dibagi menjadi 4 *time window size* (0, 1, 2, dan 3) yang akan digunakan untuk menentukan apakah suatu perusahaan mengalami *financial distress* atau tidak, serta membagi data menjadi 80% data *training* dan 20% data *testing*. Jumlah unit observasi dan perbandingan perusahaan yang mengalami *financial distress* dan perusahaan dengan kondisi sehat secara finansial pada tiap *size* ditunjukkan pada Tabel 3.3 berikut.

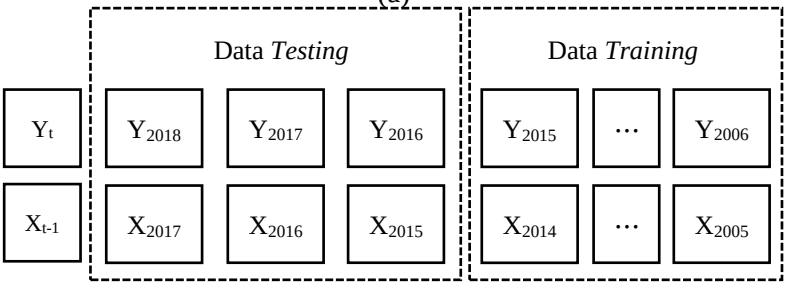
Tabel 3.3 Perbandingan Jumlah Unit Observasi Tiap *Size*.

Size	**Training**			**Testing**		
	n	**Y = 0**	**Y = 1**	**n**	**Y = 0**	**Y = 1**
0	1595	1213	382	435	304	131
1	1450	1101	349	435	304	131
2	1305	988	317	435	304	131
3	1160	872	288	435	304	131

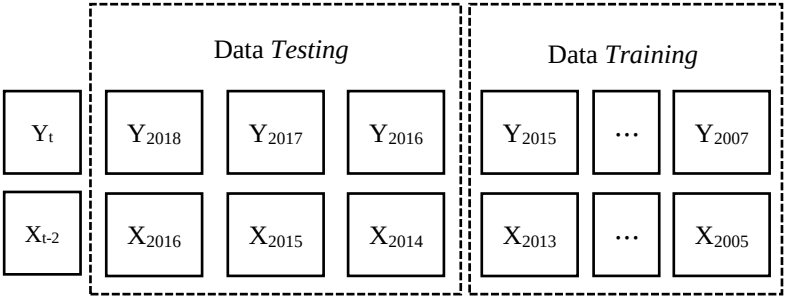
Ilustrasi konsep *drift* untuk keempat *size* pada penelitian ini ditampilkan pada Gambar 3.1 (a), (b), (c), dan (d).



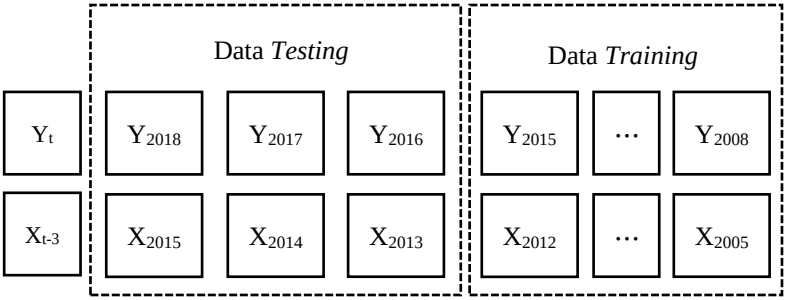
(a)



(b)



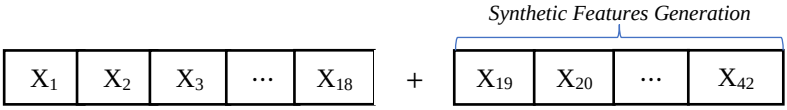
(c)



(d)

Gambar 3.1 Ilustrasi Konsep *Drift* dengan *Size* 0 (a), *Size* 1 (b), *Size* 2 (c), dan *Size* 3 (d).

Setelah data dikelompokkan menjadi 4 *time window size*, selanjutnya pada variabel pada tiap *size* dilakukan *synthetic features generation* dengan ilustrasi ditampilkan pada Gambar 3.2.



Gambar 3.2 Ilustrasi *Synthetic Features Generation*.

3.2 Langkah Analisis

Langkah-langkah analisis yang dilakukan dalam mencapai tujuan penelitian ini adalah sebagai berikut.

1. Melakukan prediksi *financial distress* menggunakan metode *Extreme Gradient Boosting* (XGB) secara serentak dan seleksi variabel dengan uraian sebagai berikut ini.

a. Menentukan *time window size* dan mengelompokkan data berdasarkan *size* dengan konsep *drift*.
b. Membagi data menjadi 80% data *training* dan 20% data *testing*.
c. Prediksi *financial distress* menggunakan *Extreme Gradient Boosting* (XGB) secara serentak dengan tahapan sebagai berikut:
   i. Membentuk model XGB menggunakan data *training*.
   ii. Melakukan prediksi dari model XGB menggunakan persamaan (2.26).
   iii. Melakukan evaluasi menggunakan data *training* dan *testing*, serta menghitung nilai *accuracy*, AUC, dan *F*1-*score* hasil prediksi menggunakan formula (2.34), (2.37), dan (2.38).
d. Prediksi *financial distress* menggunakan *Extreme Gradient Boosting* (XGB) dengan seleksi variabel.
   i. Mengeliminasi variabel yang tidak signifikan berdasarkan nilai *gain* yang didapatkan dari formula (2.31).
   ii. Membentuk model XGB dengan variabel yang telah diseleksi dan melakukan prediksi menggunakan persamaan (2.26).
   iii. Melakukan evaluasi menggunakan data *training* dan *testing*, serta menghitung nilai *accuracy*, AUC, dan *F*1-*score* hasil prediksi menggunakan formula (2.34), (2.37), dan (2.38).

2. Melakukan prediksi *financial distress* menggunakan metode klasifikasi (*Generalized Extreme Value Regression* (GEVR), *Logistic Regression*, *Support Vector Machine* (SVM), dan *Extreme Gradient Boosting* (XGB)) dan melibatkan *synthetic features generation* secara serentak dan seleksi variabel dengan uraian sebagai berikut ini.
   a. Menentukan *time window size* dan mengelompokkan data berdasarkan *size* dengan konsep *drift*.

b. Membentuk variabel baru menggunakan *synthetic features generaton*.
c. Membagi data menjadi 80% data *training* dan 20% data *testing*.
d. Prediksi *financial distress* menggunakan *Generalized Extreme Value Regression* (GEVR) dengan tahapan sebagai berikut:
   i. Membentuk model GEVR menggunakan data *training* dengan persamaan (2.3).
   ii. Melakukan prediksi dari model GEVR dengan formula (2.2).
   iii. Melakukan evaluasi menggunakan data *training* dan *testing*, serta menghitung nilai *accuracy*, AUC, dan *F*1-*score* hasil prediksi menggunakan formula (2.34), (2.37), dan (2.38).
e. Prediksi *financial distress* menggunakan *Generalized Extreme Value Regression* (GEVR) dengan seleksi variabel.
   i. Melakukan uji signifikansi parameter berdasarkan subbab 2.11 dan mengeliminasi variabel yang tidak signifikan dalam model dengan prosedur *stepwise*.
   ii. Membentuk model GEVR dengan variabel yang telah diseleksi dan melakukan prediksi.
   iii. Melakukan evaluasi menggunakan data *training* dan *testing*, serta menghitung nilai *accuracy*, AUC, dan *F*1-*score* hasil prediksi menggunakan formula (2.34), (2.37), dan (2.38).
f. Prediksi *financial distress* menggunakan *Logistic Regression* dengan tahapan sebagai berikut:
   i. Membentuk model *Logistic Regression* menggunakan data *training* dengan persamaan (2.6).
   ii. Melakukan prediksi dari model *Logistic Regression* dengan formula (2.5).
   iii. Melakukan evaluasi menggunakan data *training* dan *testing*, serta menghitung nilai *accuracy*, AUC, dan

*F*1-*score* hasil prediksi menggunakan formula (2.34), (2.37), dan (2.38).

g. Prediksi *financial distress* menggunakan *Logistic Regression* dengan seleksi variabel.
   i. Melakukan uji signifikansi parameter berdasarkan subbab 2.11 dan mengeliminasi variabel yang tidak signifikan dalam model dengan prosedur *stepwise*.
   ii. Membentuk model *Logistic Regression* menggunakan variabel yang telah diseleksi dengan persamaan (2.6) dan melakukan prediksi dengan formula (2.5).
   iii. Melakukan evaluasi menggunakan data *training* dan *testing*, serta menghitung nilai *accuracy*, AUC, dan *F*1-*score* hasil prediksi menggunakan formula (2.34), (2.37), dan (2.38).

h. Prediksi *financial distress* menggunakan *Lasso Logistic Regression* dengan tahapan sebagai berikut:
   i. Menentukan nilai λ yang digunakan untuk membentuk model.
   ii. Membentuk model *Lasso Logistic Regression* menggunakan data *training* dengan persamaan (2.10).
   iii. Melakukan prediksi dari model *Lasso Logistic Regression*.
   iv. Melakukan evaluasi menggunakan data *training* dan *testing*, serta menghitung nilai *accuracy*, AUC, dan *F*1-*score* hasil prediksi menggunakan formula (2.34), (2.37), dan (2.38).
   v. Menentukan nilai λ optimal berdasarkan hasil performansi model.

i. Prediksi *financial distress* menggunakan *Elastic-Net Logistic Regression* dengan tahapan sebagai berikut:
   i. Menentukan nilai λ dan γ yang digunakan untuk membentuk model.

ii. Membentuk model *Elastic-Net Logistic Regression* menggunakan data *training* dengan persamaan (2.11).
iii. Melakukan prediksi dari model *Elastic-Net Logistic Regression*.
iv. Melakukan evaluasi menggunakan data *training* dan *testing*, serta menghitung nilai *accuracy*, AUC, dan *F*1-*score* hasil prediksi menggunakan formula (2.34), (2.37), dan (2.38).
v. Menentukan nilai λ dan γ optimal berdasarkan hasil performansi model.

j. Prediksi *financial distress* menggunakan *Support Vector Machine* (SVM) dengan tahapan sebagai berikut:
i. Membentuk model SVM menggunakan data *training*.
ii. Melakukan prediksi dari model SVM.
iii. Melakukan evaluasi menggunakan data *training* dan *testing*, serta menghitung nilai *accuracy*, AUC, dan *F*1-*score* hasil prediksi menggunakan formula (2.34), (2.37), dan (2.38).

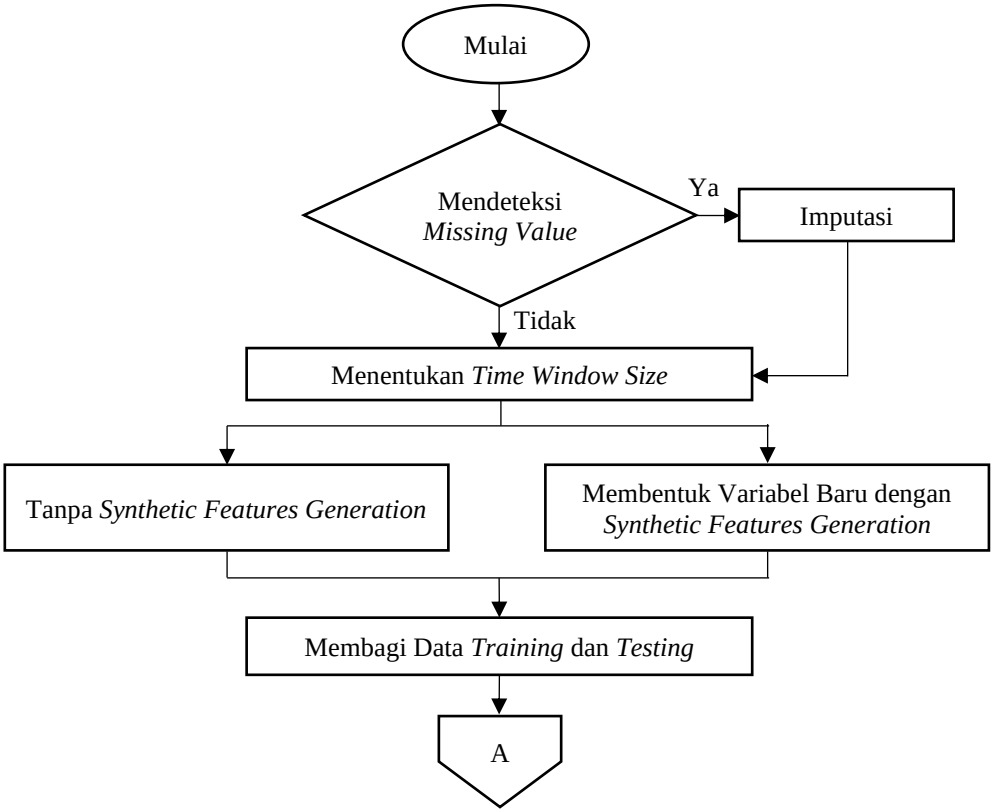
k. Prediksi *financial distress* menggunakan *Lasso Support Vector Machine* dengan tahapan sebagai berikut:
i. Menentukan nilai λ yang digunakan untuk membentuk model.
ii. Membentuk model *Lasso Support Vector Machine* menggunakan data *training* dengan persamaan (2.22).
iii. Melakukan prediksi dari model *Lasso Support Vector Machine*.
iv. Melakukan evaluasi menggunakan data *training* dan *testing*, serta menghitung nilai *accuracy*, AUC, dan *F*1-*score* hasil prediksi menggunakan formula (2.34), (2.37), dan (2.38).
v. Menentukan nilai λ optimal berdasarkan hasil performansi model.

l. Prediksi *financial distress* menggunakan *Elastic-Net Support Vector Machine* dengan tahapan sebagai berikut:
   i. Menentukan nilai λ_1 dan λ_2 yang digunakan untuk membentuk model.
   ii. Membentuk model *Elastic-Net Support Vector Machine* menggunakan data *training* dengan persamaan (2.23).
   iii. Melakukan prediksi dari model *Elastic-Net Support Vector Machine*.
   iv. Melakukan evaluasi menggunakan data *training* dan *testing*, serta menghitung nilai *accuracy*, AUC, dan *F*1-*score* hasil prediksi menggunakan formula (2.34), (2.37), dan (2.38).
   v. Menentukan nilai λ_1 dan λ_2 optimal berdasarkan hasil performansi model.
m. Prediksi *financial distress* menggunakan *Extreme Gradient Boosting* (XGB) dengan tahapan sebagai berikut:
   i. Membentuk model XGB menggunakan data *training*.
   ii. Melakukan prediksi dari model XGB dengan persamaan (2.26).
   iii. Melakukan evaluasi menggunakan data *training* dan *testing*, serta menghitung nilai *accuracy*, AUC, dan *F*1-*score* hasil prediksi menggunakan formula (2.34), (2.37), dan (2.38).
n. Prediksi *financial distress* menggunakan *Extreme Gradient Boosting* (XGB) dengan seleksi variabel.
   i. Mengeliminasi variabel yang tidak signifikan berdasarkan nilai *gain* yang didapatkan dari formula (2.31).
   ii. Membentuk model XGB dengan variabel yang telah diseleksi dan melakukan prediksi menggunakan persamaan (2.26).

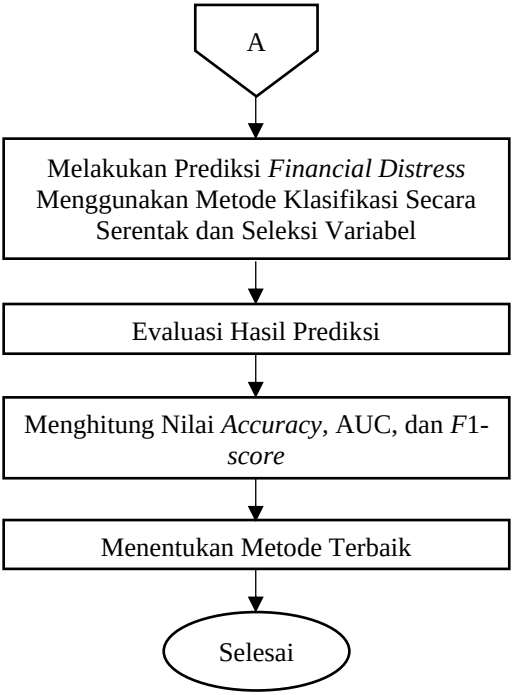
iii. Melakukan evaluasi menggunakan data *training* dan *testing*, serta menghitung nilai *accuracy*, AUC, dan *F*1-*score* hasil prediksi menggunakan formula (2.34), (2.37), dan (2.38).

3. Menentukan metode terbaik berdasarkan nilai *accuracy*, AUC, dan *F*1-*score* untuk memprediksi *financial distress* perusahaan sektor industri di Indonesia serta melakukan perbandingan model secara serentak dan seleksi variabel berdasarkan nilai *sensitivity*.

Berdasarkan langkah analisis yang telah dijelaskan sebelumnya, visualisasi diagram alirnya ditampilkan pada Gambar 3.3.



Gambar 3.3 Diagram Alir.



Gambar 3.3 Diagram Alir (Lanjutan).

Dikarenakan penelitian Tugas Akhir ini melanjutkan penelitian Savera (2020), sehingga beberapa langkah analisis yang sudah dilakukan tidak dianalisis kembali pada penelitian ini. Adapun langkah analisis yang telah dilakukan pada penelitian sebelumnya ditunjukkan pada Tabel 3.4.

Tabel 3.4 Perbandingan Langkah Analisis dengan Penelitian Sebelumnya.

Langkah Analisis	**Penelitian Savera (2020)**	**Penelitian Ini**
Mendeteksi *missing value* dan melakukan imputasi dengan *k-Nearest Neighbor* (kNN).	✓	✗

Tabel 3.4 Perbandingan Langkah Analisis dengan Penelitian Sebelumnya (Lanjutan).

Langkah Analisis	**Penelitian Savera (2020)**	**Penelitian Ini**
Menentukan *time window size* dan analisis statistika deskriptif.	✓	✗
Melakukan prediksi *financial distress* menggunakan *Generalized Extreme Value Regression* (GEVR), *Support Vector Machine* (SVM), dan *Logistic Regression* secara serentak dan seleksi variabel.	✓	✗
Melakukan prediksi *financial distress* menggunakan *Extreme Gradient Boosting* (XGB)	✗	✓
Menambahkan variabel baru dengan metode *synthetic features generation*.	✗	✓
Melakukan prediksi *financial distress* menggunakan metode klasifikasi (*Generalized Extreme Value Regression* (GEVR), *Logistic Regression*, *Support Vector Machine* (SVM), dan *Extreme Gradient Boosting* (XGB)) dengan melibatkan *synthetic features generation* secara serentak dan seleksi variabel.	✗	✓

BAB IV
ANALISIS DAN PEMBAHASAN

4.1 Metode Klasifikasi Tanpa *Synthetic Features Generation*

Prediksi *financial distress* dengan metode klasifikasi tanpa melibatkan *synthetic features generation* dilakukan dengan metode *Extreme Gradient Boosting* (XGB). Pada penelitian ini akan dibandingkan kebaikan model *Extreme Gradient Boosting* (XGB) secara serentak dan seleksi variabel pada tiap *size* dengan pembahasan sebagai berikut.

4.1.1 *Extreme Gradient Boosting* (XGB) Secara Serentak

Pada penelitian ini model *Extreme Gradient Boosting* (XGB) memberikan hasil yang optimum dengan menggunakan parameter $\lambda = 1$ dan $\gamma = 0.01$. Hasil prediksi *financial distress* perusahaan sektor industri di Indonesia menggunakan model *Extreme Gradient Boosting* (XGB) tanpa melibatkan *synthetic features generation* secara serentak pada tiap *size* ditunjukkan pada Tabel 4.1.

Tabel 4.1a *Confusion Matrix* Model XGB Tanpa *Synthetic Features Generation* Untuk *Size* 0.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1213	0	0	299	5
1	2	380	1	13	118

Tabel 4.1b *Confusion Matrix* Model XGB Tanpa *Synthetic Features Generation* Untuk *Size* 1.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1084	17	0	265	39
1	50	299	1	26	105

Tabel 4.1c *Confusion Matrix* Model XGB Tanpa *Synthetic Fatures Genereation* Untuk *Size* 2.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	970	18	0	274	30
1	54	263	1	45	86

Tabel 4.1d *Confusion Matrix* Model XGB Tanpa *Synthetic Features Generation* Untuk *Size* 3.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	864	8	0	285	19
1	42	246	1	67	64

Berdasarkan *confusion matrix* pada Tabel 4.1 dapat dilihat ketepatan hasil prediksi *financial distress* perusahaan sektor industri di Indonesia dengan model *Extreme Gradient Boosting* (XGB) pada tiap *size*. Kemudian dilakukan evaluasi hasil prediksi dari model tersebut dengan menghitung nilai *accuracy*, AUC, dan *F*1-*score* pada tiap *size* yang ditampilkan pada Tabel 4.2.

Tabel 4.2 Evaluasi Hasil Klasifikasi Model XGB Tanpa *Synthetic Features Generation*.

Size	**Ukuran**	***Training***	***Testing***
0	*Accuracy*	99,87%	95,86%
	AUC	99,74%	94,22%
	*F*1-*score*	99,92%	97,08%
1	*Accuracy*	95,38%	85,06%
	AUC	92,07%	83,66%
	*F*1-*score*	97,00%	89,08%
2	*Accuracy*	94,48%	82,76%
	AUC	90,58%	77,89%
	*F*1-*score*	96,42%	87,96%
3	*Accuracy*	95,69%	80,23%
	AUC	92,25%	71,30%
	*F*1-*score*	97,19%	86,89%

Tabel 4.2 menunjukkan bahwa model *Extreme Gradient Boosting* (XGB) sangat baik dalam memprediksi kondisi *financial distress* pada *size* 0. Semakin besar *size* maka akan semakin rendah kebaikan model *Extreme Gradient Boosting* (XGB) dalam memprediksi kondisi *financial distress* perusahaan sektor industri di Indonesia. Hasil klasifikasi dengan model *Extreme Gradient Boosting* (XGB) menggunakan data *testing* pada *size* 1, *size* 2, dan *size* 3 cenderung menurun, tetapi pada *size* 3 dengan data *training* lebih baik dibandingkan *size* 1 dan *size* 2, sehingga model pada *size* 3 ada kecenderungan *overfitting*.

4.1.2 *Extreme Gradient Boosting* (XGB) dengan Seleksi Variabel

Salah satu alasan yang mengakibatkan rendahnya hasil klasifikasi ialah terdapatnya variabel prediktor yang tidak signifikan dalam model, sehingga perlu dilakukan seleksi variabel untuk mengeliminasi variabel prediktor yang tidak signifikan dalam model. Seleksi variabel pada model *Extreme Gradient Boosting* dilakukan dengan melihat nilai *gain*. Variabel yang memiliki nilai *gain* sangat kecil (mendekati nol) akan dianggap sebagai variabel yang tidak berpengaruh signifikan dalam model. Nilai *gain* pada tiap variabel prediktor untuk tiap *size* ditunjukkan pada Tabel 4.3.

Tabel 4.3 Nilai *Gain* Tiap Variabel Prediktor Model XGB Tanpa *Synthetic Features Generation*.

Size	**Variabel**	***Gain***	***Size***	**Variabel**	***Gain***
0	ETD	0.361	1	EBITA	0.296
	EBITA	0.292		OPM	0.116
	NPM	0.148		ROE	0.081
	OPM	0.060		ETD	0.070
	ROE	0.033		NPM	0.058
	STA	0.023		DAR	0.054
	QR	0.022		STA	0.041
	DER	0.017		QR	0.040
	WCS	0.009		CR	0.040
	DAR	0.008		DER	0.037
	CR	0.008		WCS	0.035
	ITR	0.006		WCA	0.034
	WCA	0.006		ITR	0.026
	rgdpg	0.002		DSIR	0.022
	USD/IDR	0.002		rgdpg	0.014
	DSIR	0.002		USD/IDR	0.013
	bi7drr	0.001		Inflasi	0.013
	Inflasi	0.000		bi7drr	0.011
2	EBITA	0.199	3	EBITA	0.240
	OPM	0.128		ETD	0.122
	ETD	0.113		OPM	0.105
	NPM	0.087		STA	0.068
	ROE	0.070		QR	0.059

Tabel 4.3 Nilai Gain Tiap Variabel Prediktor Model XGB Tanpa *Synthetic Features Generation* (Lanjutan).

Size	**Variabel**	***Gain***	***Size***	**Variabel**	***Gain***
2	STA	0.066	3	WCS	0.054
	WCS	0.056		ROE	0.053
	DAR	0.055		DAR	0.052
	QR	0.039		NPM	0.046
	ITR	0.035		DER	0.041
	WCA	0.032		ITR	0.034
	rgdpg	0.024		CR	0.032
	DER	0.022		WCA	0.030
	CR	0.022		rgdpg	0.018
	DSIR	0.018		DSIR	0.016
	USD/IDR	0.018		bi7drr	0.015
	Inflasi	0.014		USD/IDR	0.009
	bi7drr	0.003		Inflasi	0.006

Berdasarkan Tabel 4.3 dapat diketahui variabel prediktor yang berpengaruh signifikan dalam model *Extreme Gradient Boosting* (XGB) pada tiap *size*. Eliminasi variabel dilakukan dengan cara melakukan klasifikasi tanpa menggunakan variabel prediktor yang memiliki nilai *gain* terkecil dan eliminasi variabel berhenti ketika evaluasi hasil klasifikasi memiliki nilai yang lebih kecil dibandingkan nilai sebelumnya. Variabel yang berpengaruh signifikan dalam model *Extreme Gradient Boosting* (XGB) pada tiap *size* ditunjukkan pada Tabel 4.4.

Tabel 4.4 Variabel Prediktor yang Signifikan dalam Model XGB Tanpa *Synthetic Features Generation*.

Size	**Variabel**
0	EBITA, STA, ITR, ROE, NPM, OPM, DER, DAR, CR, QR, ETD, WCA, WCS
1	EBITA, STA, ITR, DSIR, ROE, NPM, OPM, DER, DAR, CR, QR, ETD, WCA, WCS, rgdpg, USD/IDR
2	EBITA, STA, ITR, DSIR, ROE, NPM, OPM, DER, DAR, CR, QR, ETD, WCA, WCS, rgdpg, USD/IDR
3	EBITA, STA, ITR, DSIR, ROE, NPM, OPM, DER, DAR, CR, QR, ETD, WCA, WCS, rgdpg

Hasil prediksi *financial distress* perusahaan sektor industri di Indonesia menggunakan model *Extreme Gradient Boosting* (XGB) tanpa melibatkan *synthetic features generation* dengan seleksi variabel pada tiap *size* ditunjukkan pada *confusion matrix* pada Tabel 4.5.

Tabel 4.5a *Confusion Matrix* Model XGB Tanpa *Synthetic Features Generation* dengan Seleksi Variabel Untuk *Size* 0.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1213	0	0	299	5
1	2	380	1	13	118

Tabel 4.5b *Confusion Matrix* Model XGB Tanpa *Synthetic Features Generation* dengan Seleksi Variabel Untuk *Size* 1.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1087	14	0	272	32
1	47	302	1	34	97

Tabel 4.5c *Confusion Matrix* Model XGB Tanpa *Synthetic Features Generation* dengan Seleksi Variabel Untuk *Size* 2.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	972	16	0	282	22
1	51	266	1	45	86

Tabel 4.5d *Confusion Matrix* Model XGB Tanpa *Synthetic Features Generation* dengan Seleksi Variabel Untuk *Size* 3.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	861	11	0	281	23
1	56	232	1	59	72

Berdasarkan *confusion matrix* pada Tabel 4.5 dapat dilihat ketepatan hasil prediksi *financial distress* perusahaan sektor industri di Indonesia menggunakan model *Extreme Gradient Boosting* (XGB) dengan seleksi variabel pada tiap *size*. Selanjutnya dilakukan evaluasi hasil prediksi dari model tersebut dengan menghitung nilai *accuracy*, AUC, dan *F*1-*score* pada tiap *size* yang ditampilkan pada Tabel 4.6.

Tabel 4.6 Evaluasi Hasil Klasifikasi Model XGB Tanpa *Synthetic Features Generation* dengan Seleksi Variabel.

Size	**Ukuran**	***Training***	***Testing***
0	*Accuracy*	99,87%	95,86%
	AUC	99,74%	94,22%
	*F*1-*score*	99,92%	97,08%
1	*Accuracy*	95,79%	84,83%
	AUC	92,63%	81,76%
	*F*1-*score*	97,27%	89,18%
2	*Accuracy*	94,87%	84,60%
	AUC	91,15%	79,21%
	*F*1-*score*	96,67%	89,38%
3	*Accuracy*	94,22%	81,15%
	AUC	89,65%	73,70%
	*F*1-*score*	96,25%	87,27%

Tabel 4.6 menunjukkan bahwa berdasarkan nilai *accuracy*, AUC, dan *F*1-*score*, model *Extreme Gradient Boosting* (XGB) dengan seleksi variabel sangat baik dalam memprediksi kondisi *financial distress* pada *size* 0. Hasil klasifikasi menggunakan model *Extreme Gradient Boosting* (XGB) dengan seleksi variabel meningkatkan kebaikan model prediksi *financial distress* perusahaan sektor industri di Indonesia dibandingkan model secara serentak terutama pada *size* 1, *size* 2, dan *size* 3, sedangkan pada *size* 0 cenderungan memberikan hasil klasifikasi yang sama.

4.2 Metode Klasifikasi dengan *Synthetic Features Generation*

Prediksi *financial distress* dengan metode klasifikasi dan melibatkan *synthetic features generation* dilakukan dengan metode *Generalized Extreme Value Regression* (GEVR), *Logistic Regression*, *Support Vector Machine* (SVM), dan *Extreme Gradient Boosting* (XGB). *Synthetic features generation* merupakan metode yang mengkombinasikan dua variabel menggunakan operasi aritmatika seperti penjumlahan, pengurangan, perkalian, dan pembagian. Menggunakan metode *synthetic features generation* terbentuk variabel baru sebesar 936 variabel. Kelemahan metode tersebut ialah memungkinkan terbentuk variabel baru yang memiliki korelasi yang tinggi, sehingga sebelum dilakukan analisis dengan metode klasifikasi

terlebih dahulu dilakukan eliminasi variabel yang memiliki korelasi yang cukup tinggi. Pada penelitian ini ditetapkan batasan korelasi yang digunakan ialah 50%, sehingga variabel yang memiliki korelasi di atas 50% akan dieliminasi dari variabel penelitian. Variabel baru hasil *synthetic features generation* yang telah dieliminasi ditampilkan pada Lampiran 1. Menggunakan variabel pada Lampiran 1 tersebut, pada penelitian ini akan dibandingkan kebaikan model *Generalized Extreme Value Regression* (GEVR), *Logistic Regression*, *Support Vector Machine* (SVM), dan *Extreme Gradient Boosting* (XGB) secara serentak dan seleksi variabel pada tiap *size* dengan pembahasan sebagai berikut.

4.2.1 *Generalized Extreme Value Regression* dengan *Synthetic Features Generation* Secara Serentak

Pada penelitian ini model *Generalized Extreme Value Regression* (GEVR) dilakukan dengan menggunakan parameter τ bernilai 0,1 karena merupakan τ yang paling optimum pada penelitian ini. Sehingga hasilnya dapat dibandingkan dan dapat melihat pengaruh metode *synthetic features generation* dalam hasil klasifikasi model *Generalized Extreme Value Regression* (GEVR). Fungsi *link* dari model *Generalized Extreme Value Regression* (GEVR) pada tiap *size* ditunjukkan pada Tabel 4.7.

Tabel 4.7 Model GEVR dengan *Synthetic Features Generation* Tiap *Size*.

Size	Model
0	$g(x) = 14,090 - 39,510\text{EBITA} - 2,533\text{STA} - 0,048\text{ITR} - 0,016\text{DSIR} - 1,922\text{ROE} - 5,695\text{NPM} - 21,400\text{OPM} - 0,005\text{DER} + 1,830\text{DAR} - 0,160\text{CR} + 0,171\text{QR} - 5,024\text{ETD} + 2,705\text{WCA} - 0,855\text{WCS} + 0,215\text{rgdpg} + 0,215\text{bi7drr} + 0,0001\text{USD/IDR} - 0,082\text{Inflasi} + 112(\text{EBITA} \times \text{EBITA}) - 0,949(\text{EBITA} \times \text{DAR}) - 3,536(\text{EBITA} \times \text{WCA}) - 0,029(\text{EBITA/WCS}) + 0,148(\text{STA} \times \text{STA}) + 3,503(\text{STA/ITR}) + 0,00004(\text{STA/WCS}) - 0,038(\text{ITR} \times \text{DSIR}) - 3,424(\text{ITR} \times \text{OPM}) - 0,00015(\text{ITR/NPM}) + 0,151(\text{ITR/QR}) + 0,005(\text{DSIR/STA}) + 15,980(\text{ROE} \times \text{NPM}) + 0,870(\text{ROE/QR}) - 0,069(\text{NPM/ROE}) - 0,015(\text{OPM/NPM}) -$

Tabel 4.7 Model GEVR dengan *Synthetic Features Generation* Tiap *Size* (Lanjutan).

Size	**Model**
	$0,006(\text{DER}\times\text{WCS})-0,296(\text{QR}/\text{CR})-0,004(\text{ETD}/\text{EBITA})-$ $8,181(\text{ETD}/\text{QR})-1,019(\text{WCA}\times\text{WCA})+0,002(\text{WCA}/\text{NPM})-$ $0,002(\text{WCS}/\text{OPM})-0,227(\text{bi7drr}/\text{Inflasi})$
1	$g(x)=19,150-15,940\text{EBITA}-0,507\text{STA}-0,001\text{ITR}-0,003\text{DSIR}-0,004\text{ROE}-$ $1,020\text{NPM}+1,066\text{OPM}+0,015\text{DER}+0,835\text{DAR}+0,035\text{CR}-0,009\text{QR}+$ $0,186\text{ETD}+0,231\text{WCA}-0,063\text{WCS}-0,756\text{rgdpg}+0,007\text{bi7drr}-$ $0,0002\text{USD/IDR}-0,090\text{Inflasi}-26,76(\text{EBITA}\times\text{EBITA})+$ $3,749(\text{EBITA}\times\text{DAR})+3,578(\text{EBITA}\times\text{WCA})-0,007(\text{EBITA}/\text{WCS})+$ $0,024(\text{STA}\times\text{STA})+1,807(\text{STA}/\text{ITR})-0,00002(\text{STA}/\text{WCS})-$ $0,031(\text{ITR}\times\text{DSIR})+0,0001(\text{ITR}\times\text{OPM})+14\times10^{-7}(\text{ITR}/\text{NPM})+$ $0,025(\text{ITR}/\text{QR})+0,0004(\text{DSIR}/\text{STA})-0,172(\text{ROE}\times\text{NPM})+$ $0,028(\text{ROE}/\text{QR})+0,046(\text{NPM}/\text{ROE})-0,0003(\text{OPM}/\text{NPM})+$ $0,058(\text{DER}\times\text{WCS})-0,374(\text{QR}/\text{CR})-0,013(\text{ETD}/\text{EBITA})+$ $0,2(\text{ETD}/\text{QR})-0,011(\text{WCA}\times\text{WCA})+0,0002(\text{WCA}/\text{NPM})-$ $0,0004(\text{WCS}/\text{OPM})-1,001(\text{bi7drr}/\text{Inflasi})$
2	$g(x)=6,225-11,160\text{EBITA}-0,237\text{STA}-0,0003\text{ITR}+0,0007\text{DSIR}+0,180\text{ROE}+$ $0,545\text{NPM}-1,913\text{OPM}+0,005\text{DER}+1,268\text{DAR}-0,150\text{CR}+0,365\text{QR}-$ $0,836\text{ETD}-0,923\text{WCA}-0,306\text{WCS}+0,063\text{rgdpg}+0,242\text{bi7drr}+$ $0,00032\text{USD/IDR}-0,221\text{Inflasi}-29,830(\text{EBITA}\times\text{EBITA})-$ $3,174(\text{EBITA}\times\text{DAR})-2,485(\text{EBITA}\times\text{WCA})+0,024(\text{EBITA}/\text{WCS})+$ $0,003(\text{STA}\times\text{STA})+0,815(\text{STA}/\text{ITR})-0,0006(\text{STA}/\text{WCS})-$ $0,021(\text{ITR}\times\text{DSIR})+0,0003(\text{ITR}\times\text{OPM})-0,00002(\text{ITR}/\text{NPM})+$ $0,003(\text{ITR}/\text{QR})-0,0002(\text{DSIR}/\text{STA})+1,029(\text{ROE}\times\text{NPM})+$ $0,033(\text{ROE}/\text{QR})+0,305(\text{NPM}/\text{ROE})+0,002(\text{OPM}/\text{NPM})-$ $0,00007(\text{DER}\times\text{WCS})-0,824(\text{QR}/\text{CR})-0,002(\text{ETD}/\text{EBITA})-$ $0,113(\text{ETD}/\text{QR})-0,683(\text{WCA}\times\text{WCA})+0,0003(\text{WCA}/\text{NPM})+$ $0,0002(\text{WCS}/\text{OPM})-1,058(\text{bi7drr}/\text{Inflasi})$
3	$g(x)=17,530-6,208\text{EBITA}-0,287\text{STA}-0,002\text{ITR}-0,008\text{DSIR}+0,020\text{ROE}+$ $1,227\text{NPM}-0,880\text{OPM}+0,007\text{DER}+0,528\text{DAR}-0,018\text{CR}+0,037\text{QR}+$ $0,401\text{ETD}+0,174\text{WCA}+0,025\text{WCS}-0,764\text{rgdpg}-0,089\text{bi7drr}-$ $0,0006\text{USD/IDR}-0,0421\text{Inflasi}-19,560(\text{EBITA}\times\text{EBITA})+$

Tabel 4.7 Model GEVR dengan *Synthetic Features Generation* Tiap *Size* (Lanjutan).

Size	Model
	$3,061(\text{EBITA}\times\text{DAR})-11,570(\text{EBITA}\times\text{WCA})+0,011(\text{EBITA}/\text{WCS})+$ $0,013(\text{STA}\times\text{STA})+2,295(\text{STA}/\text{ITR})-0,0002(\text{STA}/\text{WCS})-$ $0,012(\text{ITR}\times\text{DSIR})+0,00006(\text{ITR}\times\text{OPM})-0,00001(\text{ITR}/\text{NPM})+$ $0,023(\text{ITR}/\text{QR})+0,002(\text{DSIR}/\text{STA})+1,065(\text{ROE}\times\text{NPM})+$ $0,059(\text{ROE}/\text{QR})+0,018(\text{NPM}/\text{ROE})+0,001(\text{OPM}/\text{NPM})+$ $0,014(\text{DER}\times\text{WCS})-0,718(\text{QR}/\text{CR})-0,011(\text{ETD}/\text{EBITA})-$ $0,330(\text{ETD}/\text{QR})+0,452(\text{WCA}\times\text{WCA})-0,0009(\text{WCA}/\text{NPM})+$ $0,005(\text{WCS}/\text{OPM})-1,200(\text{bi7drr}/\text{Inflasi})$

Hasil estimasi parameter dan uji signifikansi parameter model *Generalized Extreme value Regression* (GEVR) pada tiap *size* ditampilkan pada Lampiran 2. Berdasarkan uji signifikansi parameter dapat diketahui bahwa tidak semua variabel signifikan dalam model. Variabel yang signifikan dalam model ditunjukkan pada Tabel 4.8.

Tabel 4.8 Variabel Signifikan Dalam Model GEVR Secara Serentak.

Size	Variabel
0	EBITA, STA, ROE, NPM, OPM, DAR, ETD, WCA, EBITA×EBITA, ITR×DSIR, ITR×OPM, ITR/NPM, ITR/QR, DSIR/STA, ROE×NPM, ROE/QR, OPM/NPM, ETD/QR, WCA×WCA, WCA/NPM
1	EBITA, STA, DAR, rgdpg, USD/IDR, EBITA×EBITA, EBITA/WCS, STA×STA, STA/ITR, ITR×DSIR, ITR/QR, ETD/EBITA, bi7drr/Inflasi
2	EBITA, DAR, QR, bi7drr, USD/IDR, Inflasi, EBITA×EBITA, EBITA/WCS, ITR×DSIR, bi7drr/Inflasi
3	DSIR, DAR, rgdpg, USD/IDR, EBITA×EBITA, EBITA×WCA, EBITA/WCS, STA/ITR, ITR×DSIR, ITR/QR, DSIR/STA, QR/CR, WCS/OPM, bi7drr/Inflasi

Berdasarkan model *Generalized Extreme Value Regression* (GEVR) pada *size* 0 dapat diketahui bahwa EBITA, STA, ROE, NPM, OPM, dan ETD berpengaruh signifikan dan bertanda negatif, artinya setiap kenaikan nilai EBITA, STA, ROE, NPM,

OPM, dan ETD akan menurunkan risiko perusahaan sektor industri di Indonesia mengalami kondisi *financial distress*. Sedangkan DAR dan WCA berpengaruh signifikan dan bertanda positif, artinya semakin tinggi nilai DAR dan WCA akan meningkatkan risiko perusahaan sektor industri di Indonesia mengalami *financial distress*. Berdasarkan Tabel 4.8 dapat diketahui bahwa indikator ekonomi makro tidak berpengaruh signifikan dalam model *Generalized Extreme Value Regression* (GEVR) secara serentak pada *size* 0. Hasil klasifikasi menggunakan model *Generalized Extreme Value Regression* (GEVR) dengan melibatkan *synthetic features generation* secara serentak pada tiap *size* ditunjukkan pada Tabel 4.9.

Tabel 4.9a *Confusion Matrix* Model GEVR dengan *Synthetic Features Generation* Untuk *Size* 0.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1130	83	0	285	19
1	53	329	1	19	112

Tabel 4.9b *Confusion Matrix* Model GEVR dengan *Synthetic Features Generation* Untuk *Size* 1.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1014	87	0	289	15
1	122	227	1	54	77

Tabel 4.9c *Confusion Matrix* Model GEVR dengan *Synthetic Features Generation* Untuk *Size* 2.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	818	170	0	161	143
1	66	251	1	6	125

Tabel 4.9d *Confusion Matrix* Model GEVR dengan *Synthetic Features Generation* Untuk *Size* 3.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	811	61	0	299	5
1	138	150	1	120	11

Confusion matrix pada Tabel 4.9 menunjukkan ketepatan hasil klasifikasi menggunakan model *Generalized Extreme Value Regression* (GEVR) secara serentak pada tiap *size*. Kemudian dilakukan evaluasi dengan menghitung *accuracy*, AUC, dan *F*1-*score* pada tiap *size* untuk mengetahui kebaikan model dalam memprediksi *financial distress* perusahaan sektor industri di Indonesia yang ditampilkan pada Tabel 4.10.

Tabel 4.10 Evaluasi Hasil Klasifikasi Model GEVR dengan *Synthetic Features Generation* Secara Serentak.

Size	**Ukuran**	***Training***	***Testing***
0	*Accuracy*	91,47%	91,26%
	AUC	89,64%	89,62%
	*F*1-*score*	94,32%	93,75%
1	*Accuracy*	85,59%	84,14%
	AUC	78,57%	76,92%
	*F*1-*score*	90,66%	89,34%
2	*Accuracy*	81,92%	65,75%
	AUC	80,99%	74,19%
	*F*1-*score*	87,39%	68,37%
3	*Accuracy*	82,84%	71,26%
	AUC	72,54%	53,38%
	*F*1-*score*	89,07%	82,71%

Berdasarkan evaluasi hasil klasifikasi pada model *Generalized Extreme Value Regression* (GEVR) yang ditunjukkan pada Tabel 4.10 dapat diketahui nilai *accuracy*, AUC, dan *F*1-*score* model *Generalized Extreme Value Regression* (GEVR) pada tiap *size*. Tingginya nilai *F*1-*score* menunjukkan kebaikan model tersebut dalam memprediksi kondisi perusahaan yang sehat. Secara serentak model *Generalized Extreme Value Regression* sangat baik dalam memprediksi kondisi *financial distress* pada *size* 0. Terdapat perbedaan yang cukup besar diantara kebaikan hasil klasifikasi data *training* dan data *testing* pada *size* 2 dan 3, hal ini mengindikasikan ada kecenderungan *overfitting*. Hal tersebut dapat terjadi karena pada model terdapat beberapa variabel yang tidak signifikan dalam model, sehingga perlu dilakukan seleksi variabel.

4.2.2 *Generalized Extreme Value Regression* (GEVR) dengan *Synthetic Features Generation* dan Seleksi Variabel

Kemudian dilakukan prediksi *financial distress* perusahaan sektor industri di Indonesia dengan model *Generalized Extreme Value Regression* (GEVR) menggunakan variabel yang signifikan dalam model dari hasil seleksi variabel dengan menggunakan prosedur *stepwise* yaitu dengan *backward elimination*. *Backward elimination* dimulai dengan model yang kompleks dan secara berurutan menghapus variabel yang memiliki *p-value* yang besar. Menggunakan nilai $\alpha = 0{,}1$, variabel yang terpilih dari prosedur *stepwise* ditunjukkan pada Tabel 4.11.

Tabel 4.11 Variabel Signifikan Dalam Model GEVR dengan *Stepwise*.

Size	Variabel	*Step*
0	EBITA, STA, DSIR, ROE, NPM, OPM, DAR, ETD, WCA, WCS, EBITA×EBITA, ITR×DSIR, ITR×OPM, ITR/NPM, ITR/QR, DSIR/STA, ROE×NPM, ROE/QR, OPM/NPM, ETD/QR, WCA/NPM	21
1	EBITA, STA, ITR, DSIR, NPM, OPM, DER, DAR, WCA, rgdpg, USD/IDR, Inflasi, EBITA×EBITA, EBITA×DAR, EBITA/WCS, STA×STA, STA/ITR, ITR×DSIR, ITR/QR, DSIR/STA, DER×WCS, ETD/EBITA, bi7drr/Inflasi	20
2	EBITA, STA, ROE, DAR, QR, ETD, WCA, USD/IDR, Inflasi, EBITA×EBITA, EBITA/WCS, STA/ITR, ITR×DSIR, WCA×WCA, bi7drr/Inflasi	27
3	EBITA, STA, ITR, DSIR, NPM, DER, DAR, CR, rgdpg, bi7drr, USD/IDR, EBITA×EBITA, EBITA×WCA, STA×STA, STA/ITR, ITR×DSIR, ITR/QR, DSI/STA, ROE×NPM, ROE/QR, QR/CR, ETD/QR, WCS/OPM, bi7drr/Inflasi	22

Berdasarkan variabel yang signifikan dalam model *Generalized Extreme Value Regression* (GEVR) dengan *stepwise*, maka dapat dibentuk fungsi *link* dari model *Generalized Extreme value Regression* (GEVR) pada tiap *size* yang ditampilkan pada Tabel 4.12.

Tabel 4.12 Model GEVR dengan *Synthetic Features Generation* dan Seleksi Variabel Tiap *Size*.

Size	**Model**
0	$g(x) = 16,45 - 37,95\text{EBITA} - 1,219\text{STA} - 0,004\text{DSIR} - 1,407\text{ROE} - 5,271\text{NPM} - 20,73\text{OPM} + 1,321\text{DAR} - 4,40\text{ETD} + 1,472\text{WCA} - 0,866\text{WCS} + 108,9(\text{EBITA} \times \text{EBITA}) - 0,038(\text{ITR} \times \text{DSIR}) - 0,542(\text{ITR} \times \text{OPM}) - 0,00008(\text{ITR}/\text{NPM}) + 0,087(\text{ITR}/\text{QR}) - 0,002(\text{DSIR}/\text{STA}) + 13,17(\text{ROE} \times \text{NPM}) + 0,564(\text{ROE}/\text{QR}) - 0,012(\text{OPM}/\text{NPM}) - 5,404(\text{ETD}/\text{QR}) - 0,001(\text{WCA}/\text{NPM})$
1	$g(x) = 18,52 - 16,91\text{EBITA} - 0,53\text{STA} - 0,0008\text{ITR} - 0,007\text{DSIR} + 1,2\text{OPM} + 0,017\text{DER} + 0,74\text{DAR} + 0,45\text{WCA} - 0,76\text{rgdpg} - 0,0002\text{USD/IDR} - 0,091\text{Inflasi} - 20,36(\text{EBITA} \times \text{EBITA}) + 3,53(\text{EBITA} \times \text{DAR}) - 0,009(\text{EBITA}/\text{WCS}) + 0,03(\text{STA} \times \text{STA}) + 2,85(\text{STA}/\text{ITR}) - 0,03(\text{ITR} \times \text{DSIR}) + 0,021(\text{ITR}/\text{QR}) + 0,001(\text{DSIR}/\text{STA}) + 0,05(\text{DER} \times \text{WCS}) - 0,01(\text{ETD}/\text{EBITA}) - 1,03(\text{bi7drr}/\text{Inflasi})$
2	$g(x) = 7,71 - 14,93\text{EBITA} - 0,12\text{STA} + 0,24\text{ROE} + 0,56\text{DAR} + 0,23\text{QR} - 0,54\text{ETD} - 1,30\text{WCA} + 0,0001\text{USD/IDR} - 0,05\text{Inflasi} - 30,89(\text{EBITA} \times \text{EBITA}) + 0,02(\text{EBITA}/\text{WCS}) + 1,19(\text{STA}/\text{ITR}) - 0,02(\text{ITR} \times \text{DSIR}) - 0,69(\text{WCA} \times \text{WCA}) - 0,70(\text{bi7drr}/\text{Inflasi})$
3	$g(x) = 17,46 - 13,18\text{EBITA} - 0,64\text{STA} - 0,001\text{ITR} - 0,01\text{DSIR} + 1,68\text{NPM} + 0,02\text{DER} + 0,66\text{DAR} - 0,54\text{rgdpg} - 0,13\text{bi7drr} - 0,001\text{USD/IDR} - 20,87(\text{EBITA} \times \text{EBITA}) - 11,62(\text{EBITA} \times \text{WCA}) + 0,03(\text{STA} \times \text{STA}) + 3,21(\text{STA}/\text{ITR}) - 0,02(\text{ITR} \times \text{DSIR}) + 0,02(\text{ITR}/\text{QR}) + 0,002(\text{DSIR}/\text{STA}) + 0,68(\text{ROE} \times \text{NPM}) + 0,06(\text{ROE}/\text{QR}) - 0,58(\text{QR}/\text{CR}) - 0,34(\text{ETD}/\text{QR}) + 0,004(\text{WCS}/\text{OPM}) - 1,13(\text{bi7drr}/\text{Inflasi})$

Hasil estimasi parameter dan uji signifikansi parameter model *Generalized Extreme value Regression* (GEVR) dengan seleksi variabel pada tiap *size* ditampilkan pada Lampiran 3. Variabel EBITA, STA, DSIR, ROE, NPM, OPM, ETD, dan WCS berpengaruh signifikan dan bertanda negatif dalam model *Generalized Extreme Value Regression* (GEVR) dengan *stepwise* pada *size* 0. Hal tersebut menunjukkan bahwa setiap kenaikan 1% EBITA, STA, DSIR, ROE, NPM, OPM, ETD, dan WCS akan menurunkan peluang perusahaan sektor industri di Indonesia

mengalami *financial distress*. Sedangkan pada variabel DAR dan WCA berpengaruh signifikan dan bertanda positif, artinya setiap kenaikan 1% DAR dan WCA akan meningkatkan peluang perusahaan sektor industri di Indonesia mengalami *financial distress*. Hasil klasifikasi model *Generalized Extreme Value Regression* dengan *synthetic features generation* dan seleksi variabel pada tiap *size* ditunjukkan pada Tabel 4.13.

Tabel 4.13a *Confusion Matrix* Model GEVR dengan *Synthetic Features Generation* dan Seleksi Variabel Untuk *Size* 0.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1145	68	0	285	19
1	39	343	1	16	115

Tabel 4.13b *Confusion Matrix* Model GEVR dengan *Synthetic Features Generation* dan Seleksi Variabel Untuk *Size* 1.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1014	87	0	289	15
1	108	241	1	53	78

Tabel 4.13c *Confusion Matrix* Model GEVR dengan *Synthetic Features Generation* dan Seleksi Variabel Untuk *Size* 2.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	844	144	0	227	77
1	85	232	1	20	111

Tabel 4.13d *Confusion Matrix* Model GEVR dengan *Synthetic Features Generation* dan Seleksi Variabel Untuk *Size* 3.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	832	40	0	302	2
1	169	119	1	117	14

Berdasarkan *confusion matrix* pada Tabel 4.13 dapat dilihat ketepatan hasil prediksi *financial distress* perusahaan sektor industri di Indonesia menggunakan model *Generalized Extreme Value Regression* (GEVR) dengan *synthetic features generation* dan seleksi variabel pada tiap *size*. Selanjutnya dilakukan evaluasi

hasil prediksi dari model tersebut untuk mengetahui kebaikan model dengan menghitung nilai *accuracy*, AUC, dan *F*1-*score* pada tiap *size*. Hasil evaluasi model *Generalized Extreme Value Regression* (GEVR) dengan *synthetic features generation* dan seleksi variabel ditampilkan pada Tabel 4.14.

Tabel 4.14 Evaluasi Hasil Klasifikasi Model GEVR dengan *Synthetic Features Generation* dan Seleksi Variabel.

Size	**Ukuran**	***Training***	***Testing***
0	*Accuracy*	93,29%	91,95%
	AUC	92,09%	90,77%
	*F*1-*score*	95,54%	94,21%
1	*Accuracy*	86,55%	84,37%
	AUC	80,58%	77,30%
	*F*1-*score*	91,23%	89,47%
2	*Accuracy*	82,45%	77,70%
	AUC	79,31%	79,70%
	*F*1-*score*	88,05%	82,40%
3	*Accuracy*	81,98%	72,64%
	AUC	68,37%	55,01%
	*F*1-*score*	88,84%	83,54%

Tabel 4.14 menunjukkan bahwa terjadi peningkatan kebaikan model *Generalized Extreme Value Regression* (GEVR) setelah dilakukan seleksi variabel. Berdasarkan Tabel 4.14 dapat diketahui bahwa setiap kenaikan *size* cenderung menurunkan nilai performansi model *Generalized Extreme Value Regression* (GEVR) dengan *synthetic features generation* dan seleksi variabel. Nilai *accuracy*, AUC, dan *F*1-*score* terbaik ialah model pada *size* 0. Terdapat perbedaan yang cukup jauh antara nilai AUC dan *F*1-*score* dari hasil evaluasi model pada *size* 3, hal tersebut menunjukkan bahwa model cenderung tepat dalam memprediksi perusahaan yang dalam kondisi sehat dibandingkan memprediksi perusahaan dalam kondisi *financial distress*.

4.2.3 *Logistic Regression* dengan *Synthetic Features Generation* Secara Serentak

Model *Logistic Regression* dibentuk melalui data *training* pada tiap *size*, dimana hasil estimasi parameter dan pengujian signifikansi ditampilkan pada Lampiran 4. Fungsi *link* dari model

Logistic Regression dengan *synthetic features generation* secara serentak ditunjukkan pada Tabel 4.15.

Tabel 4.15 Model *Logistic Regression* dengan *Synthetic Features Generation* Secara Serentak.

Size	**Model**
0	$g(x) = 3,63\times10^{15} - 9,04\times10^{15}\,\text{EBITA} - 2,53\times10^{14}\,\text{STA} + 3,61\times10^{11}\,\text{ITR} -$ $1,59\times10^{12}\,\text{DSIR} - 8,22\times10^{13}\,\text{ROE} - 1,57\times10^{15}\,\text{NPM} - 4,12\times10^{15}\,\text{OPM}$ $-5,80\times10^{12}\,\text{DER} + 2,81\times10^{14}\,\text{DAR} - 3,24\times10^{13}\,\text{CR} + 3,98\times10^{13}\,\text{QR} -$ $1,60\times10^{14}\,\text{ETD} - 3,17\times10^{13}\,\text{WCA} - 5,31\times10^{13}\,\text{WCS} + 4,45\times10^{13}\,\text{rgdpg} +$ $3,28\times10^{13}\,\text{bi7drr} + 1,95\times10^{9}\,\text{USD/IDR} + 1,23\times10^{11}\,\text{Inflasi} +$ $1,55\times10^{16}\left(\text{EBITA}\times\text{EBITA}\right) + 1,94\times10^{15}\left(\text{EBITA}\times\text{DAR}\right) +$ $3,49\times10^{15}\left(\text{EBITA}\times\text{WCA}\right) - 3,29\times10^{12}\left(\text{EBITA}/\text{WCS}\right) +$ $9,43\times10^{12}\left(\text{STA}\times\text{STA}\right) + 5,49\times10^{14}\left(\text{STA}/\text{ITR}\right) + 7\times10^{10}\left(\text{STA}/\text{WCS}\right)$ $-9,36\times10^{12}\left(\text{ITR}\times\text{DSIR}\right) + 3,61\times10^{11}\left(\text{ITR}\times\text{OPM}\right) -$ $2,44\times10^{10}\left(\text{ITR}/\text{NPM}\right) + 6,28\times10^{12}\left(\text{ITR}/\text{QR}\right) + 5,59\times10^{11}\left(\text{DSIR}/\text{STA}\right)$ $+2,92\times10^{15}\left(\text{ROE}\times\text{NPM}\right) + 8,84\times10^{13}\left(\text{ROE}/\text{QR}\right) -$ $3,01\times10^{13}\left(\text{NPM}/\text{ROE}\right) - 3,35\times10^{12}\left(\text{OPM}/\text{NPM}\right) +$ $4,02\times10^{12}\left(\text{DER}\times\text{WCS}\right) - 9,69\times10^{13}\left(\text{QR}/\text{CR}\right) -$ $3,31\times10^{11}\left(\text{ETD}/\text{EBITA}\right) - 1,24\times10^{15}\left(\text{ETD}/\text{QR}\right) -$ $1,61\times10^{14}\left(\text{WCA}\times\text{WCA}\right) + 1,67\times10^{11}\left(\text{WCA}/\text{NPM}\right) +$ $1,88\times10^{11}\left(\text{WCS}/\text{OPM}\right) - 2,63\times10^{13}\left(\text{bi7drr}/\text{Inflasi}\right)$
1	$g(x) = 20,55 - 23,23\text{EBITA} - 0,85\text{STA} - 0,001\text{ITR} - 0,003\text{DSIR} + 0,05\text{ROE} -$ $0,46\text{NPM} + 0,85\text{OPM} + 0,04\text{DER} + 1,44\text{DAR} + 0,02\text{CR} + 0,03\text{QR} -$ $0,23\text{ETD} + 0,24\text{WCA} - 0,10\text{WCS} - 1,11\text{rgdpg} + 0,08\text{bi7drr} -$ $0,0002\text{USD/IDR} - 0,211\text{Inflasi} - 29,76\left(\text{EBITA}\times\text{EBITA}\right) +$ $2,31\left(\text{EBITA}\times\text{DAR}\right) + 0,47\left(\text{EBITA}\times\text{WCA}\right) - 0,002\left(\text{EBITA}/\text{WCS}\right) +$ $0,05\left(\text{STA}\times\text{STA}\right) + 1,53\left(\text{STA}/\text{ITR}\right) - 0,0001\left(\text{STA}/\text{WCS}\right) -$ $0,02\left(\text{ITR}\times\text{DSIR}\right) + 0,0002\left(\text{ITR}\times\text{OPM}\right) - 0,000001\left(\text{ITR}/\text{NPM}\right) +$ $0,03\left(\text{ITR}/\text{QR}\right) + 0,0001\left(\text{DSIR}/\text{STA}\right) - 0,09\left(\text{ROE}\times\text{NPM}\right) +$ $0,03\left(\text{ROE}/\text{QR}\right) + 0,10\left(\text{NPM}/\text{ROE}\right) + 0,0001\left(\text{OPM}/\text{NPM}\right) +$ $0,09\left(\text{DER}\times\text{WCS}\right) - 0,48\left(\text{QR}/\text{CR}\right) - 0,02\left(\text{ETD}/\text{EBITA}\right) +$ $0,30\left(\text{ETD}/\text{QR}\right) - 0,51\left(\text{WCA}\times\text{WCA}\right) + 0,0004\left(\text{WCA}/\text{NPM}\right) +$ $0,00005\left(\text{WCS}/\text{OPM}\right) - 2,05\left(\text{bi7drr}/\text{Inflasi}\right)$

Tabel 4.15 Model *Logistic Regression* dengan *Synthetic Features Generation* Secara Serentak (Lanjutan).

Size	**Model**
2	$g(x) = 1,52\times10^{15} - 2,72\times10^{15}\,\text{EBITA} - 3,39\times10^{14}\,\text{STA} - 3,05\times10^{11}\,\text{ITR} -$ $2,02\times10^{11}\,\text{DSIR} - 1,95\times10^{14}\,\text{ROE} + 5,46\times10^{14}\,\text{NPM} - 2,87\times10^{15}\,\text{OPM} +$ $6,64\times10^{12}\,\text{DER} + 5,51\times10^{14}\,\text{DAR} - 8,67\times10^{13}\,\text{CR} + 1,63\times10^{14}\,\text{QR} -$ $2,39\times10^{14}\,\text{ETD} - 5,18\times10^{14}\,\text{WCA} - 1,72\times10^{14}\,\text{WCS} + 1,43\times10^{14}\,\text{rgdpg} +$ $1,47\times10^{14}\,\text{bi7drr} - 3,16\times10^{11}\,\text{USD/IDR} - 1,18\times10^{14}\,\text{Inflasi} -$ $6,30\times10^{15}\left(\text{EBITA}\times\text{EBITA}\right) - 1,04\times10^{15}\left(\text{EBITA}\times\text{DAR}\right) +$ $2,60\times10^{15}\left(\text{EBITA}\times\text{WCA}\right) + 5,75\times10^{12}\left(\text{EBITA}/\text{WCS}\right) +$ $1,96\times10^{13}\left(\text{STA}\times\text{STA}\right) + 1,14\times10^{15}\left(\text{STA}/\text{ITR}\right) - 3,02\times10^{11}\left(\text{STA}/\text{WCS}\right) -$ $1,35\times10^{13}\left(\text{ITR}\times\text{DSIR}\right) + 1,81\times10^{11}\left(\text{ITR}\times\text{OPM}\right) +$ $7,25\times10^{9}\left(\text{ITR}/\text{NPM}\right) + 8,68\times10^{11}\left(\text{ITR}/\text{QR}\right) - 7,64\times10^{10}\left(\text{DSIR}/\text{STA}\right) +$ $6,19\times10^{14}\left(\text{ROE}\times\text{NPM}\right) + 3,17\times10^{13}\left(\text{ROE}/\text{QR}\right) +$ $1,34\times10^{14}\left(\text{NPM}/\text{ROE}\right) - 2,30\times10^{12}\left(\text{OPM}/\text{NPM}\right) -$ $1,37\times10^{13}\left(\text{DER}\times\text{WCS}\right) - 1,19\times10^{14}\left(\text{QR}/\text{CR}\right) - 3,84\times10^{11}\left(\text{ETD}/\text{EBITA}\right)$ $-2,33\times10^{14}\left(\text{ETD}/\text{QR}\right) - 4,28\times10^{14}\left(\text{WCA}\times\text{WCA}\right) +$ $2,05\times10^{11}\left(\text{WCA}/\text{NPM}\right) + 2,93\times10^{10}\left(\text{WCS}/\text{OPM}\right) -$ $4,96\times10^{14}\left(\text{bi7drr}/\text{Inflasi}\right)$
3	$g(x) = 17,63 - 11,08\text{EBITA} - 0,63\text{STA} - 0,001\text{ITR} - 0,01\text{DSIR} -$ $0,10\text{ROE} + 1,72\text{NPM} - 2,20\text{OPM} + 0,02\text{DER} + 0,58\text{DAR} -$ $0,04\text{CR} + 0,05\text{QR} + 0,29\text{ETD} - 0,26\text{WCA} + 0,02\text{WCS} -$ $0,49\text{rgdpg} - 0,22\text{bi7drr} - 0,001\text{USD/IDR} + 0,07\text{Inflasi} -$ $21,57\left(\text{EBITA}\times\text{EBITA}\right) + 0,10\left(\text{EBITA}\times\text{DAR}\right) -$ $12,58\left(\text{EBITA}\times\text{WCA}\right) + 0,02\left(\text{EBITA}/\text{WCS}\right) +$ $0,03\left(\text{STA}\times\text{STA}\right) + 2,80\left(\text{STA}/\text{ITR}\right) - 0,0003\left(\text{STA}/\text{WCS}\right) -$ $0,02\left(\text{ITR}\times\text{DSIR}\right) + 0,0005\left(\text{ITR}\times\text{OPM}\right) -$ $0,00001\left(\text{ITR}/\text{NPM}\right) + 0,02\left(\text{ITR}/\text{QR}\right) + 0,002\left(\text{DSIR}/\text{STA}\right) +$ $0,86\left(\text{ROE}\times\text{NPM}\right) + 0,07\left(\text{ROE}/\text{QR}\right) + 0,03\left(\text{NPM}/\text{ROE}\right) -$ $0,002\left(\text{OPM}/\text{NPM}\right) - 0,002\left(\text{DER}\times\text{WCS}\right) - 0,73\left(\text{QR}/\text{CR}\right) -$ $0,002\left(\text{ETD}/\text{EBITA}\right) - 0,45\left(\text{ETD}/\text{QR}\right) + 0,07\left(\text{WCA}\times\text{WCA}\right) -$ $0,0002\left(\text{WCA}/\text{NPM}\right) + 0,003\left(\text{WCS}/\text{OPM}\right) - 76\left(\text{bi7drr}/\text{Inflasi}\right)$

Hasil estimasi parameter dan pengujian signifikansi parameter ditampilkan pada Lampiran 4. Berdasarkan hasil pengujian signifikansi parameter dapat diketahui bahwa pada model *Logistic Regression* dengan *synthetic features generation* secara serentak terdapat variabel yang tidak signifikan dalam model. Variabel-variabel yang signifikan dalam model pada tiap *size* ditunjukkan pada Tabel 4.16.

Tabel 4.16 Variabel Signifikan Dalam Model *Logistic Regression* Secara Serentak.

Size	**Variabel**
0	EBITA, STA, ITR, DSIR, ROE, NPM, OPM, DER, DAR, CR, QR, ETD, WCA, WCS, rgdpg, bi7drr, USD/IDR, Inflasi, EBITA×EBITA, EBITA×DAR, EBITA×WCA,EBITA×WCS, STA×STA, STA×ITR, STA×WCS, ITR×DSIR, ITR×OPM, ITR/NPM, ITR/QR, DSIR/STA, ROE×NPM, ROE/QR, NPM/ROE, OPM/NPM, DER×WCS, QR/CR, ETD/EBITA, ETD/QR, WCA×WCA, WCA/NPM, WCS/OPM, bi7drr/Inflasi
1	EBITA, STA, DER, DAR, rgdpg, Inflasi, EBITA×EBITA, STA×STA, ITR×DSIR, ITR/QR, ETD/EBITA, bi7drr/Inflasi
2	EBITA, STA, ITR, DSIR, ROE, NPM, OPM, DER, DAR, CR, QR, ETD, WCA, WCS, rgdpg, bi7drr, USD/IDR, Inflasi, EBITA×EBITA, EBITA×DAR, EBITA×WCA,EBITA×WCS, STA×STA, STA×ITR, STA×WCS, ITR×DSIR, ITR×OPM, ITR/NPM, ITR/QR, DSIR/STA, ROE×NPM, ROE/QR, NPM/ROE, OPM/NPM, DER×WCS, QR/CR, ETD/EBITA, ETD/QR, WCA×WCA, WCA/NPM, WCS/OPM, bi7drr/Inflasi
3	EBITA, STA, DSIR, DAR, USD/IDR, EBITA×EBITA, STA×ITR, ITR×DSIR, DSIR/STA, WCS/OPM

Berdasarkan Tabel 4.16 dapat dilihat bahwa pada *size* 0 dan *size* 2 memiliki model dengan seluruh variabelnya signifikan dalam model. Variabel EBITA, STA, DSIR, ROE, NPM, OPM, DER, CR, ETD, WCA, dan WCS memiliki koefisien yang bertanda negatif dalam model *Logistic Regression* secara serentak. Hal tersebut menunjukkan bahwa setiap kenaikan 1% pada tiap variabel tersebut akan menurunkan peluang perusahaan sektor industri di Indonesia mengalami *financial distress*. Sedangkan

pada variabel ITR, DAR, QR, *real* GDP *growth*, BI 7-*day repo rate*, nilai tukar dollar, dan inflasi memiliki koefisien yang bertanda positif, artinya setiap kenaikan variabel-variabel tersebut akan meningkatkan peluang perusahaan mengalami *financial distress*. Setelah mendapatkan model *Logistic Regression* dari data *training*, kemudian dapat dilakukan klasifikasi menggunakan data *training* dan data *testing* pada tiap *size* untuk mengetahui kebaikan model dalam memprediksi *financial distress* perusahaan sektor industri di Indonesia. Hasil klasifikasi model *Logistic Regression* dengan *synthetic features generation* secara serentak pada tiap *size* ditunjukkan pada Tabel 4.17.

Tabel 4.17a *Confusion Matrix* Model *Logistic Regression* dengan *Synthetic Features Generation* Secara Serentak Untuk *Size* 0.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1076	137	0	274	30
1	13	369	1	7	124

Tabel 4.17b *Confusion Matrix* Model *Logistic Regression* dengan *Synthetic Features Generation* Secara Serentak Untuk *Size* 1.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1038	63	0	295	9
1	146	203	1	66	65

Tabel 4.17c *Confusion Matrix* Model *Logistic Regression* dengan *Synthetic Features Generation* Secara Serentak Untuk *Size* 2.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	810	178	0	69	235
1	88	229	1	5	126

Tabel 4.17d *Confusion Matrix* Model *Logistic Regression* dengan *Synthetic Features Generation* Secara Serentak Untuk *Size* 3.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	827	45	0	299	5
1	149	139	1	114	17

Berdasarkan *confusion matrix* pada Tabel 4.17 dapat diketahui ketepatan hasil prediksi *financial distress* perusahaan sektor industri di Indonesia menggunakan model *Logistic Regression* dengan *synthetic features generation* secara serentak pada tiap *size*. Selanjutnya dilakukan evaluasi hasil prediksi dari model tersebut dengan menghitung nilai *accuracy*, AUC, dan *F*1-*score* pada tiap *size* untuk mengetahui seberapa besar kebaikan model *Logistic Regression* dalam memprediksi kondisi *financial distress* perusahaan sektor industri di Indonesia yang ditampilkan pada Tabel 4.18.

Tabel 4.18 Evaluasi Hasil Klasifikasi Model *Logistic Regression* dengan *Synthetic Features Generation* Secara Serentak.

Size	**Ukuran**	***Training***	***Testing***
0	*Accuracy*	90,60%	91,49%
	AUC	92,65%	92,39%
	*F*1-*score*	93,48%	93,68%
1	*Accuracy*	85,59%	82,76%
	AUC	76,22%	73,32%
	*F*1-*score*	90,85%	88,72%
2	*Accuracy*	79,62%	44,83%
	AUC	77,11%	59,44%
	*F*1-*score*	85,90%	36,51%
3	*Accuracy*	83,28%	72,64%
	AUC	71,55%	55,67%
	*F*1-*score*	89,50%	83,40%

Tabel 4.18 menunjukkan bahwa berdasarkan nilai *accuracy*, AUC, dan *F*1-*score*, model *Logistic Regression* dengan *synthetic features generation* secara serentak pada *size* 0 sangat baik dalam memprediksi kondisi *financial distress* perusahaan sektor industri di Indonesia. Model prediksi pada *size* 2 menunjukkan perbandingan yang sangat jauh antara hasil prediksi dengan data *training* dan data *testing*, sehingga ada kecenderungan terjadi *overfitting* pada model *Logistic Regression* dengan *synthetic features generation* pada *size* 2. Salah satu solusi dalam mengatasi model yang *overfitting* dapat dilakukan dengan mengeliminasi variabel-variabel yang memiliki korelasi yang cukup tinggi dan mengeliminasi variabel yang kurang berpengaruh dalam model.

4.2.4 *Logistic Regression* dengan *Synthetic Features Generation* dan Seleksi Variabel

Rendahnya hasil klasifikasi pada model *Logistic Regression* secara serentak dapat disebabkan karena terdapat variabel yang tidak signifikan dalam model, sehingga perlu dilakukan seleksi variabel dengan menggunakan prosedur *stepwise*. Menggunakan prosedur *stepwise* akan didapatkan model yang optimum dalam memprediksi kondisi *financial distress* perusahaan sektor industri di Indonesia pada tiap *size*. Variabel yang terpilih dari hasil prosedur *stepwise* ditunjukkan pada Tabel 4.19.

Tabel 4.19 Variabel Signifikan Dalam Model *Logistic Regression* dengan *Stepwise*.

Size	**Variabel**	***Step***
0	EBITA, STA, ITR, DSIR, ROE, NPM, OPM, DER, DAR, CR, QR, ETD, WCA, WCS, rgdpg, bi7drr, USD/IDR, Inflasi, EBITA×EBITA, EBITA×DAR, EBITA×WCA, EBITA×WCS, STA×STA, STA×ITR, STA×WCS, ITR×DSIR, ITR×OPM, ITR/NPM, ITR/QR, DSIR/STA, ROE×NPM, ROE/QR, NPM/ROE, OPM/NPM, DER×WCS, QR/CR, ETD/EBITA, ETD/QR, WCA×WCA, WCA/NPM, WCS/OPM, bi7drr/Inflasi	0
1	EBITA, STA, ITR, DER, DAR, QR, rgdpg, USD/IDR, Inflasi, EBITA×EBITA, ITR×DSIR, ITR×OPM, ITR/QR, NPM/ROE, DER×WCS, ETD/EBITA, ETD/QR, WCA×WCA, bi7drr/Inflasi	23
2	EBITA, STA, DAR, CR, QR, ETD, WCA, WCS, bi7drr, USD/IDR, Inflasi, EBITA×EBITA, EBITA×WCS, STA×ITR, STA×WCS, ITR×DSIR, DSIR/STA, ROE×NPM, ROE/QR, NPM/ROE, QR/CR, WCA×WCA, bi7drr/Inflasi	19
3	EBITA, STA, ITR, DSIR, DAR, QR, ETD, USD/IDR, EBITA×EBITA, EBITA×WCA, STA×ITR, ITR×DSIR, ITR/QR, DSIR/STA, QR/CR, ETD/EBITA, WCS/OPM, bi7drr/Inflasi	24

Berdasarkan variabel yang signifikan dalam model *Logistic Regression* dengan *stepwise*, maka dapat dibentuk fungsi *link* dari

model *Logistic Regression* pada tiap *size* yang ditampilkan pada Tabel 4.20.

Tabel 4.20 Model *Logistic Regression* dengan *Synthetic Features Generation* dan Seleksi Variabel.

Size	**Model**
0	$g(x) = 3,63\times10^{15} - 9,04\times10^{15}\,\text{EBITA} - 2,53\times10^{14}\,\text{STA} + 3,61\times10^{11}\,\text{ITR} -$ $1,59\times10^{12}\,\text{DSIR} - 8,22\times10^{13}\,\text{ROE} - 1,57\times10^{15}\,\text{NPM} - 4,12\times10^{15}\,\text{OPM}$ $-5,80\times10^{12}\,\text{DER} + 2,81\times10^{14}\,\text{DAR} - 3,24\times10^{13}\,\text{CR} + 3,98\times10^{13}\,\text{QR} -$ $1,60\times10^{14}\,\text{ETD} - 3,17\times10^{13}\,\text{WCA} - 5,31\times10^{13}\,\text{WCS} + 4,45\times10^{13}\,\text{rgdpg} +$ $3,28\times10^{13}\,\text{bi7drr} + 1,95\times10^{9}\,\text{USD/IDR} + 1,23\times10^{11}\,\text{Inflasi} +$ $1,55\times10^{16}(\text{EBITA}\times\text{EBITA}) + 1,94\times10^{15}(\text{EBITA}\times\text{DAR}) +$ $3,49\times10^{15}(\text{EBITA}\times\text{WCA}) - 3,29\times10^{12}(\text{EBITA}/\text{WCS}) +$ $9,43\times10^{12}(\text{STA}\times\text{STA}) + 5,49\times10^{14}(\text{STA}/\text{ITR}) + 7\times10^{10}(\text{STA}/\text{WCS})$ $-9,36\times10^{12}(\text{ITR}\times\text{DSIR}) + 3,61\times10^{11}(\text{ITR}\times\text{OPM}) -$ $2,44\times10^{10}(\text{ITR}/\text{NPM}) + 6,28\times10^{12}(\text{ITR}/\text{QR}) + 5,59\times10^{11}(\text{DSIR}/\text{STA})$ $+2,92\times10^{15}(\text{ROE}\times\text{NPM}) + 8,84\times10^{13}(\text{ROE}/\text{QR}) -$ $3,01\times10^{13}(\text{NPM}/\text{ROE}) - 3,35\times10^{12}(\text{OPM}/\text{NPM}) +$ $4,02\times10^{12}(\text{DER}\times\text{WCS}) - 9,69\times10^{13}(\text{QR}/\text{CR}) - 3,31\times10^{11}(\text{ETD}/\text{EBITA})$ $-1,24\times10^{15}(\text{ETD}/\text{QR}) - 1,61\times10^{14}(\text{WCA}\times\text{WCA}) + 1,67\times10^{11}(\text{WCA}/\text{NPM})$ $+1,88\times10^{11}(\text{WCS}/\text{OPM}) - 2,63\times10^{13}(\text{bi7drr}/\text{Inflasi})$
1	$g(x) = 19,72 - 21\text{EBITA} - 0,42\text{STA} - 0,001\text{ITR} + 0,04\text{DER} +$ $1,63\text{DAR} + 0,04\text{QR} - 1,06\text{rgdpg} - 0,0002\text{USD/IDR} - 0,16\text{Inflasi} -$ $28,15(\text{EBITA}\times\text{EBITA}) - 0,02(\text{ITR}\times\text{DSIR}) + 0,0002(\text{ITR}\times\text{OPM}) -$ $0,02(\text{ITR}/\text{QR}) - 0,11(\text{NPM}/\text{ROE}) + 0,11(\text{DER}\times\text{WCS}) -$ $0,016(\text{ETD}/\text{EBITA}) + 0,3(\text{ETD}/\text{QR}) - 0,66(\text{WCA}\times\text{WCA}) -$ $1,82(\text{bi7drr}/\text{Inflasi})$
2	$g(x) = 6,99 - 15,93\text{EBITA} - 0,19\text{STA} + 1,15\text{DAR} - 0,13\text{CR} + 0,32\text{QR} -$ $0,69\text{ETD} - 0,94\text{WCA} - 0,29\text{WCS} + 0,24\text{bi7drr} + 0,0001\text{USD/IDR} -$ $0,23\text{Inflasi} - 28,6(\text{EBITA}\times\text{EBITA}) - 0,03(\text{EBITA}/\text{WCS}) +$ $0,89(\text{STA}/\text{ITR}) - 0,001(\text{STA}/\text{WCS}) - 0,02(\text{ITR}\times\text{DSIR}) -$ $0,0002(\text{DSIR}/\text{STA}) + 0,92(\text{ROE}\times\text{NPM}) + 0,03(\text{ROE}/\text{QR}) +$

Tabel 4.20 Model *Logistic Regression* dengan *Synthetic Features Generation* dan Seleksi Variabel (Lanjuran).

Size	**Model**
	$0,29(\mathrm{NPM/ROE}) - 0,78(\mathrm{QR/CR}) - 0,61(\mathrm{WCA} \times \mathrm{WCA})$ $-1,11(\mathrm{bi7drr/Inflasi})$
3	$g(x) = 10,41 - 13,28\mathrm{EBITA} - 0,28\mathrm{STA} - 0,001\mathrm{ITR} - 0,006\mathrm{DSIR} +$ $0,73\mathrm{DAR} + 0,02\mathrm{QR} + 0,60\mathrm{ETD} - 0,0003\mathrm{USD/IDR} -$ $20,43(\mathrm{EBITA} \times \mathrm{EBITA}) - 12,68(\mathrm{EBITA} \times \mathrm{WCA}) +$ $1,98(\mathrm{STA/ITR}) - 0,02\ (\mathrm{ITR} \times \mathrm{DSIR}) + 0,02(\mathrm{ITR/QR}) +$ $0,002(\mathrm{DSIR/STA}) - 0,81(\mathrm{QR/CR}) - 0,004(\mathrm{ETD/EBITA})$ $+0,003(\mathrm{WCS/OPM}) - 0,61(\mathrm{bi7drr/Inflasi})$

Hasil estimasi parameter dan pengujian signifikansi ditampilkan pada Lampiran 5. Setelah mendapatkan model *Logistic Regression* dengan *synthetic feattures generation* dan seleksi variabel, maka dapat dilakukan prediksi kondisi keuangan perusahaan sektor industri di Indonesia pada tiap *size*. Hasil klasifikasi ditunjukkan pada *confusion matrix* yang ditampilkan pada Tabel 4.21.

Tabel 4.21a *Confusion Matrix* Model *Logistic Regression* dengan *Synthetic Features Generation* dan Seleksi Variabel Untuk *Size* 0.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1076	137	0	274	30
1	13	369	1	7	124

Tabel 4.21b *Confusion Matrix* Model *Logistic Regression* dengan *Synthetic Features Generation* dan Seleksi Variabel Untuk *Size* 1.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1037	64	0	295	9
1	152	197	1	67	64

Tabel 4.21c *Confusion Matrix* Model *Logistic Regression* dengan *Synthetic Features Generation* dan Seleksi Variabel Untuk *Size* 2.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	941	47	0	258	46
1	150	167	1	34	97

Tabel 4.21d *Confusion Matrix* Model *Logistic Regression* dengan *Synthetic Features Generation* dan Seleksi Variabel Untuk *Size* 3.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	821	51	0	298	6
1	155	133	1	107	24

Hasil klasifikasi yang ditampilkan pada Tabel 4.21 kemudian dapat dilakukan perhitungan nilai *accuracy*, AUC, dan *F*1-*score* untuk mengevaluasi kebaikan model *Logistic Regression* dengan *synthetic features generation* dan seleksi variabel dengan prosedur *stepwise*. Hasil perhitungan tersebut dapat menjadi perbandingan dengan hasil kebaikan model secara serentak. Hasil evaluasi model *Logistic Regression* dengan *synthetic features generation* dan seleksi variabel ditunjukkan pada Tabel 4.22.

Tabel 4.22 Evaluasi Hasil Klasifikasi Model *Logistic Regression* dengan *Synthetic Features Generation* dan Seleksi Variabel.

Size	**Ukuran**	***Training***	***Testing***
0	*Accuracy*	90,60%	91,49%
	AUC	92,65%	92,39%
	*F*1-*score*	93,48%	93,68%
1	*Accuracy*	85,10%	82,53%
	AUC	75,32%	72,95%
	*F*1-*score*	90,57%	88,59%
2	*Accuracy*	84,90%	81,61%
	AUC	73,96%	79,46%
	*F*1-*score*	90,52%	86,58%
3	*Accuracy*	82,24%	74,02%
	AUC	70,17%	58,17%
	*F*1-*score*	88,85%	84,06%

Berdasarkan evaluasi hasil klasifikasi model *Logistic Regression* dengan *synthetic features generation* dan seleksi variabel pada Tabel 4.22 dapat dilihat bahwa setiap kenaikan *size* menurunkan performansi model dalam memprediksi kondisi keuangan perusahaan sektor industri di Indonesia. Seleksi variabel dengan prosedur *stepwise* mengakibatkan perubahan yang cukup signifikan pada kebaikan model *size* 2 dan masalah *overfitting* telah diatasi. Tingginya nilai *F*1-*score* dibandingkan nilai AUC

menunjukkan kebaikan model tersebut dalam memprediksi kondisi perusahaan yang sehat.

4.2.5 *Lasso Logistic Regression* dengan *Synthetic Features Generation*

Metode *Lasso Logistic Regression* merupakan metode untuk mengeliminasi variabel yang tidak relevan dengan mengecilkan nilai koefisiennya menjadi 0. Pada penelitian ini nilai λ yang digunakan merupakan nilai λ optimum yang memberikan ketepatan klasifikasi terbaik. Nilai koefisien pada model *Lasso Logistic Regression* untuk tiap *size* ditampilkan pada Lampiran 6. Adapun variabel yang terpilih dalam model *Lasso Logistic Regression* ditunjukkan pada Tabel 4.23.

Tabel 4.23 Variabel Model *Lasso Logistic Regression* dengan *Synthetic Features Generation*.

Size	**Variabel**
0	EBITA, STA, ITR, NPM, OPM, DER, DAR, QR, ETD, WCA, rgdpg, bi7drr, USD/IDR, EBITA×EBITA, EBITA×DAR, EBITA×WCS, STA×WCS, ITR×DSIR, ITR/NPM, ITR/QR, DSIR/STA, NPM/ROE, OPM/NPM, DER×WCS, QR/CR, ETD/QR, WCA×WCA, WCA/NPM, WCS/OPM
1	EBITA, STA, ITR, DAR, CR, ETD, USD/IDR, EBITA×WCS, STA×WCS, ITR×DSIR, ITR×OPM, ITR/QR, NPM/ROE, DER×WCS, QR/CR, bi7drr/Inflasi
2	EBITA, STA, DAR, QR, ETD, WCA, USD/IDR, EBITA×WCA, EBITA×WCS, STA×WCS, ITR×DSIR, ITR/QR, ROE×NPM, ROE/QR, NPM/ROE
3	EBITA, OPM, DER, DAR, QR, WCA, bi7drr, USD/IDR, EBITA×EBITA, EBITA×WCA, EBITA×WCS, STA×WCS, ITR×DSIR, ITR/QR, DSIR/STA, ROE×NPM, NPM/ROE, OPM/NPM, QR/CR, ETD/EBITA, WCS/OPM, bi7drr/Inflasi

Menggunakan variabel yang telah ditampilkan pada Tabel 4.23 pada tiap *size*, maka dapat dibentuk fungsi *link* dari model *Lasso Logistic Regression* yang ditampilkan pada Tabel 4.24.

Tabel 4.24 Model *Lasso Logistic Regression* dengan *Synthetic Features Generation*.

Size	**Model**
0	$g(x) = 14,39 - 54,59\text{EBITA} - 0,502\text{STA} + 0,004\text{ITR} - 6,20\text{NPM} - 8,03\text{OPM} + 0,003\text{DER} + 0,90\text{DAR} + 0,001\text{QR} - 0,98\text{ETD} - 1,15\text{WCA} + 0,05\text{rgdpg} + 0,06\text{bi7drr} + 0,0001\text{USD/IDR} + 57,46(\text{EBITA} \times \text{EBITA}) + 7,95(\text{EBITA} \times \text{DAR}) - 0,001(\text{EBITA/WCS}) - 0,0003(\text{STA/WCS}) - 0,04(\text{ITR} \times \text{DSIR}) - 0,0001(\text{ITR/NPM}) + 0,001(\text{ITR/QR}) + 0,00003(\text{DSIR/STA}) - 0,04(\text{NPM/ROE}) - 0,01(\text{OPM/NPM}) + 0,05(\text{DER} \times \text{WCS}) - 0,24(\text{QR/CR}) - 1,90(\text{ETD/QR}) - 0,60(\text{WCA} \times \text{WCA}) + 0,001(\text{WCA/NPM}) + 0,0003(\text{WCS/OPM})$
1	$g(x) = 4,46 - 17,17\text{EBITA} - 0,08\text{STA} + 0,000002\text{ITR} + 0,99\text{DAR} + 0,003\text{CR} - 0,01\text{ETD} + 0,00005\text{USD/IDR} - 0,001(\text{EBITA/WCS}) - 0,0002(\text{STA/WCS}) - 0,014(\text{ITR} \times \text{DSIR}) + 0,0001(\text{ITR} \times \text{OPM}) + 0,0002(\text{ITR/NPM}) + 0,04(\text{NPM/ROE}) + 0,003(\text{DER} \times \text{WCS}) - 0,20(\text{QR/CR}) - 0,06(\text{bi7drr/Inflasi})$
2	$g(x) = 2,68 - 12,23\text{EBITA} - 0,02\text{STA} + 0,50\text{DAR} + 0,001\text{QR} - 0,04\text{ETD} - 0,40\text{WCA} + 0,000002\text{USD/IDR} - 3,47(\text{EBITA} \times \text{WCA}) - 0,0008(\text{EBITA/WCS}) - 0,0003(\text{STA/WCS}) - 0,008(\text{ITR} \times \text{DSIR}) + 0,00002(\text{ITR/QR}) + 0,09(\text{ROE} \times \text{NPM}) + 0,004(\text{ROE/QR}) + 0,03(\text{NPM/ROE})$
3	$g(x) = 7,63 - 9,68\text{EBITA} - 0,07\text{OPM} + 0,001\text{DER} + 0,46\text{DAR} + 0,004\text{QR} - 0,62\text{WCA} - 0,03\text{bi7drr} - 0,0002\text{USD/IDR} - 10,57(\text{EBITA} \times \text{EBITA}) - 7,86(\text{EBITA} \times \text{WCA}) - 0,001(\text{EBITA/WCS}) - 0,0002(\text{STA/WCS}) - 0,014(\text{ITR} \times \text{DSIR}) + 0,001(\text{ITR/QR}) + 0,00003(\text{DSIR/STA}) + 0,17(\text{ROE} \times \text{NPM}) + 0,05(\text{NPM/ROE}) + 0,00004(\text{OPM/NPM}) - 0,57(\text{QR/CR}) - 0,0002(\text{ETD/EBITA}) + 0,001(\text{WCS/OPM}) - 0,45(\text{bi7drr/Inflasi})$

Tabel 4.24 menunjukkan bahwa variabel EBITA, STA, NPM, OPM, ETD, dan WCA berpengaruh signifikan dan bertanda negatif pada model *Lasso Logistic Regression* pada *size* 0. Hal tersebut menunjukkan bahwa setiap kenaikan 1% EBITA, STA, NPM, OPM, ETD, dan WCA akan menurunkan peluang

perusahaan sektor industri di Indonesia mengalami kondisi *financial distress*. Sedangkan variabel ITR, DER, DAR, QR, *real* GDP *growth*, BI 7-*day repo rate*, dan nilai tukar dollar berpengaruh signifikan dan bertanda positif. Hal tersebut menunjukkan bahwa setiap kenaikan variabel-variabel tersebut akan meningkatkan peluang perusahaan sektor industri di Indonesia mengalami *financial distress*. Berdasarkan model *Lasso Logistic Regression* pada Tabel 4.24, maka kemudian dapat dilakukan klasifikasi menggunakan data *training* dan data *testing* pada tiap *size* untuk mengetahui ketepatan hasil prediksi dengan model tersebut. Hasil klasifikasi model *Lasso Logistic Regression* dengan *synthetic features generation* ditunjukkan pada Tabel 4.25.

Tabel 4.25a *Confusion Matrix* Model *Lasso Logistic Regression* dengan *Synthetic Features Generation* Untuk *Size* 0.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1187	26	0	284	20
1	54	328	1	10	121

Tabel 4.25b *Confusion Matrix* Model *Lasso Logistic Regression* dengan *Synthetic Features Generation* Untuk *Size* 1.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1042	59	0	274	30
1	174	175	1	30	101

Tabel 4.25c *Confusion Matrix* Model *Lasso Logistic Regression* dengan *Synthetic Features Generation* Untuk *Size* 2.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	939	49	0	273	31
1	176	141	1	45	86

Tabel 4.25d *Confusion Matrix* Model *Lasso Logistic Regression* dengan *Synthetic Features Generation* Untuk *Size* 3.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	819	53	0	288	16
1	153	135	1	83	48

Berdasarkan *confusion matrix* hasil klasifikasi pada Tabel 4.25, maka dapat dilakukan analisis kebaikan model *Lasso Logistic Regression* dengan *synthetic features generation* pada tiap *size* dengan menghitung nilai *accuracy*, AUC, dan *F*1-*score*. Hasil kebaikan model tersebut dapat dibandingkan dengan model *Logistic Regression* secara serentak untuk mengetahui besarnya pengaruh eliminasi variabel dengan menggunakan model *Lasso Logistic Regression*. Evaluasi hasil klasifikasi model *Lasso Logistic Regression* dengan *synthetic features generation* pada tiap *size* ditunjukkan pada Tabel 4.26.

Tabel 4.26 Evaluasi Hasil Klasifikasi Model *Lasso Logistic Regression* dengan *Synthetic Features Generation*.

Size	***λ***	**Ukuran**	***Training***	***Testing***
0	0,0011	*Accuracy*	94,98%	93,10%
		AUC	91,86%	92,89%
		*F*1-*score*	96,74%	94,98%
1	0,0110	*Accuracy*	83,93%	86,21%
		AUC	72,39%	83,62%
		*F*1-*score*	89,94%	90,13%
2	0,0138	*Accuracy*	82,76%	82,53%
		AUC	69,76%	77,72%
		*F*1-*score*	89,30%	87,78%
3	0,0055	*Accuracy*	82,24%	77,24%
		AUC	70,40%	65,69%
		*F*1-*score*	88,83%	85,33%

Evaluasi hasil klasifikasi model *Lasso Logistic Regression* yang ditampilkan pada Tabel 4.26, menunjukkan ada kecenderungan penurunan kebaikan model setiap kenaikan *size*. Hasil evaluasi tersebut juga menunjukkan bahwa dengan menggunakan variabel hasil eliminasi dengan metode *Lasso Logistic Regression* meningkatkan kebaikan hasil prediksi dari model *Logistic Regression* secara serentak. Nilai *accuracy*, AUC, dan *F*1-*score* terbaik ialah model pada *size* 0.

4.2.6 *Elastic-Net Logistic Regression* dengan *Synthetic Features Generation*

Elastic-Net Logistic Regression mengatasi kelemahan dari model *Lasso Logistic Regression*, dimana kelompok prediktor yang berkorelasi dapat dipilih bersama-sama untuk dieliminasi. Pada penelitian ini nilai λ dan γ yang dipilih merupakan nilai λ dan γ optimum yang memberikan ketepatan klasifikasi terbaik. Nilai koefisien pada model *Elastic-Net Logistic Regression* untuk tiap *size* ditampilkan pada Lampiran 7. Adapun variabel yang terpilih dalam model *Elastic-Net Logistic Regression* ditunjukkan pada Tabel 4.27.

Tabel 4.27 Variabel Model *Elastic-Net Logistic Regression* dengan *Synthetic Features Generation*.

Size	**Variabel**
0	EBITA, STA, ITR, NPM, OPM, DER, DAR, QR, ETD, WCA, rgdpg, bi7drr, USD/IDR, EBITA×EBITA, EBITA×DAR, EBITA×WCS, STA×ITR, STA×WCS, ITR×DSIR, ITR/NPM, ITR/QR, DSIR/STA, ROE×NPM, NPM/ROE, OPM/NPM, DER×WCS, QR/CR, ETD/QR, WCA×WCA, WCA/NPM, WCS/OPM
1	EBITA, STA, ITR, DSIR, NPM, DER, DAR, CR, QR, ETD, WCA, rgdpg, bi7drr, USD/IDR, Inflasi, EBITA×EBITA, EBITA×DAR, EBITA×WCA, EBITA×WCS, STA×STA, STA×WCS, ITR×DSIR, ITR×OPM, ITR/QR, DSIR/STA, ROE×NPM, NPM/ROE, OPM/NPM, DER×WCS, QR/CR, ETD/EBITA, ETD/QR, WCA×WCA, WCA/NPM, bi7drr/Inflasi
2	EBITA, STA, ITR, DSIR, NPM, OPM, DER, DAR, QR, ETD, WCA, WCS, rgdpg, bi7drr, USD/IDR, Inflasi, EBITA×EBITA, EBITA×DAR, EBITA×WCA, EBITA×WCS, STA×ITR, STA×WCS, ITR×DSIR, ITR×OPM, ITR/NPM, ITR/QR, DSIR/STA, ROE×NPM, ROE/QR, NPM/ROE, OPM/NPM, DER×WCS, QR/CR, ETD/EBITA, ETD/QR, WCA×WCA, WCA/NPM, WCS/OPM, bi7drr/Inflasi

Tabel 4.27 Variabel Model *Elastic-Net Logistic Regression* dengan *Synthetic Features Generation* (Lanjutan).

Size	**Variabel**
3	EBITA, STA, OPM, DER, DAR, QR, WCA, bi7drr, USD/IDR, EBITA×EBITA, EBITA×WCA, EBITA×WCS, STA×WCS, ITR×DSIR, ITR/QR, DSIR/STA, ROE×NPM, NPM/ROE, QR/CR, ETD/EBITA, WCS/OPM, bi7drr/Inflasi

Berdasarkan variabel terpilih dari hasil eliminasi Tabel 4.27, maka kemudian dapat dibentuk fungsi *link* dari model *Elastic-Net Logistic Regression* pada tiap *size* yang ditampilkan pada Tabel 4.28.

Tabel 4.28 Model *Elastic-Net Logistic Regression* dengan *Synthetic Features Generation*.

Size	**Model**
0	$g(x) = 12,58 - 44,70\text{EBITA} - 0,47\text{STA} + 0,003\text{ITR} - 4,93\text{NPM} - 6,71\text{OPM} + 0,002\text{DER} + 0,94\text{DAR} + 0,002\text{QR} - 0,87\text{ETD} - 1,11\text{WCA} + 0,08\text{rgdpg} + 0,58\text{bi7drr} + 0,0001\text{USD/IDR} + 23,80(\text{EBITA} \times \text{EBITA}) + 6,99(\text{EBITA} \times \text{DAR}) - 0,001(\text{EBITA/WCS}) - 0,09(\text{STA/ITR}) - 0,0001(\text{STA/WCS}) - 0,03(\text{ITR} \times \text{DSIR}) - 0,0001(\text{ITR/NPM}) + 0,001(\text{ITR/QR}) + 0,00005(\text{DSIR/STA}) + 0,48(\text{ROE} \times \text{NPM}) - 0,03(\text{NPM/ROE}) - 0,01(\text{OPM/NPM}) + 0,04(\text{DER} \times \text{WCS}) - 0,40(\text{QR/CR}) - 1,56(\text{ETD/QR}) - 0,56(\text{WCA} \times \text{WCA}) + 0,001(\text{WCA/NPM}) + 0,001(\text{WCS/OPM})$
1	$g(x) = 11,06 - 11,99\text{EBITA} - 0,26\text{STA} - 0,0001\text{ITR} - 0,0005\text{DSIR} - 0,21\text{NPM} + 0,01\text{DER} + 1,52\text{DAR} + 0,008\text{CR} + 0,003\text{QR} - 0,08\text{ETD} - 0,06\text{WCA} - 0,38\text{rgdpg} - 0,04\text{bi7drr} - 0,00001\text{USD/IDR} - 0,041\text{Inflasi} - 7,17(\text{EBITA} \times \text{EBITA}) - 5,12(\text{EBITA} \times \text{DAR}) - 6,49(\text{EBITA} \times \text{WCA}) - 0,001(\text{EBITA/WCS}) + 0,02(\text{STA} \times \text{STA}) - 0,0002(\text{STA/WCS}) - 0,02(\text{ITR} \times \text{DSIR}) + 0,0001(\text{ITR} \times \text{OPM}) + 0,001(\text{ITR/QR}) + 0,00002(\text{DSIR/STA}) - 0,08(\text{ROE} \times \text{NPM}) + 0,07(\text{NPM/ROE}) - 0,0003(\text{OPM/NPM}) + 0,05(\text{DER} \times \text{WCS}) - 0,57(\text{QR/CR}) - 0,0002(\text{ETD/EBITA}) + 0,003(\text{ETD/QR}) - 0,41(\text{WCA} \times \text{WCA}) + 0,0003(\text{WCA/NPM}) - 0,71(\text{bi7drr/Inflasi})$

Tabel 4.28 Model *Elastic-Net Logistic Regression* dengan *Synthetic Features Generation* (Lanjutan).

Size	**Model**
2	$g(x) = 3,92 - 7,61\text{EBITA} - 0,12\text{STA} + 0,0001\text{ITR} + 0,0001\text{DSIR} - 0,22\text{NPM} - 0,83\text{OPM} + 0,001\text{DER} + 1,13\text{DAR} + 0,04\text{QR} - 0,20\text{ETD} - 0,74\text{WCA} - 0,04\text{WCS} + 0,18\text{rgdpg} + 0,08\text{bi7drr} - 0,0002\text{USD/IDR} - 0,09\text{Inflasi} - 15,42(\text{EBITA} \times \text{EBITA}) - 5,81(\text{EBITA} \times \text{DAR}) - 6,40(\text{EBITA} \times \text{WCA}) - 0,001(\text{EBITA}/\text{WCS}) + 0,46(\text{STA}/\text{ITR}) - 0,0003(\text{STA}/\text{WCS}) - 0,02(\text{ITR} \times \text{DSIR}) + 0,0001(\text{ITR} \times \text{OPM}) - 0,000003(\text{ITR}/\text{NPM}) + 0,001(\text{ITR}/\text{QR}) - 0,0001(\text{DSIR}/\text{STA}) + 0,73(\text{ROE} \times \text{NPM}) + 0,02(\text{ROE}/\text{QR}) + 0,12(\text{NPM}/\text{ROE}) + 0,0008(\text{OPM}/\text{NPM}) + 0,004(\text{DER} \times \text{WCS}) - 0,26(\text{QR}/\text{CR}) - 0,0001(\text{ETD}/\text{EBITA}) - 0,05(\text{ETD}/\text{QR}) - 0,52(\text{WCA} \times \text{WCA}) + 0,0003(\text{WCA}/\text{NPM}) + 0,0001(\text{WCS}/\text{OPM}) - 0,38(\text{bi7drr}/\text{Inflasi})$
3	$g(x) = 6,29 - 8,59\text{EBITA} - 0,0004\text{STA} - 0,15\text{OPM} + 0,0003\text{DER} + 0,41\text{DAR} + 0,03\text{QR} - 0,63\text{WCA} - 0,02\text{bi7drr} - 0,0001\text{USD/IDR} - 7,29(\text{EBITA} \times \text{EBITA}) - 7,44(\text{EBITA} \times \text{WCA}) - 0,001(\text{EBITA}/\text{WCS}) - 0,0002(\text{STA}/\text{WCS}) - 0,01(\text{ITR} \times \text{DSIR}) + 0,0005(\text{ITR}/\text{QR}) + 0,00003(\text{DSIR}/\text{STA}) + 0,18(\text{ROE} \times \text{NPM}) + 0,04(\text{ROE}/\text{QR}) - 0,53(\text{QR}/\text{CR}) - 0,0001(\text{ETD}/\text{EBITA}) + 0,001(\text{WCS}/\text{OPM}) - 0,34(\text{bi7drr}/\text{Inflasi})$

Berdasarkan Tabel 4.28 dapat diketahui bahwa variabel EBITA, STA, NPM, OPM, ETD, dan WCA berpengaruh signifikan dan bertanda negatif dalam model *Elastic-Net Logistic Regression* dengan *synthetic features generation* pada *size* 0. Hal tersebut menunjukkan bahwa setiap kenaikan 1% pada rasio keuangan tersebut akan menurunkan peluang perusahaan sektor industri di Indonesia mengalami *financial distress*. Sedangkan pada variabel ITR, DER, DAR, QR, *real* GDP *growth*, BI 7-*day repo rate*, dan nilai tukar dollar berpengaruh signifikan dan bertanda positif, artinya setiap kenaikan rasio keuangan dan indikator ekonomi makro tersebut akan meningkatkan peluang perusahaan sektor industri di Indonesia mengalami *financial distress*. Kemudian dilakukan klasifikasi kondisi keuangan

perusahaan sektor industri di Indonesia dengan menggunakan model *Elastic-Net Logistic Regression*. Model *Elastic-Net Logistic Regression* dibentuk dari data *training*, kemudian dari hasil prediksi model tersebut dilakukan evaluasi menggunakan data *training* dan data *testing* pada tiap *size*. Hasil klasifikasi model *Elastic-Net Logistic Regression* dengan *synthetic features generation* ditampilkan pada Tabel 4.29.

Tabel 4.29a *Confusion Matrix* Model *Elastic-Net Logistic Regression* dengan *Synthetic Features Generation* Untuk *Size* 0.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1189	24	0	283	21
1	56	326	1	6	125

Tabel 4.29b *Confusion Matrix* Model *Elastic-Net Logistic Regression* dengan *Synthetic Features Generation* Untuk *Size* 1.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1038	63	0	266	38
1	154	195	1	33	98

Tabel 4.29c *Confusion Matrix* Model *Elastic-Net Logistic Regression* dengan *Synthetic Features Generation* Untuk *Size* 2.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	936	52	0	227	77
1	158	159	1	18	113

Tabel 4.29d *Confusion Matrix* Model *Elastic-Net Logistic Regression* dengan *Synthetic Features Generation* Untuk *Size* 3.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	819	53	0	287	17
1	155	133	1	75	56

Confusion matrix pada Tabel 4.29 menunjukkan ketepatan hasil klasifikasi menggunakan model *Elastic-Net Logistic Regression* pada tiap *size*. Kemudian dilakukan evaluasi kebaikan model dengan menghitung *accuracy*, AUC, dan *F*1-*score* pada tiap *size* untuk mengetahui kebaikan model dalam memprediksi kondisi

keuangan perusahaan sektor industri di Indonesia yang ditampilkan pada Tabel 4.30.

Tabel 4.30 Evaluasi Hasil Klasifikasi Model *Elastic-Net Logistic Regression* dengan *Synthetic Features Generation*.

Size	λ	γ	**Ukuran**	***Training***	***Testing***
0	0,0013	0,7	*Accuracy*	94,98%	93,79%
			AUC	91,68%	94,26%
			*F*1-*score*	96,75%	95,45%
1	0,0054	0,2	*Accuracy*	85,03%	83,68%
			AUC	75,08%	81,15%
			*F*1-*score*	90,54%	88,23%
2	0,0037	0,1	*Accuracy*	83,91%	78,16%
			AUC	72,45%	80,47%
			*F*1-*score*	89,91%	82,70%
3	0,0104	0,7	*Accuracy*	82,07%	78,85%
			AUC	70,05%	68,58%
			*F*1-*score*	88,73%	86,19%

Berdasarkan Tabel 4.30 dapat diketahui bahwa model *Elastic-Net Logistic Regression* dengan *synthetic features generation* memiliki nilai *accuracy*, AUC, dan *F*1-*score* terbaik pada *size* 0. Evaluasi hasil klasifikasi model *Elastic-Net Logistic Regression* yang ditampilkan pada Tabel 4.30, menunjukkan ada kecenderungan penurunan kebaikan model setiap kenaikan *size*. Hasil evaluasi pada Tabel 4.30 juga menunjukkan bahwa dengan menggunakan variabel hasil eliminasi dengan metode *Elastic-Net Logistic Regression* meningkatkan kebaikan hasil prediksi dari model *Logistic Regression* secara serentak.

4.2.7 *Support Vector Machine* (SVM) dengan *Synthetic Features Generation*

Pada penelitian ini metode *Support Vector Machine* (SVM) yang digunakan merupakan *Support Vector Machinne* (SVM) dengan *synthetic features generation* untuk kasus yang dapat dipisahkan secara nonlinier yaitu dengan *Radial Basis Function* (RBF). Hasil klasifikasi menggunakan *Support Vector Machine* (SVM) dengan *synthetic features generation* ditampilkan pada *confusion matrix* yang ditunjukkan pada Tabel 4.31.

Tabel 4.31a *Confusion Matrix* Model SVM dengan *Synthetic Features Generation* Untuk *Size* 0.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1210	3	0	298	6
1	168	214	1	62	69

Tabel 4.31b *Confusion Matrix* Model SVM dengan *Synthetic Features Generation* Untuk *Size* 1.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1088	13	0	296	8
1	212	137	1	83	48

Tabel 4.31c *Confusion Matrix* Model SVM dengan *Synthetic Features Generation* Untuk *Size* 2.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	971	17	0	291	13
1	195	122	1	85	46

Tabel 4.31d *Confusion Matrix* Model SVM dengan *Synthetic Features Generation* Untuk *Size* 3.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	865	7	0	291	13
1	285	103	1	98	33

Berdasarkan *confusion matrix* pada Tabel 4.31 dapat diketahui hasil prediksi kondisi keuangan dengan menggunakan model *Support Vector Machine* (SVM) dengan *synthetic features generation* pada tiap *size*. Kemudian dilakukan evaluasi kebaikan hasil klasifikasi tersebut dengan menghitung nilai *accuracy*, AUC, dan *F*1-*score* pada model *Support Vector Machine* (SVM) dengan *synthetic feature generation* pada tiap *size*. Setelah melakukan evaluasi hasil klasifikasi model *Support Vector Machine* (SVM) dengan *synthetic features generation* pada tiap *size*, kemudian menentukan *size* yang memiliki ketepatan klasifikasi terbaik dari model *Support Vector Machine* (SVM) dengan *synthetic features generation*. Evaluasi hasil klasifikasi menggunakan model *Support*

Vector Machine (SVM) dengan *synthetic features generation* pada tiap *size* ditunjukkan pada Tabel 4.32.

Tabel 4.32 Evaluasi Hasil Klasifikasi Model SVM dengan *Synthetic Features Generation*.

Size	**Ukuran**	***Training***	***Testing***
0	*Accuracy*	89,28%	84,37%
	AUC	77,89%	75,35%
	*F*1-*score*	93,40%	89,76%
1	*Accuracy*	84,48%	79,08%
	AUC	69,04%	67,00%
	*F*1-*score*	90,63%	86,68%
2	*Accuracy*	83,75%	77,47%
	AUC	68,38%	65,42%
	*F*1-*score*	90,16%	85,59%
3	*Accuracy*	76,83%	74,48%
	AUC	62,87%	60,46%
	*F*1-*score*	85,56%	83,98%

Tabel 4.32 menunjukkan bahwa model *Support Vector Machine* (SVM) dengan *synthetic features generation* dengan hasil klasifikasi terbaik ialah pada *size* 0. Evaluasi hasil klasifikasi tersebut menunjukkan bahwa semakin besar *size* akan memperkecil kebaikan hasil klasifikasi. Nilai AUC model *Support Vector Machine* (SVM) dengan *synthetic features generation* pada tiap *size* cenderung rendah dibandingkan dengan nilai *F*1-*score*. Hal tersebut menunjukkan bahwa model lebih baik dalam memprediksi kondisi keuangan perusahaan sektor industri yang sehat dibandingkan kondisi perusahaan sektor industri yang mengalami *financial distress*.

4.2.8 *Lasso Support Vector Machine* dengan *Synthetic Features Generation*

Metode *Lasso Support Vector Machine* merupakan metode untuk mengeliminasi variabel yang tidak relevan dalam model dengan mengecilkan nilai koefisiennya menjadi 0. Pada penelitian ini nilai λ yang digunakan merupakan nilai λ optimum yang memberikan ketepatan klasifikasi terbaik. Nilai koefisien pada model *Lasso Support Vector Machine* dengan *synthetic features generation* untuk tiap *size* ditampilkan pada Lampiran 8. Variabel

yang terpilih dalam model *Lasso Support Vector Machine* ditunjukkan pada Tabel 4.33.

Tabel 4.33 Variabel Model *Lasso* SVM dengan *Synthetic Features Generation*.

Size	**Variabel**
0	EBITA, STA, ITR, DSIR, NPM, OPM, DAR, CR, QR, ETD, WCA, WCS, rgdpg, bi7drr, USD/IDR, Inflasi, EBITA×DAR, EBITA×WCA, EBITA×WCS, STA×ITR, STA×WCS, ITR×DSIR, ITR×OPM, ITR/NPM, DSIR/STA, ROE×NPM, NPM/ROE, OPM/NPM, DER×WCS, QR/CR, ETD/EBITA, ETD/QR, WCA×WCA, WCA/NPM, WCS/OPM, bi7drr/Inflasi
1	EBITA, STA, OPM, DER, DAR, CR, ETD, WCA, WCS, rgdpg, bi7drr, USD/IDR, Inflasi, EBITA×EBITA, EBITA×DAR, EBITA×WCA, EBITA×WCS, STA×STA, STA×ITR, STA×WCS, ITR×DSIR, ITR×OPM, ITR/NPM, ROE×NPM, NPM/ROE, OPM/NPM, DER×WCS, QR/CR, ETD/EBITA, ETD/QR, WCA/NPM, WCS/OPM, bi7drr/Inflasi
2	EBITA, STA, ITR, DSIR, ROE, NPM, OPM, DER, DAR, CR, QR, ETD, WCA, rgdpg, bi7drr, USD/IDR, Inflasi, EBITA×EBITA, EBITA×DAR, EBITA×WCS, STA×ITR, STA×WCS, ITR×DSIR, ITR×OPM, ITR/QR, DSIR/STA, ROE×NPM, ROE/QR, NPM/ROE, OPM/NPM, DER×WCS, QR/CR, WCA×WCA, WCA/NPM, WCS/OPM, bi7drr/Inflasi
3	EBITA, STA, ITR, DSIR, ROE, NPM, OPM, DER, DAR, CR, QR, ETD, WCA, WCS, rgdpg, bi7drr, USD/IDR, Inflasi, EBITA×EBITA, EBITA×DAR, EBITA×WCA, EBITA×WCS, STA×STA, STA×ITR, STA×WCS, ITR×DSIR, ITR×OPM, ITR/NPM, ITR/QR, DSIR/STA, ROE×NPM, ROE/QR, NPM/ROE, OPM/NPM, DER×WCS, QR/CR, ETD/EBITA, ETD/QR, WCA/NPM, WCS/OPM, bi7drr/Inflasi

Menggunakan variabel terpilih yang ditampilkan pada Tabel 4.33, maka dapat dilakukan prediksi kondisi keuangan perusahaan sektor industri di Indonesia dengan menggunakan model *Lasso Support Vector Machine* dengan *synthetic features generation* yang ditampilkan pada Tabel 4.34.

Tabel 4.34a *Confusion Matrix* Model *Lasso* SVM dengan *Synthetic Features Generation* Untuk *Size* 0.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1197	48	0	289	14
1	16	334	1	15	117

Tabel 4.34b *Confusion Matrix* Model *Lasso* SVM dengan *Synthetic Features Generation* Untuk *Size* 1.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1017	136	0	262	37
1	84	213	1	42	94

Tabel 4.34c *Confusion Matrix* Model *Lasso* SVM dengan *Synthetic Features Generation* Untuk *Size* 2.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	949	242	0	282	108
1	39	75	1	22	23

Tabel 4.34d *Confusion Matrix* Model *Lasso* SVM dengan *Synthetic Features Generation* Untuk *Size* 3.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	815	155	0	293	84
1	57	133	1	11	47

Berdasarkan *confusion matrix* model *Lasso Support Vector Machine* dengan *synthetic features generation* tiap *size* pada Tabel 4.34, dapat dilihat ketepatan hasil klasifikasi kondisi keuangan perusahaan sektor industri di Indonesia menggunakan model *Lasso Support Vector Machine* dengan *synthetic features generation* pada tiap *size*. Kemudian dilakukan perhitungan nilai *accuracy*, AUC, dan *F*1-*score* pada model tiap *size* untuk mengetahui kebaikan model *Lasso Support Vector Machine* dengan *synthetic features generation* dalam memprediksi kondisi keuangan perusahaan sektor industri di Indonesia. Evaluasi hasil klasifikasi model *Lasso Support Vector Machine* dengan *synthetic features generation* pada tiap *size* ditunjukkan pada Tabel 4.35.

Tabel 4.35 Evaluasi Hasil Klasifikasi Model *Lasso* SVM dengan *Synthetic Features Generation*.

Size	***λ***	**Ukuran**	***Training***	***Testing***
0	0,0031	*Accuracy*	95,99%	93,33%
		AUC	95,79%	92,01%
		*F*1-*score*	97,40%	95,22%
1	0,0026	*Accuracy*	84,83%	81,84%
		AUC	79,96%	78,37%
		*F*1-*score*	90,24%	86,90%
2	0,0025	*Accuracy*	78,47%	70,11%
		AUC	72,74%	61,71%
		*F*1-*score*	87,10%	81,27%
3	0,0023	*Accuracy*	81,72%	78,16%
		AUC	77,01%	79,38%
		*F*1-*score*	88,49%	86,05%

Berdasarkan Tabel 4.35 dapat dilihat bahwa dengan melakukan klasifikasi menggunakan variabel terpilih dalam model *Lasso Support Vector Machine* meningkatkan kebaikan hasil klasifikasi dibandingkan dengan seluruh variabel pada model *Support Vector Machine*. Model *Lasso Support Vector Machine* dengan *synthetic features generation* yang memiliki nilai *accuracy*, AUC, dan *F*1-*score* terbaik ialah model *Lasso Support Vector Machine* dengan *synthetic features generation* pada *size* 0. Sedangkan model yang memiliki nilai *accuracy*, AUC, dan *F*1-*score* terendah ialah model *Lasso Support Vector Machine* dengan *synthetic features generation* pada *size* 2.

4.2.9 *Elastic-Net Support Vector Machine* dengan *Synthetic Features Generation*

Elastic-Net Support Vector Machine mengatasi kelemahan dari model *Lasso Support Vector Machine*, dimana kelompok prediktor yang berkorelasi dapat dipilih secara bersama-sama untuk dieliminasi. Pada penelitian ini nilai λ_1 dan λ_2 yang dipilih merupakan nilai λ_1 dan λ_2 optimum yang memberikan ketepatan klasifikasi terbaik. Nilai koefisien pada model *Elastic-Net Support Vector Machine* untuk tiap *size* ditampilkan pada Lampiran 9.

Variabel terpilih dalam model *Elastic-Net Support Vector Machine* dengan *synthetic features generation* ditunjukkan pada Tabel 4.36.

Tabel 4.36 Variabel Model *Elastic-Net Support Vector Machine* dengan *Synthetic Features Generation*.

Size	**Variabel**
0	ITR, DSIR, DER, QR, USD/IDR, EBITA×WCS, STA×STA, STA×WCS, ITR×OPM, ITR/NPM, ITR/QR, DSIR/STA, ROE/QR, NPM/ROE, OPM/NPM, QR/CR, ETD/EBITA, ETD/QR, WCA/NPM, WCS/OPM
1	ITR, DSIR, DER, CR, USD/IDR, STA×STA, STA×WCS, ITR×OPM, ITR/NPM, ITR/QR, DSIR/STA, OPM/NPM, ETD/EBITA, WCA/NPM, WCS/OPM
2	ITR, DSIR, USD/IDR, STA×WCS, ITR×OPM, ITR/NPM, ITR/QR, DSIR/STA, ETD/EBITA, WCA/NPM
3	DSIR, DER, USD/IDR, EBITA×WCS, STA×WCS, ITR×OPM, ITR/NPM, ITR/QR, DSIR/STA, ETD/EBITA, WCS/OPM

Berdasarkan Tabel 4.36 dapat dilihat bahwa variabel terpilih dalam model *Elastic-Net Support Vector Machine* pada tiap *size* yang memiliki jumlah variabel paling banyak ialah model pada *size* 2, sedangkan yang memiliki jumlah variabel paling sedikit ialah model pada *size* 3. Kemudian dengan menggunakan variabel terpilih pada model *Elastic-Net Support Vector Machine* pada tiap *size* tersebut dilakukan klasifikasi kondisi keuangan perusahaan sektor industri di Indonesia dengan menggunakan model yang terbentuk dari data *training* dan selanjutnya dilakukan evaluasi menggunakan data *training* dan data *testing* untuk mengetahui ketepatan hasil klasifikasi model pada tiap *size*. Hasil klasifikasi model *Elastic-Net Support Vector Machine* dengan *synthetic features generation* pada tiap *size* ditampilkan pada Tabel 4.37.

Tabel 4.37a *Confusion Matrix* Model *Elastic-Net* SVM dengan *Synthetic Features Generation* Untuk *Size* 0.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1183	118	0	223	5
1	30	264	1	81	126

Tabel 4.37b *Confusion Matrix* Model *Elastic-Net* SVM dengan *Synthetic Features Generation* Untuk *Size* 1.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1085	297	0	295	114
1	16	52	1	9	17

Tabel 4.37c *Confusion Matrix* Model *Elastic-Net* SVM dengan *Synthetic Features Generation* Untuk *Size* 2.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	984	299	0	302	127
1	4	18	1	2	4

Tabel 4.37d *Confusion Matrix* Model *Elastic-Net* SVM dengan *Synthetic Features Generation* Untuk *Size* 3.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	860	242	0	290	104
1	12	46	1	14	27

Confusion matrix pada Tabel 4.37 menunjukkan ketepatan hasil klasifikasi menggunakan model *Elastic-net Support Vector Machine* pada tiap *size*. Untuk mengetahui kebaikan hasil klasifikasi model pada tiap *size* tersebut dapat dilihat melalui nilai *accuracy*, AUC, dan *F*1-*score* yang dihasilkan pada model *Elastic-Net Support Vector Machine* pada tiap *size*. Evaluasi hasil klasifikasi menggunakan model *Elastic-Net Support Vector Machine* dengan *synthetic features generation* pada tiap *size* ditunjukkan pada Tabel 4.38.

Tabel 4.38 Evaluasi Hasil Klasifikasi Model *Elastic-Net* SVM dengan *Synthetic Features Generation*.

Size	λ_1	λ_2	**Ukuran**	***Training***	***Testing***
0	42,6	0,001	*Accuracy*	90,72%	80,23%
			AUC	90,36%	79,34%
			*F*1-*score*	94,11%	83,83%
1	231	0,001	*Accuracy*	78,41%	71,72%
			AUC	77,49%	68,76%
			*F*1-*score*	87,39%	82,75%

Tabel 4.38 Evaluasi Hasil Klasifikasi Model *Elastic-Net* SVM dengan *Synthetic Features Generation* (Lanjutan).

Size	$\boldsymbol{\lambda_1}$	$\boldsymbol{\lambda_2}$	**Ukuran**	***Training***	***Testing***
			Accuracy	76,78%	70,34%
2	842	0,004	AUC	79,26%	68,53%
			*F*1-*score*	86,66%	82,40%
			Accuracy	78,10%	72,87%
3	1321	1	AUC	78,68%	69,73%
			*F*1-*score*	87,13%	83,09%

Berdasarkan evaluasi hasil klasifikasi model *Elastic-Net Support Vector Machine* pada Tabel 4.38 dapat diketahui bahwa model yang memiliki nilai *accuracy*, AUC, dan *F*1-*score* terbaik ialah model *Elastic-Net Support Vector Machine* pada *size* 0. Hasil evaluasi ketepatan klasifikasi dengan menggunakan data *testing* pada model *size* 0 memberikan hasil yang cukup jauh berbeda dari data *training*, sehingga ada kecenderungan model yang dihasilkan *overfitting*. Kemudian berdasarkan evaluasi hasil klasifikasi tersebut dapat diketahui bahwa model *Elastic-Net Support Vector Machine* lebih baik dalam memprediksi kondisi keuangan perusahaan sektor industri di Indonesia dibandingkan dengan hasil ketepatan klasifikasi model *Support Vector Machine*. Tetapi hasil klasifikasi dengan model *Elastic-Net Support Vector Machine* kurang baik jika dibandingkan dengan hasil ketepatan klasifikasi dengan model *Lasso Support Vector Machine*.

4.2.10 *Extreme Gradient Boosting* (XGB) dengan *Synthetic Features Generation* Secara Serentak

Pada penelitian ini model *Extreme Gradient Boosting* (XGB) dengan *synthetic features generation* menggunakan nilai λ dan γ optimum yang memberikan model dengan ketepatan hasil klasifikasi terbaik. Hasil prediksi kondisi keuangan perusahaan sektor industri di Indonesia dengan menggunakan model *Extreme Gradient Boosting* (XGB) dengan *synthetic features generation* ditunjukkan pada Tabel 4.39.

Tabel 4.39a *Confusion Matrix* Model XGB dengan *Synthetic Features Generation* Untuk *Size* 0.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1211	2	0	298	6
1	6	376	1	14	117

Tabel 4.39b *Confusion Matrix* Model XGB dengan *Synthetic Features Generation* Untuk *Size* 1.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1035	66	0	271	33
1	97	252	1	27	104

Tabel 4.39c *Confusion Matrix* Model XGB dengan *Synthetic Features Generation* Untuk *Size* 2.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	926	62	0	271	33
1	126	191	1	36	95

Tabel 4.39d *Confusion Matrix* Model XGB dengan *Synthetic Features Generation* Untuk *Size* 3.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	833	39	0	284	20
1	96	192	1	68	63

Berdasarkan hasil klasifikasi model *Extreme Gradient Boosting* (XGB) yang ditunjukkan pada *confusion matrix* pada Tabel 4.39, maka dapat dilakukan evaluasi ketepatan hasil klasifikasi model *Extreme Gradient Boosting* (XGB) dengan *synthetic features generation* secara serentak pada tiap *size* yang ditunjukkan pada Tabel 4.40.

Tabel 4.40 Evaluasi Hasil Klasifikasi Model XGB dengan *Synthetic Features Generation* Secara Serentak.

Size	λ	γ	**Ukuran**	***Training***	***Testing***
			Accuracy	99,50%	95,40%
0	0,01	0	AUC	99,13%	93,67%
			*F*1-*score*	99,67%	96,75%

Tabel 4.40 Evaluasi Hasil Klasifikasi Model XGB dengan *Synthetic Features Generation* Secara Serentak (Lanjutan).

Size	$\boldsymbol{\lambda}$	$\boldsymbol{\gamma}$	**Ukuran**	***Training***	***Testing***
			Accuracy	88,76%	86,21%
1	2	0.05	AUC	83,11%	84,27%
			*F*1-*score*	92,70%	90,03%
			Accuracy	85,59%	84,14%
2	0.03	0	AUC	76,99%	80,83%
			*F*1-*score*	90,78%	88,71%
			Accuracy	88,36%	79,77%
3	1	0.01	AUC	81,10%	70,76%
			*F*1-*score*	92,50%	86,59%

Tabel 4.40 menunjukkan nilai λ dan γ optimum yang memberikan ketepatan hasil klasifikasi terbaik pada model tiap *size*. Berdasarkan Evaluasi hasil klasifikasi tersebut dapat diketahui bahwa model *Extreme Gradient Boosting* (XGB) dengan *synthetic features generation* yang memberikan ketepatan hasil klasifikasi terbaik ialah model pada *size* 0. Tabel 4.40 juga menunjukkan bahwa semakin besar *size* akan menurunkan kebaikan model *Extreme Gradient Boosting* (XGB) dengan *synthetic features generation* dalam memprediksi kondisi keuangan perusahaan sektor industri di Indonesia.

4.2.11 *Extreme Gradient Boosting* (XGB) dengan *Synthetic Features Generation* dan Seleksi Variabel

Dalam model *Extreme Gradient Boosting* (XGB) dengan *synthetic features generation* secara serentak memungkinkan adanya variabel yang tidak signifikan dalam model dan mengakibatkan hasil klasifikasi kurang baik. Sehingga kemudian perlu dilakukan seleksi variabel untuk mengeliminasi variabel yang tidak signifikan dalam model. Seleksi variabel pada model *Extreme Gradient Boosting* dilakukan dengan melihat nilai *gain*. Variabel yang memiliki nilai *gain* sangat kecil akan dianggap sebagai variabel yang tidak berpengaruh signifikan dalam model. Nilai *gain* pada tiap variabel untuk tiap *size* ditampilkan pada Lampiran 10. Variabel terpilih yang digunakan dalam model

Extreme Gradient Boosting (XGB) dengan *synthetic features generation* pada tiap *size* ditunjukkan pada Tabel 4.41.

Tabel 4.41 Variabel Model XGB dengan *Synthetic Features Generation* dan Seleksi Variabel.

Size	**Variabel**
0	EBITA, STA, ITR, DSIR, ROE, NPM, OPM, DER, DAR, CR, QR, ETD, WCA, WCS, rgdpg, USD/IDR, EBITA×EBITA, EBITA×DAR, EBITA×WCA, EBITA×WCS, STA×STA, STA×ITR, STA×WCS, ITR×OPM, ITR/NPM, ITR/QR, DSIR/STA, ROE×NPM, ROE/QR, NPM/ROE, OPM/NPM, DER×WCS, QR/CR, ETD/EBITA, ETD/QR, WCA×WCA, WCA/NPM, WCS/OPM
1	EBITA, STA, ITR, DSIR, ROE, OPM, DER, DAR, CR, QR, ETD, WCA, WCS, rgdpg, USD/IDR, EBITA×EBITA, EBITA×DAR, EBITA×WCA, EBITA×WCS, STA×STA, STA×ITR, STA×WCS, ITR×OPM, ITR/NPM, ITR/QR, DSIR/STA, ROE×NPM, NPM/ROE, OPM/NPM, DER×WCS, QR/CR, ETD/EBITA, ETD/QR, WCA×WCA, WCA/NPM, WCS/OPM, bi7drr/Inflasi
2	EBITA, STA, ROE, NPM, OPM, DER, DAR, CR, QR, ETD, WCA, WCS, rgdpg, USD/IDR, Inflasi, EBITA×EBITA, EBITA×DAR, EBITA×WCA, EBITA×WCS, STA×STA, STA×ITR, STA×WCS, ITR×OPM, ITR/NPM, ITR/QR, ROE/QR, NPM/ROE, OPM/NPM, DER×WCS, QR/CR, ETD/EBITA, ETD/QR, WCA×WCA, WCA/NPM, WCS/OPM, bi7drr/Inflasi
3	EBITA, STA, ITR, DSIR, ROE, NPM, OPM, DER, DAR, CR, QR, ETD, WCA, WCS, rgdpg, bi7drr, USD/IDR, EBITA×EBITA, EBITA×DAR, EBITA×WCA, EBITA×WCS, STA×STA, STA×ITR, STA×WCS, ITR×OPM, ITR/NPM, ITR/QR, DSIR/STA, ROE×NPM, ROE/QR, NPM/ROE, OPM/NPM, DER×WCS, QR/CR, ETD/EBITA, ETD/QR, WCA×WCA, WCA/NPM, WCS/OPM, bi7drr/Inflasi

Tabel 4.41 menunjukkan variabel-variabel yang berpengaruh signifikan dalam model *Extreme Gradient Boosting* (XGB) dengan *synthetic features generation* pada tiap *size*. Hasil

klasifikasi kondisi keuangan perusahaan sektor industri di Indonesia menggunakan variabel-variabel yang berpengaruh signifikan dalam model *Extreme Gradient Boosting* (XGB) dengan *synthetic features generation* ditunjukkan pada Tabel 4.42.

Tabel 4.42a *Confusion Matrix* Model XGB dengan *Synthetic Features Generation* dan Seleksi Variabel Untuk *Size* 0.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1213	0	0	299	5
1	0	382	1	13	118

Tabel 4.42b *Confusion Matrix* Model XGB dengan *Synthetic Features Generation* dan Seleksi Variabel Untuk *Size* 1.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	1084	17	0	282	22
1	61	288	1	29	102

Tabel 4.42c *Confusion Matrix* Model XGB dengan *Synthetic Features Generation* dan Seleksi Variabel Untuk *Size* 2.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	964	24	0	283	21
1	62	255	1	46	85

Tabel 4.42d *Confusion Matrix* Model XGB dengan *Synthetic Features Generation* dan Seleksi Variabel Untuk *Size* 3.

Training	Aktual		***Testing***	Aktual	
Prediksi	0	1	Prediksi	0	1
0	847	25	0	284	20
1	65	223	1	63	68

Confusion matrix pada Tabel 4.42 menunjukkan ketepatan hasil klasifikasi model *Extreme Gradient Boosting* (XGB) dengan *synthetic features generation* dan seleksi variabel. Kemudian dilakukan evaluasi kebaikan model *Extreme Gradient Boosting* (XGB) dengan *synthetic features generation* dan seleksi variabel pada tiap *size* dengan menghitung nilai *accuracy*, AUC, dan *F*1-*score*. Evaluasi Hasil klasifikasi model *Extreme Gradient Boosting*

(XGB) dengan *synthetic features generation* dan seleksi variabel pada tiap *size* ditunjukkan pada Tabel 4.43.

Tabel 4.43 Evaluasi Hasil Klasifikasi Model XGB dengan *Synthetic Features Generation* dan Seleksi Variabel.

Size	**Ukuran**	***Training***	***Testing***
0	*Accuracy*	100,00%	95,86%
	AUC	100,00%	94,22%
	*F*1-*score*	100,00%	97,08%
1	*Accuracy*	94,62%	88,28%
	AUC	90,49%	85,31%
	*F*1-*score*	96,53%	91,71%
2	*Accuracy*	93,41%	84,60%
	AUC	89,01%	78,99%
	*F*1-*score*	95,73%	89,42%
3	*Accuracy*	92,24%	80,92%
	AUC	87,28%	72,66%
	*F*1-*score*	94,96%	87,25%

Berdasarkan Tabel 4.43 dapat dilihat bahwa dengan melakukan eliminasi variabel pada model *Extreme Gradient Boosting* (XGB) dapat meningkatkan performansi model. Tabel 4.43 menunjukkan bahwa model *Extreme Gradient Boosting* (XGB) dengan *synthetic features generation* yang memiliki nilai *accuracy*, AUC, dan *F*1-*score* terbaik ialah model pada *size* 0. Evaluasi hasil klasifikasi tersebut menunjukkan semakin besar *size* maka performansi model cenderung semakin rendah.

4.3 Menentukan Model Terbaik

Setelah mendapatkan model dan mengetahui hasil klasifikasi kondisi keuangan perusahaan sektor industri di Indonesia menggunakan model yang tidak melibatkan *synthetic features generation* dan model yang melibatkan *synthetic features generation*, maka dapat dilakukan perbandingan hasil evaluasi klasifikasi menggunakan data *training* dan data *testing* pada masing-masing model. Model terbaik merupakan model yang memiliki nilai *accuracy*, AUC, dan *F*1-*score* terbaik. Dalam penentuan model prediksi terbaik akan dibagi menjadi 2 bagian yaitu model secara serentak dan model dengan seleksi variabel.

4.3.1 Perbandingan Model Secara Serentak

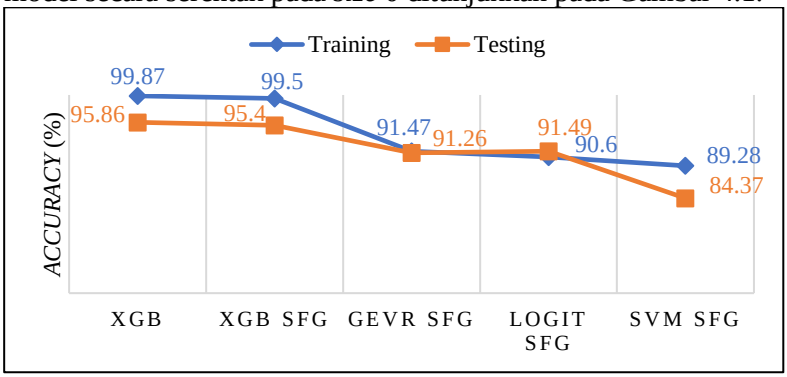
Pada penelitian ini model yang dilakukan secara serentak ialah model *Extreme Gradient Boosting* (XGB) dengan dan tanpa melibatkan *synthetic features generation*, *Generalized Extreme Value regression* (GEVR) dengan *synthetic features generation*, *Logistic Regression* dengan *synthetic features generation*, dan *Support Vector Machine* (SVM) dengan *synthetic features generation*. Perbandingan evaluasi hasil klasifikasi model secara serentak ditunjukkan pada Tabel 4.44.

Tabel 4.44 Perbandingan Model Secara Serentak.

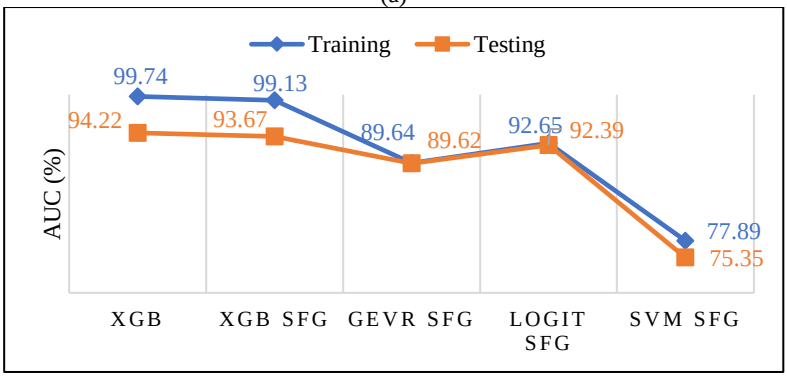
Model	***Size***	***Accuracy* (%)**		**AUC (%)**		***F1-score* (%)**	
		Train	***Test***	***Train***	***Test***	***Train***	***Test***
XGB	0	99,87	95,86	99,74	94,22	99,92	97,08
	1	95,38	85,06	92,07	83,66	97,00	89,08
	2	94,48	82,76	90,58	77,89	96,42	87,96
	3	95,69	80,23	92,25	71,30	97,19	86,89
XGB SFG	0	99,50	95,40	99,13	93,67	99,67	96,75
	1	88,76	86,21	83,11	84,27	92,70	90,03
	2	85,59	84,14	76,99	80,83	90,78	88,71
	3	88,36	79,77	81,10	70,76	92,50	86,59
GEVR SFG	0	91,47	91,26	89,64	89,62	94,32	93,75
	1	85,59	84,14	78,57	76,92	90,66	89,34
	2	81,92	65,75	80,99	74,19	87,39	68,37
	3	82,84	71,26	72,54	53,38	89,07	82,71
LOGIT SFG	0	90,60	91,49	92,65	92,39	93,48	93,68
	1	85,59	82,76	76,22	73,32	90,85	88,72
	2	79,62	44,83	77,11	59,44	85,90	36,51
	3	83,28	72,64	71,55	55,67	89,50	83,40
SVM SFG	0	89,28	84,37	77,89	75,35	93,40	89,76
	1	84,48	79,08	69,04	67,00	90,63	86,68
	2	83,75	77,47	68,38	65,42	90,16	85,59
	3	76,83	74,48	62,87	60,46	85,56	83,98

Berdasarkan Tabel 4.44 dapat dilihat bahwa model *Extreme Gradient Boosting* (XGB) tanpa *synthetic features generation* secara serentak pada *size* 0, 1, 2, dan 3 memberikan ketepatan klasifikasi terbaik dalam memprediksi kondisi keuangan perusahaan sektor industri di Indonesia. Kemudian diikuti oleh

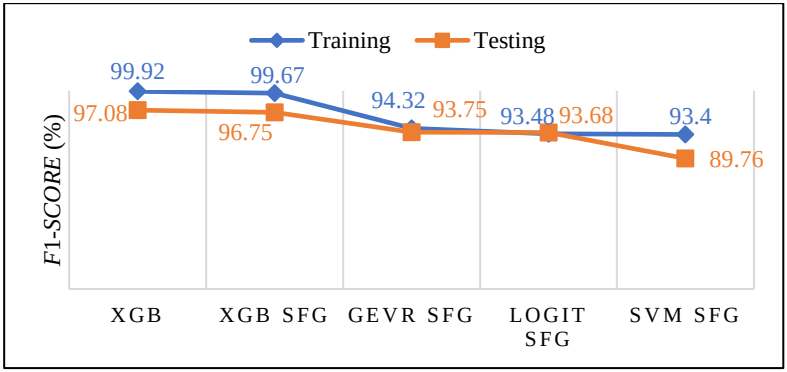
model *Extreme Gradient Boosting* (XGB) dengan *synthetic features generation* yang juga memiliki performansi model yang cukup baik. Jika dibandingkan dengan akurasi model dari penelitian Savera (2020) yang ditampilkan pada Lampiran 11, maka dapat diketahui bahwa model *Extreme Gradient Boosting* dengan dan tanpa melibatkan *synthetic features generation* secara serentak memberikan akurasi terbaik pada tiap *size* dibandingkan model *Generalized Extreme Value Regression* (GEVR) tanpa melibatkan *synthetic features generation* yang merupakan model prediksi terbaik pada penelitian tersebut. Ilustrasi perbandingan model secara serentak pada *size* 0 ditunjukkan pada Gambar 4.1.



(a)



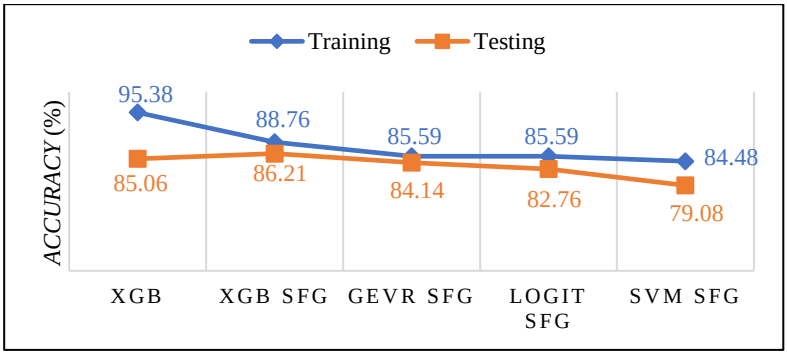
(b)



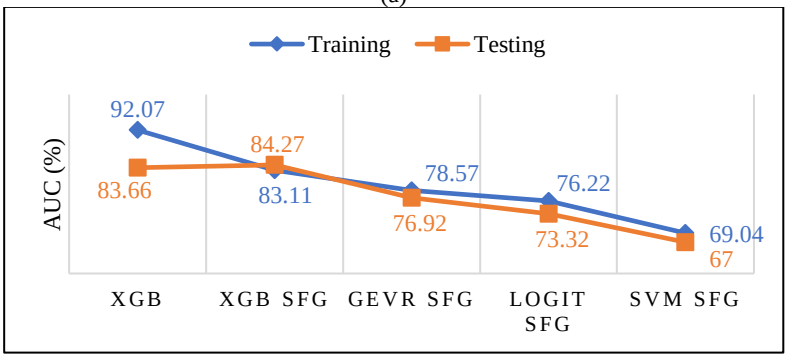
(c)

Gambar 4.1 Perbandingan *Accuracy* (a), AUC (b), dan *F*1-*score* (c) Model Secara Serentak Untuk *Size* 0.

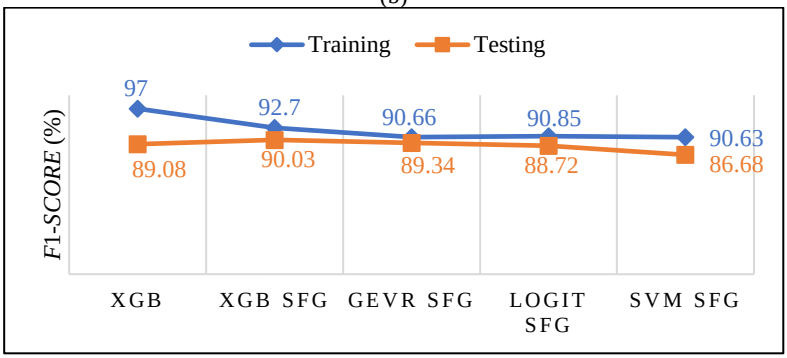
Gambar 4.1 (a), (b), dan (c) menunjukkan perbandingan nilai *accuracy*, AUC, dan *F*1-*score* dari evaluasi hasil klasifikasi menggunakan data *training* dan data *testing* pada tiap model secara serentak untuk *size* 0. Ketiga nilai kebaikan model tersebut memiliki karakteristik yang berbeda. *Accuracy* menunjukkan rata-rata ketepatan hasil klasifikasi dengan asumsi data memiliki kategori seimbang (*balance*). Sedangkan untuk data dengan kategori yang tidak seimbang (*imbalanced*) sebaiknya dievaluasi menggunakan AUC dan *F*1-*score*. Perbedaan nilai AUC dan *F*1-*score* terletak pada kelebihan model dalam memprediksi kondisi keuangan perusahaan yang sehat dan yang mengalami *financial distress*. Semakin baik model dalam memprediksi kondisi perusahaan yang mengalami *financial distress*, maka nilai AUC akan semakin baik. Sebaliknya apabila model semakin baik dalam memprediksi kondisi perusahaan yang sehat, maka nilai *F*1-*score* akan semakin baik. Berdasarkan Gambar 4.1 dapat dilihat bahwa model *Extreme Gradient Boosting* (XGB) tanpa melibatkan *synthetic features generation* memiliki nilai *accuracy*, AUC, dan *F*1-*score* terbaik dibandingkan model lainnya. Kemudian ilustrasi perbandingan model secara serentak untuk *size* 1 ditunjukkan pada Gambar 4.2.



(a)



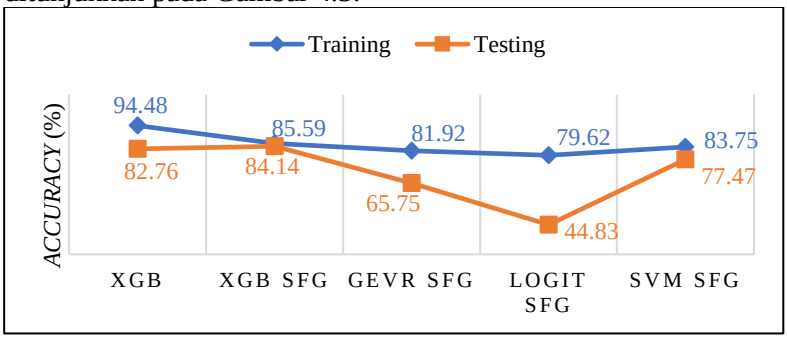
(b)



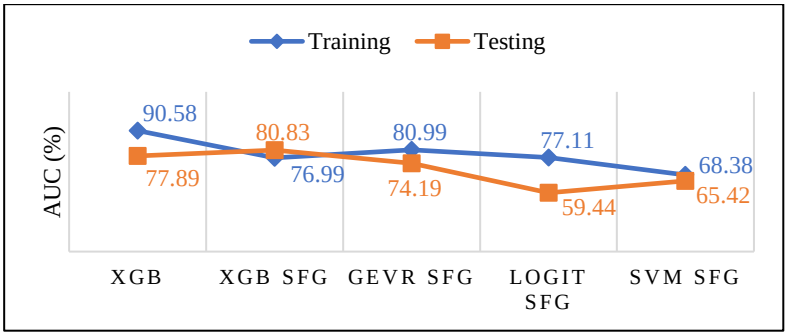
(c)

Gambar 4.2 Perbandingan *Accuracy* (a), AUC (b), dan *F1-score* (c) Model Secara Serentak Untuk *Size* 1.

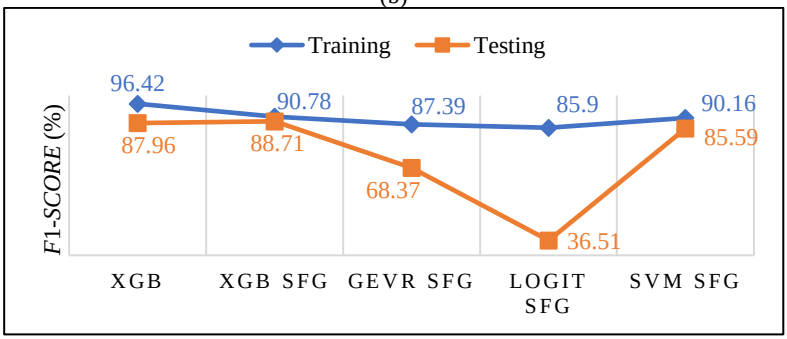
Gambar 4.2 (a), (b), dan (c) menunjukkan perbandingan nilai *accuracy*, AUC, dan *F*1-*score* pada tiap model secara serentak untuk *size* 1. Berdasarkan Gambar 4.2 tersebut dapat diketahui bahwa model *Extreme Gradient Boosting* (XGB) tanpa melibatkan *synthetic features generation* merupakan model terbaik dalam memprediksi kondisi keuangan perusahaan sektor industri di Indonesia dengan nilai *accuracy*, AUC, dan *F*1-*score* terbaik dibandingkan dengan model lainnya. Model terbaik selanjutnya ialah model *Extreme Gradient Boosting* (XGB) dengan melibatkan *synthetic features generation*. Kelebihan model tersebut dibandingkan model *Extreme Gradient Boosting* (XGB) tanpa melibatkan *synthetic features generation* adalah hasil evaluasi menggunakan data *training* dan data *testing* cenderung tidak jauh berbeda. Model secara serentak untuk *size* 1 yang memiliki nilai *accuracy*, AUC, dan *F*1-*score* terendah dibandingkan model lainnya ialah model *Support Vector Machine* (SVM) dengan melibatkan *synthetic features generation* dibandingkan model lainnya atau dapat dikatakan bahwa model *Support Vector Machine* (SVM) dengan melibatkan *synthetic features generation* lemah dalam memprediksi kondisi keuangan perusahaan sektor industri di Indonesia dengan menggunakan rasio keuangan dan indikator ekonomi makro pada 1 tahun sebelumnya. Selanjutnya ilustrasi perbandingan model secara serentak untuk *size* 2 ditunjukkan pada Gambar 4.3.



(a)



(b)

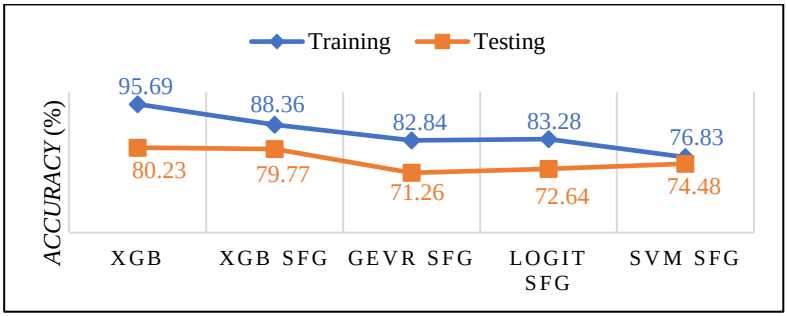


(c)

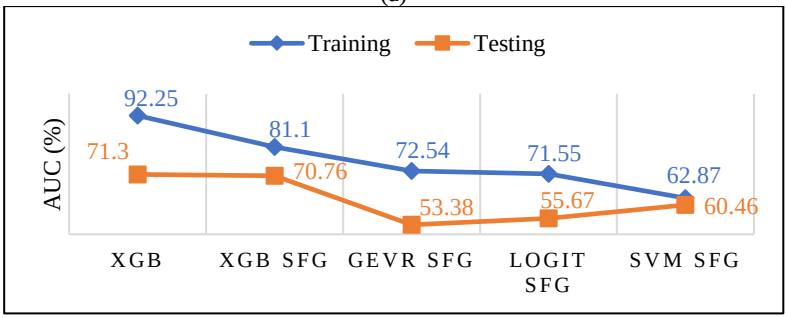
Gambar 4.3 Perbandingan *Accuracy* (a), AUC (b), dan *F*1-*score* (c) Model Secara Serentak Untuk *Size* 2.

Gambar 4.3 (a), (b), dan (c) menunjukkan perbandingan nilai *accuracy*, AUC, dan *F*1-*score* model secara serentak untuk *size* 2. Model *Extreme Gradient Boosting* (XGB) tanpa melibatkan *synthetic features generation* merupakan model prediksi terbaik pada *size* 2 berdasarkan nilai *accuracy*, AUC, dan *F*1-*score*. Model terbaik selanjutnya ialah model *Extreme Gradient Boosting* (XGB) dengan melibatkan *synthetic features generation*. Model *Extreme Gradient Boosting* (XGB) dengan melibatkan *synthetic features generation* merupakan model yang paling konsisten dibandingkan model lainnya, karena hasil evaluasi menggunakan data *training* dan data *testing* tidak terlalu jauh berbeda. Model yang memiliki nilai *accuracy*, AUC, dan *F*1-*score* terendah dan cenderung *overfitting* pada *size* 2 ialah model *Logistic Regression* dengan

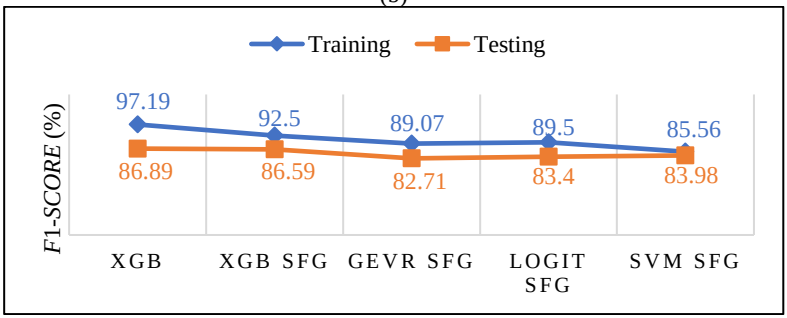
melibatkan *synthetic features generation*. Kemudian ilustrasi perbandingan model secara serentak untuk *size* 3 ditunjukkan pada Gambar 4.4 berikut.



(a)



(b)



(c)

Gambar 4.4 Perbandingan *Accuracy* (a), AUC (b), dan *F*1-*score* (c) Model Secara Serentak Untuk *Size* 3.

Gambar 4.4 (a), (b), dan (c) menunjukkan perbandingan nilai *accuracy*, AUC, dan *F*1-*score* hasil evaluasi menggunakan data *training* dan *testing* pada tiap model secara serentak untuk *size* 3. Berdasarkan Gambar 4.4 dapat diketahui bahwa model *Extreme Gradient Boosting* (XGB) tanpa melibatkan *synthetic features generation* merupakan model terbaik dalam memprediksi kondisi keuangan perusahaan sektor industri di Indonesia menggunakan data rasio keuangan dan indikator ekonomi makro pada 3 tahun sebelumnya. Sedangkan model yang paling konsisten atau dapat dikatakan hasil evaluasi menggunakan data *training* dan *testing* tidak jauh berbeda ialah model *Support Vector Machine* (SVM) dengan melibatkan *synthetic features generation*.

4.3.2 Perbandingan Model dengan Seleksi Variabel

Pada penelitian ini model yang dilakukan dengan seleksi variabel ialah model *Extreme Gradient Boosting* (XGB) dengan dan tanpa melibatkan *synthetic features generation*, *stepwise Generalized Extreme Value Regression* (GEVR), *stepwise Logistic Regression*, *Lasso Logistic Regression*, *Elastic-Net Logistic Regression*, *Lasso Support Vector Machine*, dan *Elastic-Net Support Vector Machine*. Perbandingan evaluasi hasil klasifikasi model dengan seleksi variabel ditunjukkan pada Tabel 4.45.

Tabel 4.45 Perbandingan Model dengan Seleksi Variabel.

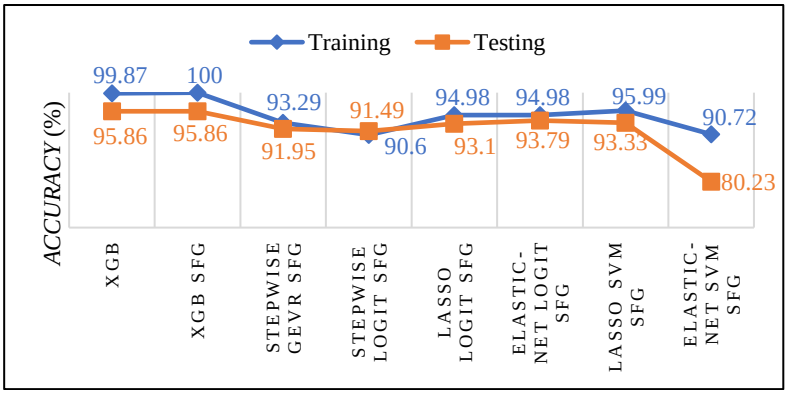
Model	***Size***	***Accuracy* (%)**		**AUC (%)**		***F*1-*score* (%)**	
		Train	***Test***	***Train***	***Test***	***Train***	***Test***
XGB	0	99,87	95,86	99,74	94,22	99,92	97,08
	1	95,79	84,83	92,63	81,76	97,27	89,18
	2	94,87	84,60	91,15	79,21	96,67	89,38
	3	94,22	81,15	89,65	73,70	96,25	87,27
XGB SFG	0	100,00	95,86	100,00	94,22	100,00	97,08
	1	94,62	88,28	90,49	85,31	96,53	91,71
	2	93,41	84,60	89,01	78,99	95,73	89,42
	3	92,24	80,92	87,28	72,66	94,96	87,25
Stepwise GEVR SFG	0	93,29	91,95	92,09	90,77	95,54	94,21
	1	86,55	84,37	80,58	77,30	91,23	89,47
	2	82,45	77,70	79,31	79,70	88,05	82,40
	3	81,98	72,64	68,37	55,01	88,84	83,54

Tabel 4.45 Perbandingan Model dengan Seleksi Variabel (Lanjutan).

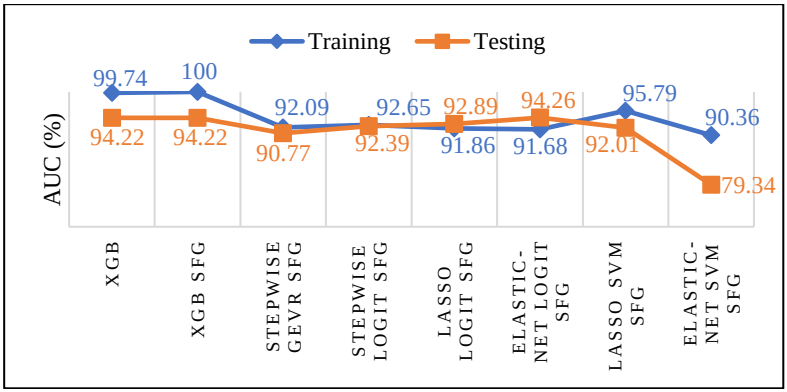
Model	***Size***	***Accuracy* (%)**		**AUC (%)**		***F*1-*score* (%)**	
		Train	***Test***	***Train***	***Test***	***Train***	***Test***
Stepwise LOGIT SFG	0	90,60	91,49	92,65	92,39	93,48	93,68
	1	85,10	82,53	75,32	72,95	90,57	88,59
	2	84,90	81,61	73,96	79,46	90,52	86,58
	3	82,24	74,02	70,17	58,17	88,85	84,06
Lasso LOGIT SFG	0	94,98	93,10	91,86	92,89	96,74	94,98
	1	83,93	86,21	72,39	83,62	89,94	90,13
	2	82,76	82,53	69,76	77,72	89,30	87,78
	3	82,24	77,24	70,40	65,69	88,83	85,33
Elastic-Net LOGIT SFG	0	94,98	93,79	91,68	94,26	96,75	95,45
	1	85,03	83,68	75,08	81,15	90,54	88,23
	2	83,91	78,16	72,45	80,47	89,91	82,70
	3	82,07	78,85	70,05	68,58	88,73	86,19
Lasso SVM SFG	0	95,99	93,33	95,79	92,01	97,40	95,22
	1	84,83	81,84	79,96	78,37	90,24	86,90
	2	78,47	70,11	72,74	61,71	87,10	81,27
	3	81,72	78,16	77,01	79,38	88,49	86,05
Elastic-Net SVM SFG	0	90,72	80,23	90,36	79,34	94,11	83,83
	1	78,41	71,72	77,49	68,76	87,39	82,75
	2	76,78	70,34	79,26	68,53	86,66	82,40
	3	78,10	72,87	78,68	69,73	87,13	83,09

Berdasarkan Tabel 4.45 dapat dilihat bahwa hasil evaluasi menggunakan data *training* terbaik ialah pada model *Extreme Gradient Boosting* dengan *synthetic features generation* pada *size* 0, sedangkan untuk hasil evaluasi menggunakan data *testing* terbaik dengan menggunakan model *Extreme Gradient Boosting* (XGB) dengan dan tanpa *synthetic features generation* pada *size* 0 memberikan performansi yang sama. Model dengan seleksi variabel terbaik untuk *size* 1 berdasarkan hasil evaluasi dengan data *training* ialah model *Extreme Gradient Boosting* (XGB) tanpa melibatkan *synthetic features generation*, sedangkan model terbaik untuk *size* 1 berdasarkan hasil evaluasi dengan data *testing* ialah model *Extreme Gradient Boosting* (XGB) dengan melibatkan *synthetic features generation*. Kemudian model dengan seleksi variabel terbaik untuk *size* 2 dan 3 ialah model *Extreme Gradient*

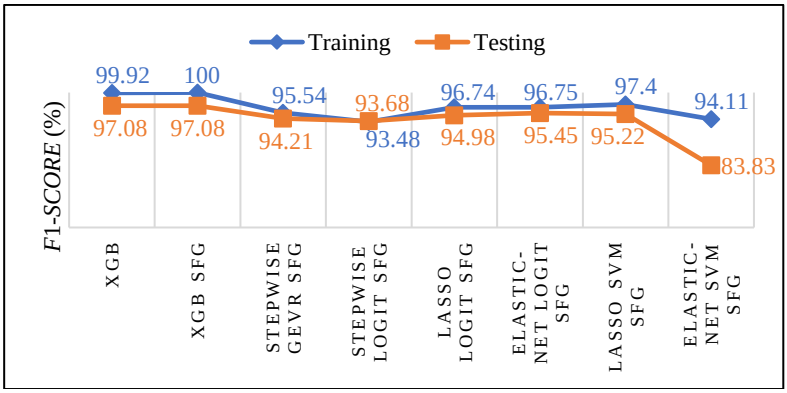
Boosting (XGB) tanpa melibatkan *synthetic features generation*. Model terbaik selanjutnya untuk *size* 2 dan 3 ialah model *Extreme Gradient Boosting* (XGB) dengan melibatkan *synthetic features generation*, dimana kelebihannya jika dibandingkan dengan model *Extreme Gradient Boosting* (XGB) yang tanpa melibatkan *synthetic features generation* ialah hasil evaluasi dengan data *training* dan data *testing* cenderung lebih dekat atau dapat dikatakan model lebih konsisten. Rasio keuangan dan variabel makro ekonomi yang berpengaruh signifikan dalam model *Extreme Gradient Boosting* (XGB) dengan *synthetic features generation* ditunjukkan pada Tabel 4.41. Jika dibandingkan dengan hasil penelitian Savera (2020) yang ditampilkan pada Lampiran 11, maka dapat disimpulkan bahwa model *Extreme Gradient Boosting* dengan dan tanpa melibatkan *synthetic features generation* dengan seleksi variabel pada tiap *size* memberikan performansi terbaik dalam memprediksi kondisi *financial distress* perusahaan sektor industri di Indonesia dibandingkan model *Generalized Extreme Value Regression* (GEVR) yang merupakan model terbaik dalam penelitian tersebut. Ilustrasi perbandingan model dengan seleksi variabel pada *size* 0 ditampilkan pada Gambar 4.5.



(a)



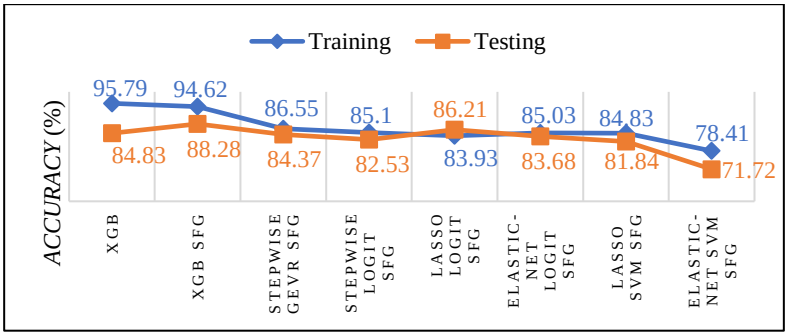
(b)



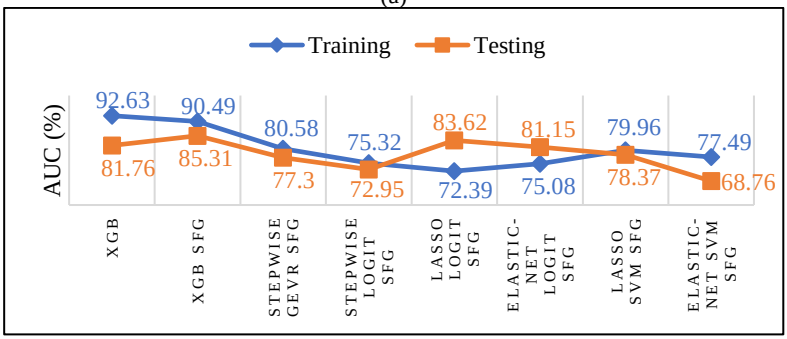
(c)

Gambar 4.5 Perbandingan *Accuracy* (a), AUC (b), dan *F*1-*score* (c) Model dengan Seleksi Variabel Untuk *Size* 0.

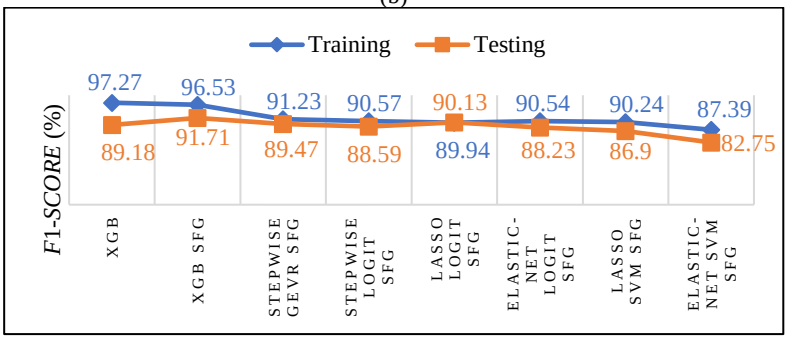
Gambar 4.5 (a), (b), dan (c) menunjukkan perbandingan nilai *accuracy*, AUC, dan *F*1-*score* dari evaluasi hasil klasifikasi menggunakan data *training* dan data *testing* pada masing-masing model untuk *size* 0. Berdasarkan nilai *accuracy*, AUC, dan *F*1-*score* dapat diketahui bahwa model *Extreme Gradient Boosting* (XGB) dengan *synthetic features generation* dan seleksi variabel memberikan ketepatan prediksi terbaik. Selanjutnya ilustrasi perbandingan *accuracy*, AUC, dan *F*1-*score* model dengan seleksi variabel pada *size* 1 ditunjukkan pada Gambar 4.6.



(a)



(b)

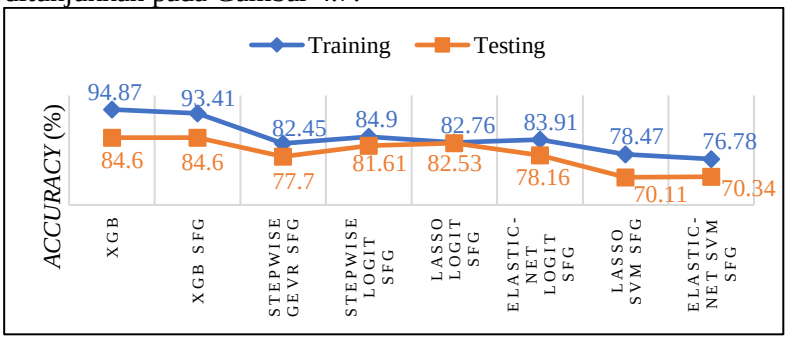


(c)

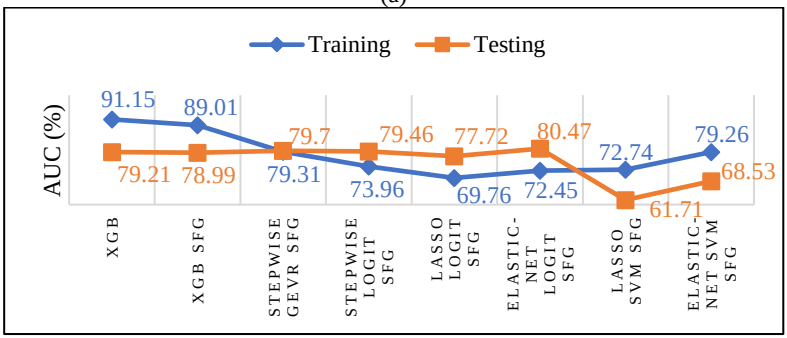
Gambar 4.6 Perbandingan *Accuracy* (a), AUC (b), dan *F1-score* (c) Model dengan Seleksi Variabel Untuk *Size* 1.

Gambar 4.6 (a), (b), dan (c) menunjukkan perbandingan hasil evaluasi model menggunakan data *training* dan data *testing*

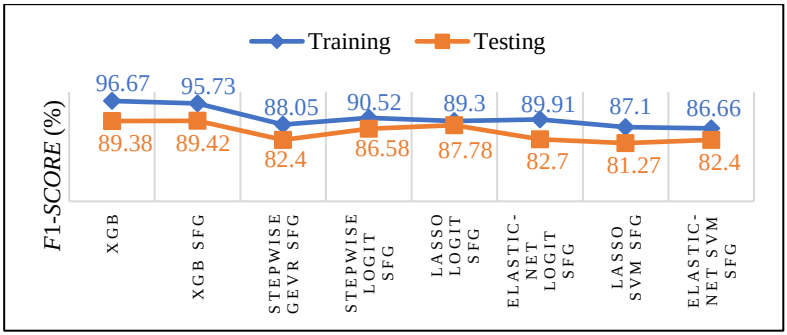
berdasarkan *accuracy*, AUC, dan *F*1-*score* pada tiap model dengan seleksi variabel untuk *size* 1. Hasil evaluasi model menggunakan data *training* menunjukkan bahwa model *Extreme Gradient Boosting* (XGB) tanpa melibatkan *synthetic features generation* merupakan model terbaik dalam memprediksi kondisi keuangan perusahaan sektor industri di Indonesia dibandingkan model dengan seleksi variabel lainnya. Sedangkan hasil evaluasi menggunakan data *testing* menunjukkan bahwa model *Extreme Gradient Boosting* (XGB) dengan melibatkan *synthetic features generation* memberikan hasil prediksi terbaik dibandingkan model prediksi dengan seleksi variabel lainnya. Kemudian ilustrasi perbandingan kebaikan model dengan seleksi variabel untuk *size* 2 ditunjukkan pada Gambar 4.7.



(a)



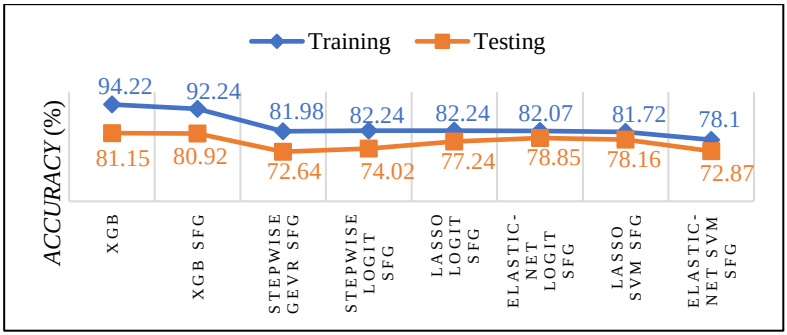
(b)



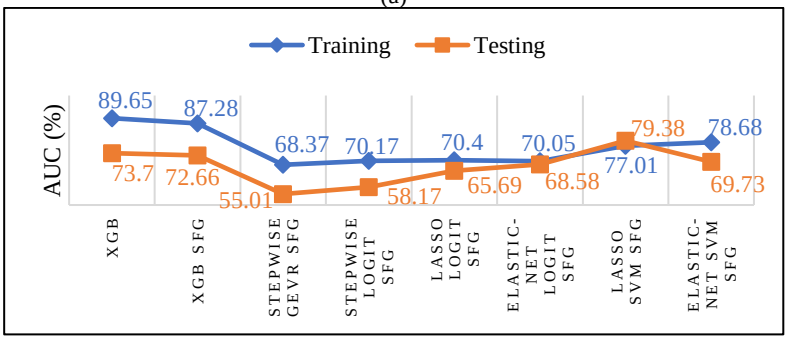
(c)

Gambar 4.7 Perbandingan *Accuracy* (a), AUC (b), dan *F*1-*score* (c) Model dengan Seleksi Variabel Untuk *Size* 2.

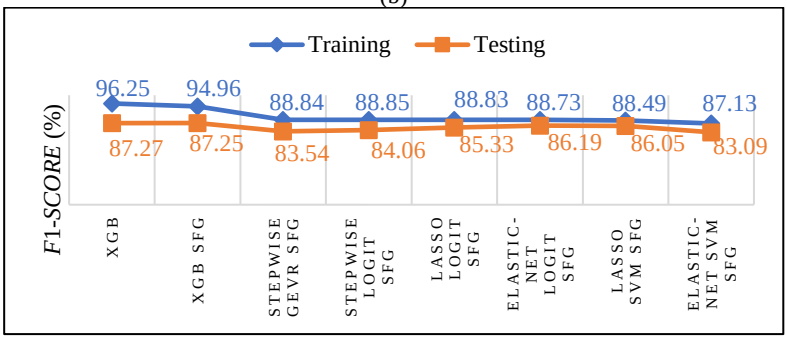
Berdasarkan Gambar 4.7 (a), (b), dan (c) dapat diketahui perbandingan hasil evaluasi pada tiap model dengan seleksi variabel untuk *size* 2 menggunakan data *training* dan data *testing*. Gambar 4.7 menunjukkan bahwa model *Extreme Gradient Boosting* tanpa melibatkan *synthetic features generation* merupakan model dengan akurasi terbaik dari hasil evaluasi menggunakan data *training*. Model *Extreme Gradient Boosting* (XGB) dengan melibatkan *synthetic features generation* berada diposisi kedua dari hasil evaluasi menggunakan data *training*. Jika dibandingkan dari hasil evaluasi menggunakan data *testing*, model *Extreme Gradient Boosting* (XGB) dengan melibatkan *synthetic features generation* lebih baik daripada tanpa terlibat karena perbandingan hasil evaluasi antara data *training* dan data *testing* lebih kecil. Adapun model yang paling konsisten atau memiliki perbandingan yang sangat kecil antara hasil evaluasi menggunakan data *training* dan data *testing* dibandingkan model lainnya pada *size* 2 ialah model *Lasso Logistic Regression* dengan melibatkan *synthetic features generation*. Selanjutnya ilustrasi perbandingan *accuracy*, AUC, dan *F*1-*score* pada tiap model dengan seleksi variabel untuk *size* 3 ditunjukkan pada Gambar 4.8 berikut.



(a)



(b)



(c)

Gambar 4.8 Perbandingan *Accuracy* (a), AUC (b), dan *F*1-*score* (c) Model dengan Seleksi Variabel Untuk *Size* 3.

Gambar 4.8 (a), (b), dan (c) menunjukkan perbandingan kebaikan model dengan seleksi variabel untuk *size* 3. Berdasarkan Gambar

4.8 dapat diketahui bahwa model *Extreme Gradient Boosting* (XGB) tanpa melibatkan *synthetic features generation* merupakan model terbaik dalam memprediksi kondisi keuangan perusahaan sektor industri di Indonesia untuk *size* 3 berdasarkan hasil evaluasi menggunakan data *training* dan data *testing*. Sedangkan model yang memberikan ketepatan klasifikasi terendah ialah pada model *Generalized Extreme Value Regression* (GEVR) dengan seleksi variabel menggunakan prosedur *stepwise* dan melibatkan *synthetic features generation*.

4.3.3 Perbandingan Model Secara Serentak dan Model dengan Seleksi Variabel

Pada penelitian ini data penelitian memiliki kategori yang tidak seimbang (*imbalanced*), sehingga salah satu ukuran untuk mengetahui kebaikan suatu model klasifikasi ialah dengan menghitung nilai *sensitivity*. *Sensitivity* merupakan rasio yang menunjukkan perbandingan antara prediksi benar positif (perusahaan dengan kondisi keuangan sehat) dan keseluruhan kelas aktual positif. Perbandingan nilai *sensitivity* pada model secara serentak dan model dengan seleksi variabel ditunjukkan pada Tabel 4.46 berikut.

Tabel 4.46 Perbandingan *Sensitivity* Model Secara Serentak dan Seleksi Variabel.

Model	***Size***	***Sensitivity* (%)**	
		Training	***Testing***
XGB	0	100,00	98,36
	1	98,46	87,17
	2	98,18	90,13
	3	99,08	93,75
XGB dengan Seleksi Variabel	0	100,00	98,36
	1	98,73	89,47
	2	98,38	92,76
	3	98,74	92,43
XGB SFG	0	99,84	98,03
	1	94,01	89,14
	2	93,72	89,14
	3	95,53	93,42

Tabel 4.46 Perbandingan *Sensitivity* Model Secara Serentak dan Seleksi Variabel (Lanjutan).

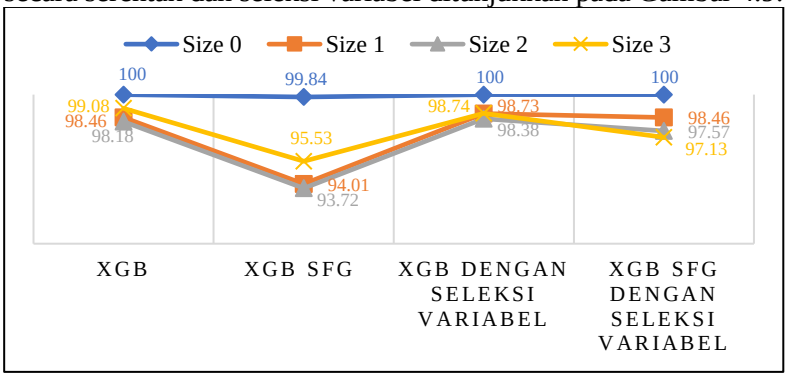
Model	***Size***	***Sensitivity* (%)**	
		Training	***Testing***
XGB SFG dengan Seleksi Variabel	0	100,00	98,36
	1	98,46	92,76
	2	97,57	93,09
	3	97,13	93,42
GEVR SFG	0	93,16	93,75
	1	92,10	95,07
	2	82,79	52,96
	3	93,00	98,36
Stepwise GEVR SFG	0	94,39	93,75
	1	92,10	95,07
	2	85,43	74,67
	3	95,41	99,34
LOGIT SFG	0	88,71	90,13
	1	94,28	97,04
	2	81,98	22,70
	3	94,84	98,36
Stepwise LOGIT SFG	0	88,71	90,13
	1	94,19	97,04
	2	95,24	84,87
	3	94,15	98,03
Lasso LOGIT SFG	0	97,86	93,42
	1	94,64	90,13
	2	95,04	89,80
	3	93,92	94,74
Elastic-Net LOGIT SFG	0	98,02	93,09
	1	94,28	87,50
	2	94,74	74,67
	3	93,92	94,41
SVM SFG	0	99,75	98,03
	1	98,82	97,37
	2	98,28	95,72
	3	99,20	95,72

Tabel 4.46 Perbandingan *Sensitivity* Model Secara Serentak dan Seleksi Variabel (Lanjutan).

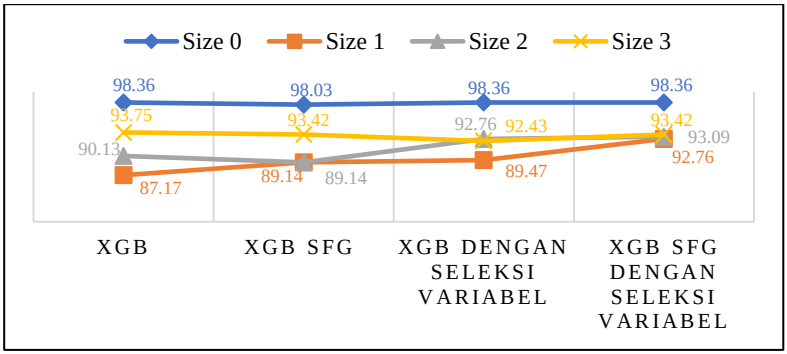
Model	***Size***	***Sensitivity* (%)**	
		Training	***Testing***
Lasso SVM SFG	0	96,14	95,38
	1	88,20	87,63
	2	79,68	72,31
	3	84,02	77,72
Elastic-Net SVM SFG	0	90,93	97,81
	1	78,51	72,13
	2	76,70	70,40
	3	78,04	73,60

Tabel 4.46 menunjukkan perbandingan nilai *sensitivity* model secara serentak dan seleksi variabel pada tiap *size*. Pada *size* 0, hasil evaluasi menggunakan data *training* dan data *testing* menunjukkan bahwa model *Extreme Gradient Boosting* (XGB) tanpa melibatkan *synthetic features generation* secara serentak dan seleksi variabel serta model *Extreme Gradient Boosting* (XGB) dengan melibatkan *synthetic features generation* dan seleksi variabel memberikan nilai *sensitivity* terbaik dibandingkan model lainnya. Hal tersebut menunjukkan bahwa model tersebut sangat baik dalam memprediksi kondisi keuangan perusahaan sektor industri di Indonesia yang sehat menggunakan data rasio keuangan dan indikator ekonomi makro pada tahun yang sama. Selanjutnya pada *size* 1, hasil evaluasi menggunakan data *training* dan data *testing* menunjukkan bahwa model *Support Vector Machine* (SVM) dengan melibatkan *synthetic features generation* merupakan model terbaik berdasarkan nilai *sensitivity* atau dapat dikatakan bahwa model *Support Vector Machine* (SVM) dengan melibatkan *synthetic features generation* sangat baik dalam memprediksi kondisi keuangan perusahaan sektor industri di Indonesia yang sehat menggunakan data rasio keuangan dan indikator ekonomi makro satu tahun sebelumnya. Kemudian hasil evaluasi model pada *size* 2 menggunakan data *training* menunjukkan bahwa model *Extreme Gradient Boosting* (XGB) tanpa melibatkan *synthetic features generation* dengan seleksi

variabel merupakan model dengan *sensitivity* terbaik dibandingkan model lainnya pada *size* 2, sedangkan hasil evaluasi menggunakan data *testing* menunjukkan bahwa model *Support Vector Machine* (SVM) dengan melibatkan *synthetic features generation* merupakan model dengan *sensitivity* terbaik. Selanjutnya pada *size* 3, hasil evaluasi menggunakan data *training* menunjukkan bahwa model *Support Vector Machine* (SVM) dengan melibatkan *synthetic features generation* merupakan model dengan *sensitivity* terbaik dibandingkan model lainnya pada *size* 3. Sedangkan hasil evaluasi menggunakan data *testing* menunjukkan bahwa model *Generalized Extreme Value Regression* (GEVR) dengan melibatkan *synthetic features generation* dan seleksi variabel dengan prosedur *stepwise* merupakan model dengan *sensitivity* terbaik dibandingkan model lainnya. Hal tersebut menunjukkan bahwa model tersebut baik dalam memprediksi kondisi keuangan perusahaan sektor industri di Indonesia yang sehat dengan menggunakan data rasio keuangan dan indikator ekonomi makro tiga tahun sebelumnya. Selanjutnya dilakukan perbandingan model secara serentak dan model dengan seleksi variabel untuk mengetahui seberapa besar pengaruh seleksi variabel dalam meningkatkan hasil prediksi berdasarkan nilai *sensitivity*. Ilustrasi perbandingan *sensitivity* model *Extreme Gradient Boosting* (XGB) tanpa dan dengan melibatkan *synthetic features generation* serta secara serentak dan seleksi variabel ditunjukkan pada Gambar 4.9.



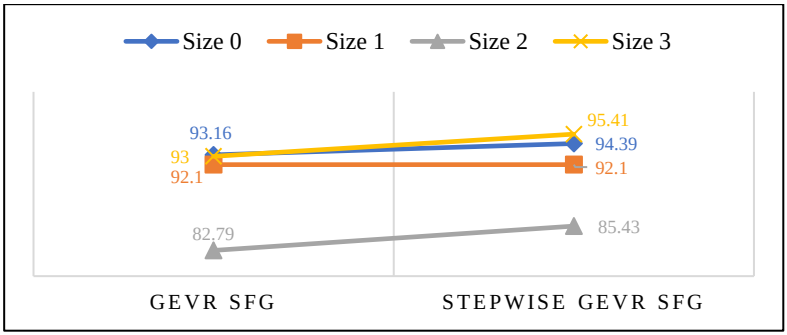
(a)



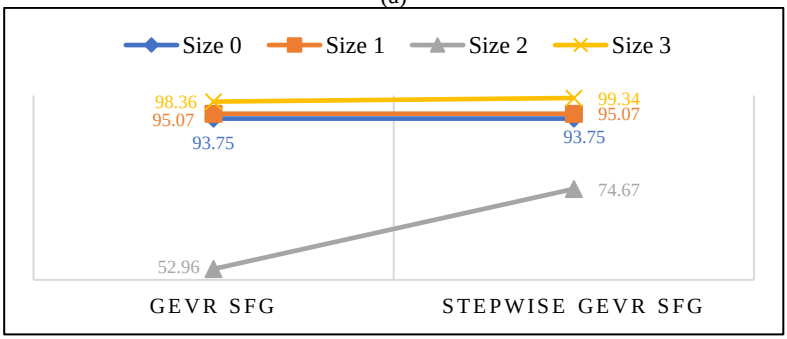
(b)

Gambar 4.9 Perbandingan *Sensitivity* Model XGB dengan dan Tanpa *Synthetic Features Generation* Secara Serentak dan Seleksi Variabel Menggunakan Data *Training* (a) dan Data *Testing* (b).

Berdasarkan Gambar 4.9 dapat diketahui bahwa pengaruh *synthetic features generation* pada model *Extreme Gradient Boosting* (XGB) tidak selalu meningkatkan *sensitivity* model pada tiap *size*. Pengaruh *synthetic features generation* pada model *Extreme Gradient Boosting* terlihat pada hasil evaluasi menggunakan data *testing* pada model yang telah dilakukan seleksi variabel yaitu pada *size* 1 dan 2, sedangkan pada *size* 0 cenderung sama dan *size* 3 cenderung turun. Gambar 4.9 juga menunjukkan bahwa seleksi variabel berdasarkan nilai *gain* mampu meningkatkan nilai *sensitivity* model *Extreme Gradient Boosting* dengan dan tanpa melibatkan *synthetic features generation*. Peningkatan hasil klasifikasi akibat *synthetic features generation* dan seleksi variabel pada *size* 0, 1, dan 2 sama jika dibandingkan berdasarkan nilai *accuracy*, AUC, dan *F*1-*score* tetapi terjadi perbedaan pada *size* 3. Hal tersebut menunjukkan bahwa model *Extreme Gradient Boosting* (XGB) dengan melibatkan *synthetic features generation* dan seleksi variabel pada *size* 3 cenderung meningkatkan ketepatan model dalam memprediksi kelas negatif (kondisi *financial distress*). Kemudian ilustrasi perbandingan nilai *sensitivity* model *Generalized Extreme Value Regression* (GEVR) dengan melibatkan *synthetic features generation* secara serentak dan seleksi variabel ditunjukkan pada Gambar 4.10.



(a)

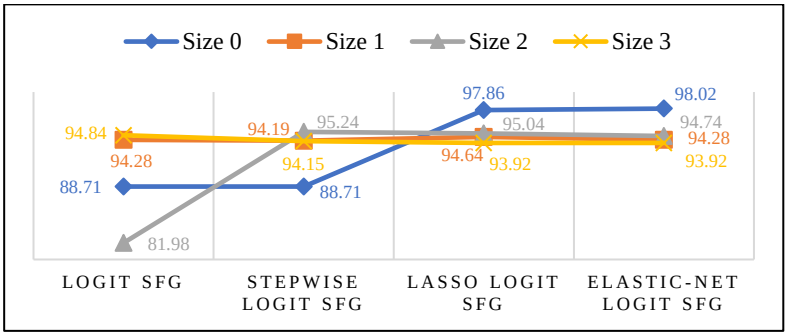


(b)

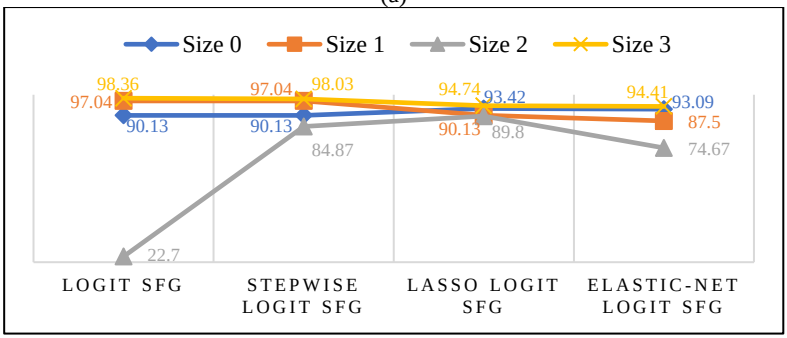
Gambar 4.10 Perbandingan *Sensitivity* Model GEVR Secara Serentak dan Seleksi Variabel Menggunakan Data *Training* (a) dan Data *Testing* (b).

Gambar 4.10 menunjukkan bahwa seleksi variabel menggunakan prosedur *stepwise* pada model *Generalized Extreme Value Regression* (GEVR) dengan melibatkan *synthetic features generation* meningkatkan nilai *sensitivity* model *Generalized Extreme Value Regression* (GEVR) dengan melibatkan *synthetic features generation* pada tiap *size*. Peningkatan nilai *sensitivity* yang cukup signifikan akibat dilakukannya seleksi variabel pada model *Generalized Extreme Value Regression* (GEVR) dengan melibatkan *synthetic features generation* terjadi pada *size* 2. Hasil tersebut juga sama jika dibandingkan dengan peningkatan nilai *accuracy*, AUC, dan *F*1-*score* dari hasil evaluasi menggunakan data *testing* yang ditampilkan pada Tabel 4.44 dan Tabel 4.45. Sehingga dari hasil tersebut dapat disimpulkan bahwa seleksi

variabel pada model *Generalized Extreme Value Regression* (GEVR) dapat meningkatkan performansi model dalam memprediksi kondisi keuangan perusahaan sektor industri di Indonesia untuk semua *size*. Selanjutnya ilustrasi perbandingan model *Logistic Regression* dengan melibatkan *synthetic features generation* secara serentak dan seleksi variabel ditampilkan pada Gambar 4.11.



(a)

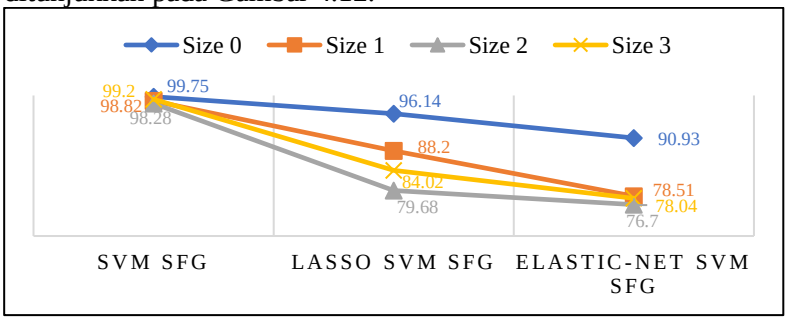


(b)

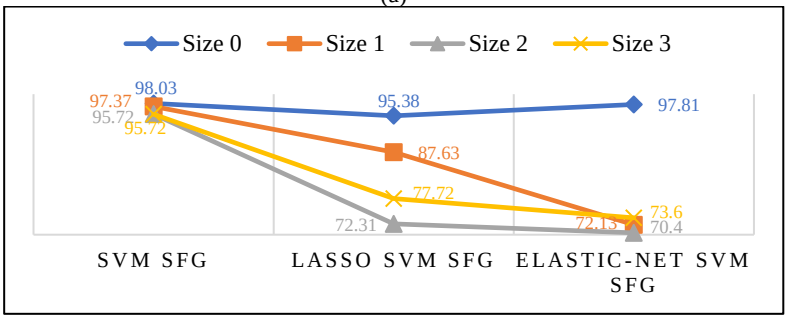
Gambar 4.11 Perbandingan *Sensitivity* Model LOGIT Secara Serentak dan Seleksi Variabel menggunakan Data *Training* (a) dan Data *Testing* (b).

Berdasarkan Gambar 4.11 dapat diketahui bahwa seleksi variabel pada model *Logistic Regression* dengan melibatkan *synthetic features generation* menggunakan prosedur *stepwise*, *Lasso*, dan *Elastic-Net* dapat meningkatkan nilai *sensitivity* model untuk *size* 0 dan 2, sedangkan terjadi penurunan nilai *sensitivity* pada model dengan *size* 1 dan 3. Peningkatan nilai *sensitivity* yang

cukup signifikan akibat seleksi variabel pada model *Logistic Regression* dengan melibatkan *synthetic features generation* terjadi pada *size* 2. Jika dibandingkan dengan pengaruh seleksi variabel berdasarkan nilai *accuracy*, AUC, dan *F*1-*score* dari hasil evaluasi menggunakan data *testing* yang ditampilkan pada Tabel 4.44 dan Tabel 4.45, maka dapat diketahui bahwa seleksi variabel pada model *Logistic Regression* dengan melibatkan *synthetic features generation* cenderung meningkatkan ketepatan model dalam memprediksi kelas negatif (kondisi *financial distress*) yang mengakibatkan nilai *accuracy*, AUC, dan *F*1-*score* cenderung meningkat untuk tiap *size*. Kemudian ilustrasi perbandingan nilai *sensitivity* model *Support Vector Machine* (SVM) pada tiap *size* ditunjukkan pada Gambar 4.12.



(a)



(b)

Gambar 4.12 Perbandingan *Sensitivity* Model SVM Secara Serentak dan Seleksi Variabel Menggunakan Data *Training* (a) dan Data *Testing* (b).

Gambar 4.12 menunjukkan bahwa seleksi variabel dalam model *Support Vector Machine* (SVM) dengan melibatkan *synthetic features generation* menggunakan metode *Lasso* dan *Elastic-Net* cenderung menurunkan nilai *sensitivity* pada model *Support Vector Machine* (SVM) dengan melibatkan *synthetic features generation* secara serentak pada tiap *size*. Jika dibandingkan dengan pengaruh seleksi variabel berdasarkan nilai *accuracy*, AUC, dan *F*1-*score* pada Tabel 4.44 dan Tabel 4.45, maka dapat diketahui bahwa seleksi variabel pada model *Support Vector Machine* (SVM) dengan melibatkan *synthetic features generation* cenderung meningkatkan ketepatan model dalam memprediksi kelas negatif (kondisi *financial distress*) yang mengakibatkan nilai *accuracy*, AUC, dan *F*1-*score* cenderung meningkat untuk tiap *size*.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan pada penelitian ini, maka diperoleh kesimpulan sebagai berikut:

1. Prediksi *financial distress* perusahaan sektor industri di Indonesia dengan menggunakan model *Extreme Gradient Boosting* (XGB) tanpa melibatkan *synthetic features generation* secara serentak dan seleksi variabel menghasilkan ketepatan prediksi terbaik pada *size* 0. Hasil evaluasi model *Extreme Gradient Boosting* (XGB) dengan seleksi variabel menggunakan data *testing* pada *size* 0 menunjukkan nilai *accuracy* sebesar 95,86%, AUC sebesar 94,22%, dan *F*1-*score* sebesar 97,08%. Kemudian pada *size* 1 menunjukkan nilai *accuracy* sebesar 85,06%, AUC sebesar 83,66%, dan *F*1-*score* sebesar 89,08%. Selanjutnya pada *size* 2 menunjukkan nilai *accuracy* sebesar 82,76%, AUC sebesar 77,89%, dan *F*1-*score* sebesar 87,96%. Sedangkan pada *size* 3 menunjukkan nilai *accuracy* sebesar 80,23%, AUC sebesar 71,30%, dan *F*1-*score* sebesar 86,89%.
2. Prediksi *financial distress* perusahaan sektor industri di Indonesia menggunakan metode klasifikasi dan melibatkan *synthetic features generation* dilakukan secara serentak dan seleksi variabel. Model yang dilakukan secara serentak dengan melibatkan *synthetic features generation* ialah model *Extreme Gradient Boosting* (XGB), *Generalized Extreme Value Regression* (GEVR), *Logistic Regression*, dan *Support Vector Machine* (SVM). Sedangkan model yang dilakukan dengan seleksi variabel dan melibatkan *synthetic features generation* ialah model *Extreme Gradient Boosting* (XGB),

stepwise Generalized Extreme Value Regression, *stepwise Logistic Regression*, *Lasso Logistic Regression*, *Elastic-Net Logistic Regression*, *Lasso Support Vector Machine*, dan *Elastic-Net Support Vector Machine*. Berdasarkan hasil evaluasi ketepatan klasifikasi dapat disimpulkan bahwa data dengan *time window size* 0 membentuk model terbaik dalam memprediksi *financial distress* perusahaan sektor industri di Indonesia. Hal tersebut menunjukkan bahwa semakin besar *time window size* yang digunakan akan menurunkan performansi model dalam memprediksi *financial distress* perusahaan sektor industri di Indonesia.

3. Berdasarkan nilai *accuracy*, AUC, dan *F*1-*score* dari hasil evaluasi model menggunakan data *testing* pada *size* 0 dan *size* 1, disimpulkan bahwa model *Extreme Gradient Boosting* (XGB) dengan melibatkan *synthetic features generation* dan seleksi variabel menghasilkan ketepatan klasifikasi terbaik dalam memprediksi kondisi *financial distress* perusahaan sektor industri di Indonesia, sedangkan pada *size* 2 dan *size* 3 didapatkan bahwa model *Extreme Gradient Boosting* (XGB) tanpa melibatkan *synthetic features generation* dengan seleksi variabel merupakan model dengan ketepatan klasifikasi terbaik untuk memprediksi kondisi *financial distress* perusahaan sektor industri di Indonesia. Kemudian berdasarkan hasil analisis diketahui bahwa pengaruh *synthetic features generation* dan seleksi variabel meningkatkan performansi klasifikasi pada model *Extreme Gradient Boosting* (XGB). Hasil evaluasi menggunakan data *testing* pada *size* 0 menunjukkan nilai *accuracy* sebesar 95,86%, AUC sebesar 94,22%, dan *F*1-*score* sebesar 97,08%. Kemudian pada *size* 1 menunjukkan nilai *accuracy* sebesar 88,28%, AUC sebesar 85,31%, dan

*F*1-*score* sebesar 91,71%. Selanjutnya pada *size* 2 menunjukkan nilai *accuracy* sebesar 84,60%, AUC sebesar 78,99%, dan *F*1-*score* sebesar 89,42%. Sedangkan pada *size* 3 menunjukkan nilai *accuracy* sebesar 80,92%, AUC sebesar 72,66%, dan *F*1-*score* sebesar 87,25%. Berdasarkan nilai *sensitivity* dari hasil evaluasi menggunakan data *testing* dapat disimpulkan bahwa pada *size* 0 model *Extreme Gradient Boosting* (XGB) tanpa melibatkan *synthetic features generation* secara serentak dan seleksi variabel serta model *Extreme Gradient Boosting* (XGB) dengan melibatkan *synthetic features generation* dan seleksi variabel merupakan model terbaik. Kemudian model dengan *sensitivity* terbaik pada *size* 1 dan 2 ialah model *Support Vector Machine* (SVM) dengan melibatkan *synthetic features generation* secara serentak. Sedangkan model dengan *sensitivity* terbaik pada *size* 3 ialah model *Generalized Extreme Value Regression* (GEVR) dengan melibatkan *synthetic features generation* dan seleksi variabel dengan prosedur *stepwise*. Selanjutnya hasil analisis juga menunjukkan bahwa terjadi perbedaan pengaruh seleksi variabel dari hasil evaluasi menggunakan data *testing* berdasarkan nilai *accuracy*, AUC, dan *F*1-*score* dengan nilai *sensitivity* pada model *Extreme Gradient Boosting* (XGB), *Logistic Regression*, dan *Support Vector Machine* (SVM). Hal ini menunjukkan bahwa pengaruh seleksi variabel pada model tersebut cenderung meningkatkan performansi model dalam memprediksi kelas negatif atau kondisi *financial distress* yang mengakibatkan nilai *accuracy*, AUC, dan *F*1-*score* cenderung meningkat dibandingkan nilai *sensitivity* untuk tiap *size*.

5.2 Saran

Berdasarkan kesimpulan yang diperoleh, maka saran yang diberikan ialah sebagai berikut:

1. Saran untuk perusahaan sektor industri di Indonesia dan pihak-pihak yang dapat dirugikan akibat bangkrutnya perusahaan ialah untuk memperhatikan rasio-rasio keuangan seperti rasio aktivitas, profitabilitas, solvabilitas, dan likuiditas yang berpengaruh signifikan dalam model *Extreme Gradient Boosting* dengan *synthetic features generation* untuk memprediksi kondisi perusahaan sektor industri yang mengalami *financial distress*. Selain itu diperlukan peran pemerintah dalam meningkatkan indikator ekonomi makro di Indonesia untuk meminimalisir perusahaan yang mengalami *financial distress*.
2. Saran untuk penelitian selanjutnya yaitu untuk menambahkan beberapa metode klasifikasi lainnya yang mungkin dapat meningkatkan ketepatan prediksi *financial distress* perusahaan sektor industri di Indonesia. Pada penelitian ini penentuan kategori perusahaan dengan kondisi *financial distress* dilihat dari nilai ICR < 1 atau ROA < 0, sehingga untuk penelitian selanjutnya disarankan untuk dapat mengkategorikan perusahaan yang mengalami *financial distress* apabila ICR < 1 dan ROA < 0. Selain itu untuk penelitian selanjutnya dapat menggunakan konsep *drift multiple* yaitu menggabungkan dua atau lebih *size* menjadi satu. Kemudian peneliti juga menyarankan untuk menggunakan *synthetic features generation* dengan kombinasi tiga atau lebih variabel yang mungkin dapat meningkatkan ketepatan prediksi *financial distress* perusahaan sektor industri di Indonesia.

DAFTAR PUSTAKA

Afriansyah, F. (2017). *Klasifikasi Aktivitas Gerakan Bayi Setelah Operasi Berbasis Motion Vector Menggunakan Support Vector Machine.* Surabaya: Fakultas Teknologi Elektro, ITS.

Afriyeni, E. (2013). Rasio Keuangan, Financial Distress Perusahaan : Suatu Tinjauan. *14*, 41-51.

Agresti, A. (2013). *Categorical Data Analysis Third Edition.* New Jersey: John Wiley & Sons, Inc.

Almilia, L. S. (2006). Prediksi Kondisi Financial Distress Perusahaan Go-Public dengan Menggunakan Analisis Multinomial Logit. *Jurnal Ekonomi dan Bisnis*, *12*.

Asquith, P., Gertner, R., & Scharfstein, D. (1994). Anatomy of Financial Distress: An Examination of Junk-Bond Issuers. *Quarterly Journal of Economics*, 1189-1222.

Awad, M., & Khanna, R. (2015). Support Vector Machines for Classification. *Efficient Learning Machines*, 39-66.

Batista, G., & Monard, M. (2002). *A Study of K-Nearest Neighbour as an Imputation Method.* Sao Paulo: University of Sao Paulo.

Beaver, W. H., Correia, M., & McNichols, M. F. (2010). Financial Statement Analysis and the Prediction of Financial Distress. *Foundations and Trends in Accounting*, *5*, 99-173.

Calabrese, R., & Osmetti, S. A. (2011). Generalized Extreme Value Regression for Binary Rare Events Data: an Application to Credit Defaults. *UCD Geary Institute Discussion Paper Series*.

Carolina, V., Marpaung, E. I., & Pratama, D. (2017). Analisis Rasio Keuangan untuk Memprediksi Kondisi Financial DIstress (Studi Empiris pada Perusahaan Manufaktur yang Terdaftar di Bursa Efek Indonesia Periode 2014-2015). *Jurnal Akuntansi Maranatha*, *9*, 137-145.

Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System.* San Francisco: University of Washington.

Dwijayanti, S. F. (2010). Penyebab, Dampak, dan Prediksi dari Financial Distress Serta Solusi untuk Mengatasi Financial Distress. *Jurnal Akuntansi Kontemporer, 2*, 191-205.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Path for Generalized Linear Models Via Coordinate Descent. *Journal of Statistical Software 33 (1)*, 1-22.

Goodman, L. (1971). The Analysis of Multidimensional Contingency Tables: Stepwise Procedures and Direct Estimation for Building Models for Multiple Classifications. *Technometrics, 13*, 33-61.

Hardle, W., & Prastyo, D. (2014). Embedded Predictor Selection for Default Risk Calculation: A Southeast Asian Industry Study. In D. Chuen, & G. Gregoriou, *Handbook of Asian Finance: Financial Markets and Sovereign Wealth Funds 1st Edition* (pp. 131-148). San Diego: Elsevier Inc.

Hardle, W., Prastyo, D., & Hafner, C. (2014). Support Vector Machines With Evolutionary Model Selection For Default Prediction. In J. Racine, L. Su, & A. Ullah, *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics* (pp. 346-373). New York: Oxford University Press.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression.* New York: John Wiley & Sons, Inc.

Hotho, A., Nurnberger, A., & Paass, G. (2005). *A Brief Survey of Text Mining.* Kassel: University of Kassel.

Johnson, R., & Wichern, D. (2007). *Applied Multivariate Statistical Analysis Sixth Edition.* New Jersey: Pearson Education, Inc.

Kamaludin. (2012). *Manajemen Keuangan-Teori & Penerapan, Edisi Revisi.* Bandung: CV. Mandar Maju.

Kamaludin, Susena, K. C., & Usman, B. (2015). *Restrukturisasi Merger & Akuisisi.* Bandung: CV. Mandar Maju.

Kasmir. (2010). *Analisis Laporan Keuangan.* Jakarta: Rajawali Press.

Kasmir. (2019). *Analisis Laporan Keuangan Edisi Revisi.* Depok: Rajawali Pers.

Kowarik, A., & Templ, M. (2016). Imputation with the R Package VIM. *Journal of Statistical Software, 74*(7).

Lestari, M. I., & Sugiharto, T. (2007). Kinerja Bank Devisa Dan Bank Non Devisa Dan Faktor-Faktor Yang Mempengaruhinya. *Proceeding PESAT (Psikologi, Ekonomi, Sastra, Arsitek & Sipil), 2*.

Mustapha, I. B., & Saeed, F. (2016, Juli 28). Bioactive Molecule Prediction Using Extreme Gradient Boosting. *molecules*.

Pasaribu, R. B. (2008). Financial Distress Prediction In Indonesia Stock Exchange (Case Study of Trade Industry Public Company). *Journal of Economics, Business, and Accounting, 11*, 153-172.

Platt, H. D., & Platt, M. B. (2002). Predicting Corporate Financial Distress: Reflections on Choice-Based Sample Bias. *Journal of Economics and Finance, 26*, 184-188.

Platt, H. D., & Platt, M. B. (2006). Understanding Differences Between Financial Distress and Bankruptcy. *Review of Apllied Economics, 2*, 141-157.

Prastyo, D. (2015). *Regularized Support Vector Machines With Application to Corporate Default Prediction.* Manuscript.

Ross, S. A., Westerfield, R. W., & Jordan, B. D. (1993). *Fundamentals of Corporate Finance.* New York: McGraw-Hill.

Sartono, A. (2001). *Manajemen Keuangan Teori dan Aplikasi, Edisi Keempat.* Yogyakarta: BPFE.

Savera, R. N. (2020). *Prediksi Financial Distress pada Perusahaan Sektor Industri di Indonesia menggunakan Metode Support Vector Machine, Generalized Extreme Value Regression, dan Logistic Regression dengan Seleksi Variabel* . Surabaya: Departemen Statistika FSAD ITS.

Scholz, M., & Klinkenberg, R. (2007). Boosting Classifiers for Drifting Concepts. *Intelligent Data Analysis 11*, 3-28.

Syahrani, I. M. (2019). *Analisis Perbandingan Teknik Ensemble Secara Boosting (XGBoost) dan Bagging (Random Forest) pada Klasifikasi Kategori Sambatan Sekuens DNA*. Bogor: IPB.

Zhang, Y., Wang, S., & Ji, G. (2013). A Rule-Based Model for Bankruptcy Prediction Based on an Improved Genetic Ant Colony Algorithm. *Mathematical Problems in Engineering*, *2013*, 10.

Zieba, M., Tomczak, S. K., & Tomczak, J. M. (2016). Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. *Expert Systems With Applications*, 93-101.

Zou, H., & Hastie, T. (2005). Regularization and Variable Selection Via The Elastic Net. *Journal of the Royal Statistical Society*(2), 301-320.

LAMPIRAN

Lampiran 1. Variabel Baru Hasil *Synthetic Features Generation*.

Variabel	Keterangan
X_1	EBITA
X_2	STA
X_3	ITR
X_4	DSIR
X_5	ROE
X_6	NPM
X_7	OPM
X_8	DER
X_9	DAR
X_{10}	CR
X_{11}	QR
X_{12}	ETD
X_{13}	WCA
X_{14}	WCS
X_{15}	rgdpg
X_{16}	bi7drr
X_{17}	USD/IDR
X_{18}	Inflasi
X_{19}	EBITA×EBITA
X_{20}	EBITA×DAR
X_{21}	EBITA×WCA
X_{22}	EBITA/WCS
X_{23}	STA×STA
X_{24}	STA/ITR
X_{25}	STA/WCS
X_{26}	ITR×DSIR
X_{27}	ITR×OPM
X_{28}	ITR/NPM
X_{29}	ITR/QR
X_{30}	DSIR/STA
X_{31}	ROE×NPM
X_{32}	ROE/QR
X_{33}	NPM/ROE

X_{34}	OPM/NPM
X_{35}	DER×WCS
X_{36}	QR/CR
X_{37}	ETD/EBITA
X_{38}	ETD/QR
X_{39}	WCA×WCA
X_{40}	WCA/NPM
X_{41}	WCS/OPM
X_{42}	bi7drr/Inflasi

Lampiran 2. Estimasi Parameter Model GEVR Secara Serentak.

***Size* 0**				
Prediktor	**Estimate**	**Std.Error**	**W**	***P-value***
(Intercept)	1,41E+01	5,22E+00	2,697	0,006987
EBITA	-3,95E+01	9,83E+00	-4,019	5,85E-05
STA	-2,53E+00	9,63E-01	-2,631	0,008525
ITR	-4,75E-02	5,18E-02	-0,916	0,359477
DSIR	-1,57E-02	8,56E-03	-1,834	0,066584
ROE	-1,92E+00	6,21E-01	-3,094	0,001973
NPM	-5,70E+00	2,43E+00	-2,348	0,01888
OPM	-2,14E+01	5,05E+00	-4,238	2,25E-05
DER	-4,83E-03	2,94E-02	-0,164	0,869478
DAR	1,83E+00	5,67E-01	3,221	0,001279
CR	-1,60E-01	2,68E-01	-0,596	0,550906
QR	1,71E-01	2,70E-01	0,632	0,527519
ETD	-5,02E+00	6,73E-01	-7,467	8,18E-14
WCA	2,71E+00	1,17E+00	2,308	0,021011
WCS	-8,55E-01	5,38E-01	-1,589	0,112164
rgdpg	2,15E-01	3,60E-01	0,596	0,550902
bi7drr	2,15E-01	1,75E-01	1,233	0,217669
USD/IDR	1,22E-04	8,87E-05	1,376	0,16881
Inflasi	-8,24E-02	1,21E-01	-0,678	0,497474

EBITA×EBITA	1,12E+02	1,75E+01	6,397	1,58E-10
EBITA×DAR	-9,49E-01	4,16E+00	-0,228	0,81976
EBITA×WCA	-3,54E+00	7,74E+00	-0,457	0,647558
EBITA/WCS	-2,86E-02	3,13E-02	-0,914	0,360957
STA×STA	1,48E-01	2,27E-01	0,651	0,515222
STA/ITR	3,50E+00	2,73E+00	1,281	0,200115
STA/WCS	3,86E-05	4,37E-04	0,088	0,929606
ITR×DSI	-3,81E-02	1,09E-02	-3,501	0,000463
ITR×OPM	-3,42E+00	3,93E-01	-8,712	<2,00E-16
ITR/NPM	-1,53E-04	4,84E-05	-3,171	0,001519
ITR/QR	1,51E-01	3,71E-02	4,062	4,87E-05
DSI/STA	4,57E-03	2,23E-03	2,045	0,040809
ROE×NPM	1,60E+01	2,50E+00	6,398	1,58E-10
ROE/QR	8,70E-01	1,02E-01	8,53	<2,00E-16
NPM/ROE	-6,95E-02	3,75E-01	-0,185	0,852868
OPM/NPM	-1,53E-02	3,23E-03	-4,73	2,25E-06
DER×WCS	-5,60E-03	4,54E-02	-0,123	0,901772
QR/CR	-2,96E-01	9,04E-01	-0,327	0,743582
ETD/EBITA	-3,74E-03	3,45E-03	-1,086	0,277642
ETD/QR	-8,18E+00	6,66E-01	-12,286	<2,00E-16
WCA×WCA	-1,02E+00	3,78E-01	-2,694	0,007063
WCA/NPM	2,04E-03	6,49E-04	3,146	0,001653
WCS/OPM	1,83E-03	1,58E-03	1,158	0,246782
bi7drr/Inflasi	-2,27E-01	6,23E-01	-0,364	0,716034

***Size* 1**				
Prediktor	**Estimate**	**Std.Error**	**W**	***P-value***
(Intercept)	1,92E+01	4,63E+00	4,135	3,54E-05
EBITA	-1,59E+01	2,34E+00	-6,827	8,67E-12
STA	-5,07E-01	1,55E-01	-3,266	0,001092

ITR	-6,93E-04	4,40E-04	-1,575	0,115235
DSIR	-3,35E-03	1,99E-03	-1,681	0,092683
ROE	-3,84E-03	1,80E-01	-0,021	0,983003
NPM	-1,02E+00	6,47E-01	-1,573	0,115781
OPM	1,07E+00	6,32E-01	1,687	0,09164
DER	1,54E-02	8,60E-03	1,793	0,07292
DAR	8,35E-01	2,35E-01	3,552	0,000383
CR	3,49E-02	5,53E-02	0,631	0,528343
QR	-8,67E-03	6,93E-02	-0,125	0,900476
ETD	1,86E-01	2,06E-01	0,903	0,366503
WCA	2,31E-01	4,12E-01	0,561	0,574951
WCS	-6,34E-02	1,16E-01	-0,549	0,582936
rgdpg	-7,56E-01	2,05E-01	-3,687	0,000227
bi7drr	7,23E-03	7,34E-02	0,098	0,921568
USD/IDR	-1,64E-04	7,61E-05	-2,152	0,031377
Inflasi	-9,05E-02	5,39E-02	-1,678	0,093323
EBITA×EBITA	-2,68E+01	2,76E+00	-9,703	<2,00E-16
EBITA×DAR	3,75E+00	1,92E+00	1,951	0,051076
EBITA×WCA	3,58E+00	3,91E+00	0,914	0,360666
EBITA/WCS	-6,96E-03	2,12E-03	-3,282	0,001032
STA×STA	2,44E-02	1,00E-02	2,435	0,014909
STA/ITR	1,81E+00	7,01E-01	2,577	0,009953
STA/WCS	-1,76E-05	1,16E-04	-0,152	0,879114
ITR×DSI	-3,06E-02	1,09E-02	-2,797	0,005161
ITR×OPM	1,12E-04	1,59E-04	0,705	0,480755
ITR/NPM	1,51E-06	4,79E-06	0,314	0,753286
ITR/QR	2,55E-02	7,82E-03	3,256	0,00113
DSI/STA	3,46E-04	4,60E-04	0,751	0,452394
ROE×NPM	-1,72E-01	1,25E-01	-1,379	0,167827
ROE/QR	2,83E-02	5,27E-02	0,536	0,591756
NPM/ROE	4,65E-02	6,04E-02	0,769	0,441822

OPM/NPM	-2,76E-04	1,34E-03	-0,207	0,836333
DER×WCS	5,80E-02	4,34E-02	1,337	0,181212
QR/CR	-3,74E-01	2,94E-01	-1,275	0,202465
ETD/EBITA	-1,30E-02	6,20E-03	-2,092	0,036421
ETD/QR	2,00E-01	2,35E-01	0,851	0,394681
WCA×WCA	-1,06E-02	3,26E-01	-0,033	0,973958
WCA/NPM	1,97E-04	4,24E-04	0,464	0,642892
WCS/OPM	-4,26E-04	9,49E-04	-0,449	0,653457
bi7drr/Inflasi	-1,00E+00	3,56E-01	-2,812	0,004921

***Size* 2**				
Prediktor	**Estimate**	**Std.Error**	**W**	***P-value***
(Intercept)	6,23E+00	3,80E+00	1,638	0,101502
EBITA	-1,12E+01	4,58E+00	-2,436	0,014857
STA	-2,37E-01	3,25E-01	-0,728	0,466667
ITR	-2,63E-04	8,81E-04	-0,299	0,765214
DSIR	6,75E-04	2,41E-03	0,28	0,779219
ROE	1,80E-01	3,96E-01	0,453	0,650676
NPM	5,45E-01	1,26E+00	0,433	0,664697
OPM	-1,91E+00	2,46E+00	-0,779	0,435828
DER	5,02E-03	1,07E-02	0,468	0,640068
DAR	1,27E+00	4,68E-01	2,707	0,006793
CR	-1,50E-01	1,35E-01	-1,113	0,265622
QR	3,65E-01	1,73E-01	2,109	0,034912
ETD	-8,36E-01	4,74E-01	-1,762	0,078075
WCA	-9,23E-01	5,82E-01	-1,584	0,113088
WCS	-3,06E-01	2,73E-01	-1,119	0,262933
rgdpg	6,26E-02	2,30E-01	0,272	0,785984
bi7drr	2,42E-01	8,06E-02	2,995	0,002742
USD/IDR	3,16E-04	8,93E-05	3,544	0,000393

Inflasi	-2,21E-01	6,19E-02	-3,57	0,000356
EBITA×EBITA	-2,98E+01	4,29E+00	-6,958	3,45E-12
EBITA×DAR	-3,17E+00	4,06E+00	-0,782	0,434109
EBITA×WCA	-2,49E+00	5,48E+00	-0,453	0,650202
EBITA/WCS	2,44E-02	3,09E-03	7,89	3,02E-15
STA×STA	2,62E-03	5,85E-02	0,045	0,964231
STA/ITR	8,15E-01	8,91E-01	0,914	0,360568
STA/WCS	-6,10E-04	4,20E-04	-1,451	0,14688
ITR×DSI	-2,10E-02	7,68E-03	-2,738	0,006177
ITR×OPM	2,80E-04	6,62E-04	0,424	0,671886
ITR/NPM	-1,83E-05	1,36E-05	-1,351	0,176777
ITR/QR	3,49E-03	1,03E-02	0,338	0,735282
DSI/STA	-1,94E-04	5,52E-04	-0,35	0,72612
ROE×NPM	1,03E+00	7,85E-01	1,311	0,18998
ROE/QR	3,26E-02	1,48E-01	0,22	0,82565
NPM/ROE	3,05E-01	1,77E-01	1,726	0,084263
OPM/NPM	1,59E-03	1,79E-03	0,885	0,376134
DER×WCS	-7,14E-05	3,30E-02	-0,002	0,998273
QR/CR	-8,24E-01	4,60E-01	-1,794	0,072833
ETD/EBITA	-2,19E-03	8,44E-03	-0,259	0,795365
ETD/QR	-1,13E-01	5,00E-01	-0,225	0,821896
WCA×WCA	-6,83E-01	4,04E-01	-1,692	0,090598
WCA/NPM	3,38E-04	6,64E-04	0,509	0,610733
WCS/OPM	2,08E-04	1,18E-03	0,177	0,859536
bi7drr/Inflasi	-1,06E+00	4,06E-01	-2,608	0,009105

***Size* 3**				
Prediktor	**Estimate**	**Std.Error**	**W**	***P-value***
(Intercept)	1,75E+01	3,85E+00	4,552	5,31E-06
EBITA	-6,21E+00	3,69E+00	-1,681	0,092823

STA	-2,87E-01	1,76E-01	-1,634	0,102268
ITR	-1,84E-03	1,19E-03	-1,537	0,124191
DSIR	-7,98E-03	2,59E-03	-3,087	0,002024
ROE	1,99E-02	1,57E-01	0,126	0,899425
NPM	1,23E+00	7,48E-01	1,639	0,101121
OPM	-8,80E-01	1,50E+00	-0,588	0,556774
DER	7,34E-03	7,95E-03	0,923	0,355799
DAR	5,28E-01	2,66E-01	1,984	0,047205
CR	-1,84E-02	3,19E-02	-0,577	0,563765
QR	3,71E-02	4,10E-02	0,905	0,365637
ETD	4,01E-01	2,41E-01	1,665	0,095971
WCA	1,74E-01	4,20E-01	0,415	0,678039
WCS	2,54E-02	1,11E-01	0,23	0,818468
rgdpg	-7,64E-01	2,40E-01	-3,179	0,001479
bi7drr	-8,88E-02	6,74E-02	-1,317	0,187692
USD/IDR	-6,30E-04	1,57E-04	-4,022	5,77E-05
Inflasi	-4,21E-02	5,07E-02	-0,83	0,406535
EBITA×EBITA	-1,96E+01	9,45E+00	-2,069	0,038519
EBITA×DAR	3,06E+00	3,35E+00	0,914	0,360779
EBITA×WCA	-1,16E+01	4,32E+00	-2,676	0,007442
EBITA/WCS	1,13E-02	3,65E-03	3,083	0,002048
STA×STA	1,26E-02	1,16E-02	1,084	0,278317
STA/ITR	2,30E+00	8,20E-01	2,798	0,005144
STA/WCS	-1,52E-04	1,40E-04	-1,086	0,277661
ITR×DSI	-1,22E-02	5,73E-03	-2,122	0,033858
ITR×OPM	6,23E-05	9,94E-04	0,063	0,950041
ITR/NPM	-1,29E-05	1,07E-05	-1,204	0,228405
ITR/QR	2,30E-02	9,14E-03	2,519	0,011759
DSI/STA	2,32E-03	6,94E-04	3,337	0,000847
ROE×NPM	1,07E+00	6,72E-01	1,586	0,112753
ROE/QR	5,90E-02	3,45E-02	1,709	0,087479

NPM/ROE	1,75E-02	5,22E-02	0,336	0,737067
OPM/NPM	1,37E-03	1,31E-03	1,042	0,297353
DER×WCS	1,43E-02	2,05E-02	0,697	0,485698
QR/CR	-7,18E-01	2,80E-01	-2,568	0,010236
ETD/EBITA	-1,05E-02	8,21E-03	-1,284	0,199155
ETD/QR	-3,30E-01	2,75E-01	-1,201	0,229756
WCA×WCA	4,52E-01	3,06E-01	1,477	0,13969
WCA/NPM	-8,84E-04	1,37E-03	-0,647	0,517896
WCS/OPM	5,30E-03	1,24E-03	4,289	1,79E-05
bi7drr/Inflasi	-1,20E+00	3,50E-01	-3,426	0,000612

Lampiran 3. Estimasi Parameter Model *Stepwise* GEVR.

***Size* 0**				
Prediktor	**Estimate**	**Std.Error**	**W**	***P-value***
(Intercept)	1.65E+01	2.83E+00	5.813	6.15E-09
EBITA	-3.80E+01	5.87E+00	-6.467	1.00E-10
STA	-1.22E+00	8.77E-02	-13.899	<2.00E-16
DSI	-4.32E-03	2.56E-03	-1.69	0.091088
ROE	-1.41E+00	1.33E-01	-10.617	<2.00E-16
NPM	-5.27E+00	2.91E+00	-1.81	0.070242
OPM	-2.07E+01	3.55E+00	-5.835	5.38E-09
DAR	1.32E+00	3.17E-01	4.165	3.12E-05
ETD	-4.40E+00	6.21E-01	-7.09	1.34E-12
WCA	1.47E+00	6.78E-01	2.173	0.029779
WCS	-8.66E-01	3.92E-01	-2.206	0.027375
EBITA×EBITA	1.09E+02	8.54E+00	12.741	<2.00E-16
ITR×DSI	-3.77E-02	7.62E-03	-4.942	7.73E-07
ITR×OPM	-5.42E-01	1.78E-01	-3.04	0.002367
ITR/NPM	-7.88E-05	1.35E-05	-5.839	5.25E-09
ITR/QR	8.65E-02	1.65E-02	5.252	1.51E-07

DSI/STA	2.03E-03	1.06E-03	1.919	0.055037
ROE×NPM	1.32E+01	2.02E+00	6.528	6.66E-11
ROE/QR	5.64E-01	6.46E-02	8.725	<2.00E-16
OPM/NPM	-1.16E-02	1.67E-03	-6.914	4.71E-12
ETD/QR	-5.40E+00	1.09E+00	-4.965	6.88E-07
WCA/NPM	1.05E-03	3.12E-04	3.352	0.000802

***Size* 1**				
Prediktor	**Estimate**	**Std.Error**	**W**	***P-value***
(Intercept)	1.85E+01	4.58E+00	4.046	5.21E-05
EBITA	-1.69E+01	1.78E+00	-9.52	<2.00E-16
STA	-5.27E-01	1.45E-01	-3.621	0.000293
ITR	-8.37E-04	3.07E-04	-2.732	0.006292
DSI	-7.25E-03	1.77E-03	-4.089	4.33E-05
OPM	1.20E+00	3.32E-01	3.612	0.000303
DER	1.72E-02	2.43E-03	7.061	1.65E-12
DAR	7.42E-01	2.07E-01	3.587	0.000335
WCA	4.52E-01	2.75E-01	1.646	0.099687
rgdpg	-7.65E-01	2.05E-01	-3.74	0.000184
USD/IDR	-1.58E-04	7.54E-05	-2.097	0.03598
Inflasi	-9.13E-02	2.72E-02	-3.358	0.000784
EBITA×EBITA	-2.04E+01	2.71E+00	-7.516	5.64E-14
EBITA×DAR	3.53E+00	1.22E+00	2.884	0.003921
EBITA/WCS	-8.56E-03	1.96E-03	-4.372	1.23E-05
STA×STA	2.74E-02	9.26E-03	2.957	0.003102
STA/ITR	2.85E+00	6.60E-01	4.324	1.53E-05
ITR×DSI	-2.88E-02	1.08E-02	-2.67	0.007582
ITR/QR	2.11E-02	7.18E-03	2.936	0.003321
DSI/STA	1.49E-03	3.70E-04	4.024	5.73E-05
DER×WCS	5.19E-02	3.01E-02	1.722	0.085157
ETD/EBITA	-1.33E-02	5.58E-03	-2.376	0.017511

bi7drr/Inflasi	-1.03E+00	3.02E-01	-3.421	0.000623

***Size* 2**				
Prediktor	**Estimate**	**Std.Error**	**W**	***P-value***
(Intercept)	7.71E+00	2.32E+00	3.322	0.000894
EBITA	-1.49E+01	9.99E-01	-14.948	<2.00E-16
STA	1.23E-01	5.41E-02	2.272	0.023065
ROE	2.40E-01	1.36E-01	1.766	0.077458
DAR	5.64E-01	2.07E-01	2.733	0.006273
QR	2.33E-01	4.49E-02	5.198	2.01E-07
ETD	-5.37E-01	2.84E-01	-1.891	0.058652
WCA	-1.30E+00	3.32E-01	-3.901	9.59E-05
USD/IDR	1.06E-04	5.52E-05	1.914	0.055651
Inflasi	-5.49E-02	1.70E-02	-3.224	0.001263
EBITA×EBITA	-3.09E+01	2.27E+00	-13.58	<2.00E-16
EBITA/WCS	2.17E-02	2.01E-03	10.819	<2.00E-16
STA/ITR	1.19E+00	3.44E-01	3.464	0.000533
ITR×DSI	-2.03E-02	6.06E-03	-3.357	0.000787
WCA×WCA	-6.85E-01	2.39E-01	-2.863	0.004195
bi7drr/Inflasi	-7.00E-01	2.04E-01	-3.431	0.000601

***Size* 3**				
Prediktor	**Estimate**	**Std.Error**	**W**	***P-value***
(Intercept)	1.75E+01	4.29E+00	4.074	4.62E-05
EBITA	-1.32E+01	1.47E+00	-8.98	<2.00E-16
STA	-6.44E-01	1.97E-01	-3.269	0.001078
ITR	-1.15E-03	4.49E-04	-2.565	0.010309
DSI	-9.97E-03	2.82E-03	-3.532	0.000413
NPM	1.68E+00	5.17E-01	3.255	0.001132
DER	1.65E-02	8.80E-03	1.876	0.060692
DAR	6.64E-01	1.74E-01	3.826	0.00013

rgdpg	-5.40E-01	2.71E-01	-1.99	0.046563
bi7drr	-1.29E-01	4.58E-02	-2.824	0.004736
USD/IDR	-5.51E-04	1.94E-04	-2.848	0.004398
EBITA×EBITA	-2.09E+01	3.65E+00	-5.712	1.11E-08
EBITA×WCA	-1.16E+01	2.90E+00	-4.007	6.16E-05
STA×STA	3.49E-02	1.92E-02	1.818	0.069043
STA/ITR	3.21E+00	9.01E-01	3.565	0.000363
ITR×DSI	-1.80E-02	5.38E-03	-3.337	0.000847
ITR/QR	2.09E-02	7.01E-03	2.982	0.002863
DSI/STA	2.28E-03	5.94E-04	3.831	0.000128
ROE×NPM	6.82E-01	3.63E-01	1.88	0.060096
ROE/QR	6.22E-02	2.16E-02	2.883	0.00394
QR/CR	-5.83E-01	3.11E-01	-1.874	0.060941
ETD/QR	-3.37E-01	1.33E-01	-2.535	0.011236
WCS/OPM	3.52E-03	9.73E-04	3.614	0.000302
bi7drr/Inflasi	-1.13E+00	2.73E-01	-4.142	3.44E-05

Lampiran 4. Estimasi Parameter Model LOGIT Secara Serentak.

***Size* 0**				
Prediktor	**Estimate**	**Std.Error**	**W**	***P-value***
(Intercept)	3.63E+15	9.42E+07	3.86E+07	<2E-16
EBITA	-9.04E+15	4.87E+07	-1.86E+08	<2E-16
STA	-2.53E+14	4.87E+06	-5.20E+07	<2E-16
ITR	3.61E+11	6.64E+03	5.45E+07	<2E-16
DSIR	-1.59E+12	4.44E+04	-3.59E+07	<2E-16
ROE	-8.22E+13	3.31E+06	-2.48E+07	<2E-16
NPM	-1.57E+15	9.19E+06	-1.71E+08	<2E-16
OPM	-4.12E+15	1.02E+07	-4.04E+08	<2E-16
DER	-5.80E+12	1.71E+05	-3.40E+07	<2E-16
DAR	2.81E+14	6.33E+06	4.44E+07	<2E-16

CR	-3.24E+13	1.06E+06	-3.05E+07	<2E-16
QR	3.98E+13	1.07E+06	3.71E+07	<2E-16
ETD	-1.60E+14	4.51E+06	-3.55E+07	<2E-16
WCA	-3.17E+13	9.72E+06	-3.26E+06	<2E-16
WCS	-5.31E+13	2.26E+06	-2.35E+07	<2E-16
rgdpg	4.45E+13	6.36E+06	6.99E+06	<2E-16
bi7drr	3.28E+13	2.80E+06	1.17E+07	<2E-16
USD/IDR	1.95E+09	1.63E+03	1.20E+06	<2E-16
Inflasi	1.23E+11	2.05E+06	6.03E+04	<2E-16
EBITA×EBITA	1.55E+16	6.33E+07	2.45E+08	<2E-16
EBITA×DAR	1.94E+15	4.53E+07	4.27E+07	<2E-16
EBITA×WCA	3.49E+15	5.77E+07	6.05E+07	<2E-16
EBITA/WCS	-3.29E+12	8.21E+04	-4.01E+07	<2E-16
STA×STA	9.43E+12	3.99E+05	2.37E+07	<2E-16
STA/ITR	5.49E+14	1.88E+07	2.93E+07	<2E-16
STA/WCS	7.00E+10	3.27E+03	2.14E+07	<2E-16
ITR×DSIR	-9.36E+12	2.00E+05	-4.67E+07	<2E-16
ITR×OPM	3.61E+11	1.49E+03	2.42E+08	<2E-16
ITR/NPM	-2.44E+10	2.39E+02	-1.02E+08	<2E-16
ITR/QR	6.28E+12	1.48E+04	4.23E+08	<2E-16
DSI/STA	5.59E+11	2.32E+03	2.40E+08	<2E-16
ROE×NPM	2.92E+15	4.24E+06	6.89E+08	<2E-16
ROE/QR	8.84E+13	2.19E+05	4.03E+08	<2E-16
NPM/ROE	-3.01E+13	4.86E+05	-6.19E+07	<2E-16
OPM/NPM	-3.35E+12	4.96E+04	-6.75E+07	<2E-16
DER×WCS	4.02E+12	3.80E+05	1.06E+07	<2E-16
QR/CR	-9.69E+13	1.42E+06	-6.83E+07	<2E-16
ETD/EBITA	-3.31E+11	1.69E+04	-1.96E+07	<2E-16
ETD/QR	-1.24E+15	2.17E+06	-5.70E+08	<2E-16
WCA×WCA	-1.61E+14	4.31E+06	-3.73E+07	<2E-16
WCA/NPM	-1.67E+11	1.22E+04	-1.36E+07	<2E-16

WCS/OPM	1.88E+11	1.49E+04	1.27E+07	<2E-16
bi7drr/Inflasi	2.63E+13	1.09E+07	2.42E+06	<2E-16

***Size* 1**				
Prediktor	**Estimate**	**Std.Error**	**W**	***P-value***
(Intercept)	2.06E+01	4.98E+00	4.129	3.64E-05
EBITA	-2.32E+01	4.26E+00	-5.459	4.78E-08
STA	-8.51E-01	2.65E-01	-3.215	0.001306
ITR	-9.03E-04	5.91E-04	-1.528	0.126489
DSIR	-2.65E-03	2.25E-03	-1.177	0.239238
ROE	5.05E-02	3.94E-01	0.128	0.898011
NPM	-4.57E-01	8.26E-01	-0.553	0.580295
OPM	8.47E-01	7.39E-01	1.147	0.25156
DER	3.97E-02	1.83E-02	2.167	0.030271
DAR	1.44E+00	3.99E-01	3.6	0.000318
CR	1.68E-02	5.75E-02	0.292	0.770605
QR	3.37E-02	8.66E-02	0.39	0.696785
ETD	-2.33E-01	3.34E-01	-0.698	0.485197
WCA	2.35E-01	6.23E-01	0.377	0.70599
WCS	-9.69E-02	1.47E-01	-0.658	0.510292
rgdpg	-1.11E+00	3.82E-01	-2.905	0.00367
bi7drr	7.81E-02	1.40E-01	0.559	0.576395
USD/IDR	-2.34E-04	1.39E-04	-1.678	0.093281
Inflasi	-2.08E-01	1.02E-01	-2.038	0.04158
EBITA×EBITA	-2.98E+01	1.06E+01	-2.799	0.005131
EBITA×DAR	2.31E+00	4.16E+00	0.555	0.578738
EBITA×WCA	4.67E-01	7.19E+00	0.065	0.94821
EBITA/WCS	-2.26E-03	1.98E-02	-0.114	0.909247
STA×STA	4.69E-02	1.94E-02	2.419	0.015559
STA/ITR	1.53E+00	1.04E+00	1.469	0.141828
STA/WCS	-7.47E-05	2.09E-04	-0.358	0.720342

ITR×DSIR	-2.41E-02	7.02E-03	-3.431	0.000602
ITR×OPM	1.90E-04	1.97E-04	0.965	0.334592
ITR/NPM	-1.27E-06	8.82E-06	-0.144	0.885286
ITR/QR	2.71E-02	1.04E-02	2.609	0.009092
DSIR/STA	1.06E-04	1.54E-04	0.687	0.491846
ROE×NPM	-9.12E-02	2.22E-01	-0.41	0.681466
ROE/QR	2.94E-02	7.95E-02	0.37	0.711623
NPM/ROE	1.03E-01	7.39E-02	1.387	0.165342
OPM/NPM	1.18E-04	1.97E-03	0.06	0.952083
DER×WCS	8.96E-02	6.00E-02	1.492	0.135759
QR/CR	-4.79E-01	5.11E-01	-0.937	0.348781
ETD/EBITA	-2.24E-02	1.03E-02	-2.177	0.029454
ETD/QR	2.95E-01	3.17E-01	0.931	0.351884
WCA×WCA	-5.06E-01	3.51E-01	-1.443	0.149079
WCA/NPM	4.45E-04	5.58E-04	0.797	0.425303
WCS/OPM	5.31E-05	1.39E-03	0.038	0.969572
bi7drr/Inflasi	-2.05E+00	6.72E-01	-3.042	0.002349

***Size* 2**				
Prediktor	**Estimate**	**Std.Error**	**W**	***P-value***
(Intercept)	1.52E+15	1.16E+08	13156134	<2E-16
EBITA	-2.72E+15	5.63E+07	-4.8E+07	<2E-16
STA	-3.39E+14	5.42E+06	-6.3E+07	<2E-16
ITR	-3.05E+11	1.07E+04	-2.9E+07	<2E-16
DSIR	-2.02E+11	5.00E+04	-4042548	<2E-16
ROE	-1.95E+14	4.15E+06	-4.7E+07	<2E-16
NPM	5.46E+14	9.52E+06	57306238	<2E-16
OPM	-2.87E+15	1.09E+07	-2.6E+08	<2E-16
DER	6.64E+12	2.27E+05	29205791	<2E-16
DAR	5.51E+14	7.01E+06	78579510	<2E-16
CR	-8.67E+13	1.17E+06	-7.4E+07	<2E-16

QR	1.63E+14	1.19E+06	1.37E+08	<2E-16
ETD	-2.39E+14	5.13E+06	-4.7E+07	<2E-16
WCA	-5.18E+14	1.13E+07	-4.6E+07	<2E-16
WCS	-1.72E+14	2.64E+06	-6.5E+07	<2E-16
rgdpg	1.43E+14	8.16E+06	17578952	<2E-16
bi7drr	1.47E+14	2.87E+06	51304434	<2E-16
USD/IDR	3.16E+11	3.06E+03	1.03E+08	<2E-16
Inflasi	-1.18E+14	2.17E+06	-5.4E+07	<2E-16
EBITA×EBITA	-6.30E+15	6.82E+07	-9.2E+07	<2E-16
EBITA×DAR	-1.04E+15	5.65E+07	-1.8E+07	<2E-16
EBITA×WCA	2.60E+15	6.92E+07	37575687	<2E-16
EBITA/WCS	5.75E+12	9.37E+04	61369239	<2E-16
STA×STA	1.96E+13	4.23E+05	46340929	<2E-16
STA/ITR	1.14E+15	2.19E+07	51852283	<2E-16
STA/WCS	-3.02E+11	4.93E+03	-6.1E+07	<2E-16
ITR×DSIR	-1.35E+13	2.02E+05	-6.7E+07	<2E-16
ITR×OPM	1.81E+11	6.94E+03	26144140	<2E-16
ITR/NPM	7.25E+09	2.50E+02	29060467	<2E-16
ITR/QR	8.68E+11	1.52E+04	57142879	<2E-16
DSIR/STA	-7.64E+10	2.49E+03	-3.1E+07	<2E-16
ROE×NPM	6.19E+14	4.32E+06	1.43E+08	<2E-16
ROE/QR	3.17E+13	2.63E+05	1.21E+08	<2E-16
NPM/ROE	1.34E+14	4.92E+05	2.73E+08	<2E-16
OPM/NPM	-2.30E+12	6.29E+04	-3.7E+07	<2E-16
DER×WCS	-1.37E+13	4.29E+05	-3.2E+07	<2E-16
QR/CR	-1.19E+14	1.16E+07	-1E+07	<2E-16
ETD/EBITA	-3.84E+11	1.89E+04	-2E+07	<2E-16
ETD/QR	-2.33E+14	2.43E+06	-9.6E+07	<2E-16
WCA×WCA	-4.28E+14	6.64E+06	-6.4E+07	<2E-16
WCA/NPM	2.05E+11	1.55E+04	13256393	<2E-16
WCS/OPM	2.93E+10	2.59E+04	1133765	<2E-16

bi7drr/Inflasi	-4.96E+14	1.41E+07	-3.5E+07	<2E-16

***Size* 3**				
Prediktor	**Estimate**	**Std.Error**	**W**	***P-value***
(Intercept)	1.76E+01	6.23E+00	2.83	0.00466
EBITA	-1.11E+01	4.71E+00	-2.354	0.01859
STA	-6.30E-01	2.87E-01	-2.192	0.02841
ITR	-1.49E-03	1.11E-03	-1.349	0.17731
DSIR	-7.94E-03	3.52E-03	-2.259	0.0239
ROE	1.03E-01	3.24E-01	0.317	0.75154
NPM	1.72E+00	9.52E-01	1.809	0.07038
OPM	-2.20E+00	2.12E+00	-1.037	0.2996
DER	1.87E-02	1.49E-02	1.259	0.20813
DAR	5.76E-01	2.85E-01	2.018	0.04362
CR	-3.51E-02	4.18E-02	-0.84	0.40098
QR	5.33E-02	5.65E-02	0.945	0.34488
ETD	2.92E-01	3.14E-01	0.93	0.35222
WCA	-2.63E-01	5.77E-01	-0.456	0.64868
WCS	2.04E-02	1.33E-01	0.154	0.87765
rgdpg	-4.87E-01	4.24E-01	-1.149	0.25073
bi7drr	-2.19E-01	1.28E-01	-1.713	0.08672
USD/IDR	-5.99E-04	2.83E-04	-2.115	0.03443
Inflasi	6.99E-02	9.35E-02	0.748	0.45462
EBITA×EBITA	-2.16E+01	8.39E+00	-2.571	0.01013
EBITA×DAR	9.74E-02	3.30E+00	0.03	0.97644
EBITA×WCA	-1.26E+01	6.82E+00	-1.846	0.06485
EBITA/WCS	1.51E-02	9.03E-03	1.672	0.09462
STA×STA	3.48E-02	2.01E-02	1.728	0.08404
STA/ITR	2.80E+00	1.22E+00	2.292	0.02191
STA/WCS	-2.58E-04	1.90E-04	-1.357	0.17466
ITR×DSIR	-1.91E-02	7.08E-03	-2.694	0.00706

ITR×OPM	4.63E-04	9.71E-04	0.477	0.63369
ITR/NPM	-9.06E-06	1.83E-05	-0.496	0.61961
ITR/QR	2.10E-02	1.17E-02	1.804	0.07131
DSIR/STA	1.88E-03	7.07E-04	2.657	0.00789
ROE×NPM	8.62E-01	7.63E-01	1.13	0.25868
ROE/QR	6.57E-02	3.55E-02	1.853	0.06395
NPM/ROE	2.92E-02	4.69E-02	0.622	0.53399
OPM/NPM	2.34E-03	2.29E-03	1.019	0.30832
DER×WCS	-1.92E-03	2.39E-02	-0.081	0.93575
QR/CR	-7.29E-01	4.92E-01	-1.483	0.13805
ETD/EBITA	-1.88E-03	3.08E-03	-0.612	0.54053
ETD/QR	-4.47E-01	3.18E-01	-1.407	0.15949
WCA×WCA	7.05E-02	3.63E-01	0.194	0.84605
WCA/NPM	-1.46E-04	2.36E-03	-0.062	0.95065
WCS/OPM	3.49E-03	1.34E-03	2.602	0.00928
bi7drr/Inflasi	-7.60E-01	6.39E-01	-1.188	0.23472

Lampiran 5. Estimasi Parameter Model *Stepwise* LOGIT.

***Size* 0**				
Prediktor	**Estimate**	**Std.Error**	**W**	***P-value***
(Intercept)	3.63E+15	9.42E+07	3.86E+07	<2E-16
EBITA	-9.04E+15	4.87E+07	-1.86E+08	<2E-16
STA	-2.53E+14	4.87E+06	-5.20E+07	<2E-16
ITR	3.61E+11	6.64E+03	5.45E+07	<2E-16
DSIR	-1.59E+12	4.44E+04	-3.59E+07	<2E-16
ROE	-8.22E+13	3.31E+06	-2.48E+07	<2E-16
NPM	-1.57E+15	9.19E+06	-1.71E+08	<2E-16
OPM	-4.12E+15	1.02E+07	-4.04E+08	<2E-16
DER	-5.80E+12	1.71E+05	-3.40E+07	<2E-16
DAR	2.81E+14	6.33E+06	4.44E+07	<2E-16

CR	-3.24E+13	1.06E+06	-3.05E+07	<2E-16
QR	3.98E+13	1.07E+06	3.71E+07	<2E-16
ETD	-1.60E+14	4.51E+06	-3.55E+07	<2E-16
WCA	-3.17E+13	9.72E+06	-3.26E+06	<2E-16
WCS	-5.31E+13	2.26E+06	-2.35E+07	<2E-16
rgdpg	4.45E+13	6.36E+06	6.99E+06	<2E-16
bi7drr	3.28E+13	2.80E+06	1.17E+07	<2E-16
USD/IDR	1.95E+09	1.63E+03	1.20E+06	<2E-16
Inflasi	1.23E+11	2.05E+06	6.03E+04	<2E-16
EBITA×EBITA	1.55E+16	6.33E+07	2.45E+08	<2E-16
EBITA×DAR	1.94E+15	4.53E+07	4.27E+07	<2E-16
EBITA×WCA	3.49E+15	5.77E+07	6.05E+07	<2E-16
EBITA/WCS	-3.29E+12	8.21E+04	-4.01E+07	<2E-16
STA×STA	9.43E+12	3.99E+05	2.37E+07	<2E-16
STA/ITR	5.49E+14	1.88E+07	2.93E+07	<2E-16
STA/WCS	7.00E+10	3.27E+03	2.14E+07	<2E-16
ITR×DSIR	-9.36E+12	2.00E+05	-4.67E+07	<2E-16
ITR×OPM	3.61E+11	1.49E+03	2.42E+08	<2E-16
ITR/NPM	-2.44E+10	2.39E+02	-1.02E+08	<2E-16
ITR/QR	6.28E+12	1.48E+04	4.23E+08	<2E-16
DSIR/STA	5.59E+11	2.32E+03	2.40E+08	<2E-16
ROE×NPM	2.92E+15	4.24E+06	6.89E+08	<2E-16
ROE/QR	8.84E+13	2.19E+05	4.03E+08	<2E-16
NPM/ROE	-3.01E+13	4.86E+05	-6.19E+07	<2E-16
OPM/NPM	-3.35E+12	4.96E+04	-6.75E+07	<2E-16
DER×WCS	4.02E+12	3.80E+05	1.06E+07	<2E-16
QR/CR	-9.69E+13	1.42E+06	-6.83E+07	<2E-16
ETD/EBITA	-3.31E+11	1.69E+04	-1.96E+07	<2E-16
ETD/QR	-1.24E+15	2.17E+06	-5.70E+08	<2E-16
WCA×WCA	-1.61E+14	4.31E+06	-3.73E+07	<2E-16
WCA/NPM	-1.67E+11	1.22E+04	-1.36E+07	<2E-16

WCS/OPM	1.88E+11	1.49E+04	1.27E+07	<2E-16
bi7drr/Inflasi	2.63E+13	1.09E+07	2.42E+06	<2E-16

***Size* 1**				
Prediktor	**Estimate**	**Std.Error**	**W**	***P-value***
(Intercept)	1.97E+01	4.80E+00	4.108	3.99E-05
EBITA	-2.10E+01	1.70E+00	-12.326	<2.00E-16
STA	-4.15E-01	1.35E-01	-3.087	0.002023
ITR	-1.10E-03	4.29E-04	-2.552	0.01071
DER	3.87E-02	1.82E-02	2.131	0.033092
DAR	1.63E+00	2.69E-01	6.074	1.25E-09
QR	3.99E-02	3.42E-02	1.166	0.24364
rgdpg	-1.06E+00	3.73E-01	-2.836	0.004563
USD/IDR	-2.26E-04	1.36E-04	-1.667	0.095568
Inflasi	-1.59E-01	4.92E-02	-3.237	0.001209
EBITA×EBITA	-2.82E+01	7.61E+00	-3.7	0.000216
ITR×DSIR	-2.48E-02	6.68E-03	-3.709	0.000208
ITR×OPM	2.14E-04	1.42E-04	1.505	0.132219
ITR/QR	2.21E-02	8.42E-03	2.625	0.008676
NPM/ROE	1.09E-01	5.84E-02	1.865	0.062171
DER×WCS	1.11E-01	4.82E-02	2.298	0.021583
ETD/EBITA	-1.56E-02	7.63E-03	-2.049	0.040475
ETD/QR	2.28E-01	1.34E-01	1.706	0.088006
WCA×WCA	-6.57E-01	2.49E-01	-2.639	0.008312
bi7drr/Inflasi	-1.82E+00	5.53E-01	-3.289	0.001007

***Size* 2**				
Prediktor	**Estimate**	**Std.Error**	**W**	***P-value***
(Intercept)	7.00E+00	2.79E+00	2.505	0.01225
EBITA	-1.59E+01	1.64E+00	-9.746	<2.00E-16
STA	-1.94E-01	1.37E-01	-1.419	0.15604

DAR	1.15E+00	2.96E-01	3.89	0.0001
CR	-1.30E-01	1.07E-01	-1.21	0.22623
QR	3.22E-01	1.54E-01	2.096	0.03608
ETD	-6.91E-01	2.91E-01	-2.376	0.01749
WCA	-9.38E-01	5.34E-01	-1.756	0.0791
WCS	-2.88E-01	1.62E-01	-1.782	0.07481
bi7drr	2.45E-01	1.26E-01	1.941	0.0523
USD/IDR	1.70E-04	9.39E-05	1.806	0.07096
Inflasi	-2.27E-01	8.62E-02	-2.633	0.00846
EBITA×EBITA	-2.86E+01	5.65E+00	-5.063	4.14E-07
EBITA/WCS	2.56E-02	5.34E-03	4.792	1.65E-06
STA/ITR	8.87E-01	5.94E-01	1.493	0.1354
STA/WCS	-6.30E-04	3.40E-04	-1.855	0.06365
ITR×DSIR	-2.13E-02	6.67E-03	-3.186	0.00144
DSIR/STA	-1.51E-04	1.12E-04	-1.352	0.17643
ROE×NPM	9.16E-01	5.08E-01	1.804	0.07117
ROE/QR	2.52E-02	2.88E-02	0.873	0.38241
NPM/ROE	2.95E-01	1.21E-01	2.436	0.01485
QR/CR	-7.84E-01	5.38E-01	-1.457	0.14499
WCA×WCA	-6.15E-01	2.97E-01	-2.07	0.03842
bi7drr/Inflasi	-1.11E+00	4.72E-01	-2.359	0.01832

***Size* 3**				
Prediktor	**Estimate**	**Std.Error**	**W**	***P-value***
(Intercept)	1.04E+01	3.07E+00	3.396	0.000683
EBITA	-1.33E+01	1.57E+00	-8.444	<2.00E-16
STA	-2.78E-01	1.59E-01	-1.751	0.07994
ITR	-1.02E-03	5.11E-04	-1.991	0.04651
DSIR	-6.38E-03	2.96E-03	-2.155	0.031149
DAR	7.27E-01	2.26E-01	3.216	0.001298
QR	2.25E-02	2.44E-02	0.925	0.354734

ETD	5.98E-01	2.65E-01	2.258	0.023935
USD/IDR	-3.06E-04	1.53E-04	-2.005	0.044985
EBITA×EBITA	-2.04E+01	7.04E+00	-2.902	0.00371
EBITA×WCA	-1.27E+01	3.49E+00	-3.631	0.000283
STA/ITR	1.98E+00	1.03E+00	1.925	0.054255
ITR×DSIR	-1.85E-02	6.87E-03	-2.684	0.007277
ITR/QR	1.71E-02	7.87E-03	2.166	0.030344
DSIR/STA	1.72E-03	6.10E-04	2.822	0.004766
QR/CR	-8.14E-01	4.41E-01	-1.847	0.064697
ETD/EBITA	-3.65E-03	7.81E-03	-0.467	0.640803
WCS/OPM	3.30E-03	1.17E-03	2.829	0.004667
bi7drr/Inflasi	-6.07E-01	2.50E-01	-2.428	0.015175

Lampiran 6. Estimasi Parameter Model *Lasso* LOGIT.

Prediktor	**Estimate**			
	***Size* 0**	***Size* 1**	***Size* 2**	***Size* 3**
(Intercept)	1.44E+01	4.46E+00	2.68E+00	7.63E+00
EBITA	-5.46E+01	-1.72E+01	-1.22E+01	-9.68E+00
STA	-5.02E-01	-7.95E-02	-1.65E-02	0
ITR	3.93E-03	1.74E-06	0	0
DSIR	0	0	0	0
ROE	0	0	0	0
NPM	-6.20E+00	0	0	0
OPM	-8.03E+00	0	0	-6.81E-02
DER	2.88E-03	0	0	5.26E-04
DAR	8.98E-01	9.90E-01	4.98E-01	4.58E-01
CR	0	2.60E-03	0	0
QR	1.22E-03	0	1.31E-03	4.25E-03
ETD	-9.79E-01	-1.41E-02	-4.39E-02	0
WCA	-1.15E+00	0	-4.04E-01	-6.21E-01

WCS	0	0	0	0
rgdpg	5.16E-02	0	0	0
bi7drr	6.00E-02	0	0	-2.61E-02
USD/IDR	8.35E-05	5.04E-05	1.49E-05	-1.59E-04
Inflasi	0	0	0	0
EBITA×EBITA	5.75E+01	0	0	-1.06E+01
EBITA×DAR	7.95E+00	0	0	0
EBITA×WCA	0	0	-3.47E+00	-7.86E+00
EBITA/WCS	-8.92E-04	-6.90E-04	-7.54E-04	-1.06E-03
STA×STA	0	0	0	0
STA/ITR	0	0	0	0
STA/WCS	-2.85E-04	-1.62E-04	-3.45E-04	-2.18E-04
ITR×DSIR	-3.72E-02	-1.37E-02	-7.70E-03	-1.38E-02
ITR×OPM	0	5.54E-05	0	0
ITR/NPM	-1.25E-04	0	0	0
ITR/QR	8.66E-04	2.32E-04	1.56E-05	5.57E-04
DSIR/STA	3.42E-05	0	0	3.26E-05
ROE×NPM	0	0	9.44E-02	1.67E-01
ROE/QR	0	0	4.40E-03	0
NPM/ROE	-3.84E-02	3.58E-02	2.64E-02	4.99E-02
OPM/NPM	-1.30E-02	0	0	3.87E-05
DER×WCS	5.01E-02	3.44E-03	0	0
QR/CR	-2.45E-01	-1.96E-01	0	-5.71E-01
ETD/EBITA	0	0	0	-1.50E-04
ETD/QR	-1.90E+00	0	0	0
WCA×WCA	-5.95E-01	0	0	0
WCA/NPM	1.43E-03	0	0	0
WCS/OPM	3.74E-04	0	0	1.40E-03
bi7drr/Inflasi	0	-6.18E-02	0	-4.45E-01

Lampiran 7. Estimasi Parameter Model *Elastic-Net* LOGIT.

Prediktor	**Estimate**			
	***Size* 0**	***Size* 1**	***Size* 2**	***Size* 3**
(Intercept)	1.26E+01	1.11E+01	3.92E+00	6.29E+00
EBITA	-4.47E+01	-1.20E+01	-7.61E+00	-8.59E+00
STA	-4.67E-01	-2.60E-01	-1.15E-01	-3.78E-04
ITR	3.15E-03	1.49E-04	6.33E-05	0
DSIR	0	-4.57E-04	1.05E-04	0
ROE	0	0	0	0
NPM	-4.93E+00	-2.14E-01	-2.17E-01	0
OPM	-6.71E+00	0	-8.29E-01	-1.51E-01
DER	2.09E-03	1.01E-02	5.59E-04	2.71E-04
DAR	9.40E-01	1.52E+00	1.13E+00	4.10E-01
CR	0	8.38E-03	0	0
QR	2.46E-03	3.42E-03	3.92E-02	3.48E-03
ETD	-8.70E-01	-7.55E-02	-2.03E-01	0
WCA	-1.11E+00	-5.65E-02	-7.44E-01	-6.34E-01
WCS	0	0	-3.93E-02	0
rgdpg	8.28E-02	-3.83E-01	1.84E-01	0
bi7drr	5.78E-02	-4.11E-02	7.96E-02	-1.61E-02
USD/IDR	9.97E-05	-1.00E-05	1.81E-04	-1.16E-04
Inflasi	0	-4.25E-02	-8.95E-02	0
EBITA×EBITA	2.38E+01	-7.17E+00	-1.54E+01	-7.29E+00
EBITA×DAR	6.99E+00	-5.12E+00	-5.81E+00	0
EBITA×WCA	0	-6.49E+00	-6.40E+00	-7.44E+00
EBITA/WCS	-1.19E-03	-8.88E-04	-7.86E-04	-1.16E-03
STA×STA	0	1.51E-02	0	0
STA/ITR	-8.86E-02	0	4.65E-01	0
STA/WCS	-1.46E-04	-1.87E-04	-3.29E-04	-1.98E-04
ITR×DSIR	-3.37E-02	-2.01E-02	-1.75E-02	-1.18E-02

ITR×OPM	0	9.97E-05	1.24E-04	0
ITR/NPM	-1.21E-04	0	-2.94E-06	0
ITR/QR	9.72E-04	1.15E-03	8.23E-04	4.65E-04
DSIR/STA	5.24E-05	1.86E-05	-1.07E-04	2.60E-05
ROE×NPM	4.83E-01	-7.91E-02	7.25E-01	1.76E-01
ROE/QR	0	0	1.85E-02	0
NPM/ROE	-3.44E-02	6.63E-02	1.25E-01	3.96E-02
OPM/NPM	-1.23E-02	-2.56E-04	8.21E-04	0
DER×WCS	4.12E-02	5.16E-02	3.55E-03	0
QR/CR	-4.04E-01	-5.69E-01	-2.63E-01	-5.34E-01
ETD/EBITA	0	-1.87E-04	-8.68E-05	-1.19E-04
ETD/QR	-1.56E+00	2.71E-03	-4.95E-02	0
WCA×WCA	-5.61E-01	-4.13E-01	-5.24E-01	0
WCA/NPM	1.33E-03	3.09E-04	3.40E-04	0
WCS/OPM	5.94E-04	0	1.32E-04	1.19E-03
bi7drr/Inflasi	0	-7.15E-01	-3.83E-01	-3.38E-01

Lampiran 8. Estimasi Parameter Model *Lasso* SVM.

Prediktor	**Estimate**			
	***Size* 0**	***Size* 1**	***Size* 2**	***Size* 3**
(Intercept)	-1.02E+01	-1.55E+01	-2.17E-01	-1.37E+01
EBITA	2.34E+01	1.54E+01	5.07E+00	6.27E+00
STA	6.08E-02	1.04E-01	7.68E-03	8.30E-02
ITR	-1.99E-05	0	-5.14E-03	1.21E-02
DSIR	-1.41E-04	0	-9.63E-04	-2.33E-04
ROE	0	0	4.59E-03	-1.44E-01
NPM	8.85E-01	0	3.87E-01	1.38E+00
OPM	4.76E+00	6.33E-01	3.67E+00	4.69E+00
DER	0	-6.21E-05	4.77E-03	1.65E-02
DAR	-7.17E-02	-6.33E-02	-2.70E-01	-3.26E-02

CR	2.02E-04	-2.25E-04	-1.86E-01	-5.44E-03
QR	2.06E-05	0	-3.18E-01	-6.64E-03
ETD	-7.18E-03	1.46E+00	-2.53E-02	5.12E-02
WCA	1.12E+00	3.79E-01	6.85E-01	7.85E-01
WCS	-1.33E-03	-2.56E-03	0	-3.74E-03
rgdpg	-1.28E-03	1.29E-01	8.37E-04	1.83E-01
bi7drr	-5.22E-03	4.33E-02	6.80E-03	5.18E-02
USD/IDR	-7.39E-07	-1.00E-06	-3.21E-05	3.15E-04
Inflasi	-1.51E-03	6.82E-04	2.19E-02	-7.62E-04
EBITA×EBITA	0	1.74E+00	3.29E-01	3.09E+01
EBITA×DAR	-2.20E-01	-6.01E-01	8.84E-01	-3.40E-01
EBITA×WCA	5.77E-01	8.69E-01	0	2.02E+00
EBITA/WCS	8.44E-04	4.91E-04	4.47E-04	7.13E-04
STA×STA	0	-3.33E-04	0	-4.54E-03
STA/ITR	-4.37E-02	-2.79E-02	-3.32E-01	-1.93E-01
STA/WCS	1.49E-04	3.23E-04	5.93E-04	4.45E-04
ITR×DSIR	2.40E-02	3.55E-02	1.03E-02	2.03E-02
ITR×OPM	3.87E-05	-4.17E-04	4.41E-05	2.36E-02
ITR/NPM	4.34E-04	-6.03E-06	0	2.31E-06
ITR/QR	0	0	-1.00E-01	4.90E-04
DSIR/STA	-1.48E-06	0	-4.09E-04	-6.85E-05
ROE×NPM	-4.99E-03	2.28E-01	-1.39E-02	9.47E-01
ROE/QR	0	0	-1.35E-03	-9.66E-03
NPM/ROE	1.66E-02	-3.30E-03	-3.13E-01	-1.22E-02
OPM/NPM	1.28E-01	1.57E-02	-1.66E-02	-7.10E-04
DER×WCS	-6.03E-03	-5.96E-03	7.35E-02	2.56E-01
QR/CR	8.16E-03	2.51E-01	1.54E-01	7.77E-01
ETD/EBITA	5.94E-05	1.25E-01	0	2.93E-03
ETD/QR	7.93E-01	1.18E-02	0	-8.94E-02
WCA×WCA	4.13E-01	0	7.77E-03	0
WCA/NPM	-1.05E-03	-9.29E-04	-5.36E-03	7.40E-05

WCS/OPM	-4.70E-05	-5.31E-04	-1.78E-02	-3.20E-04
bi7drr/Inflasi	5.81E-03	4.05E-01	2.12E-01	7.00E-01

Lampiran 9. Estimasi Parameter Model *Elastic-Net* SVM.

Prediktor	**Estimate**			
	***Size* 0**	***Size* 1**	***Size* 2**	***Size* 3**
(Intercept)	-0.26E+00	-0.56E+00	-0.29E+00	0.32E+00
EBITA	0	0	0	0
STA	0	0	0	0
ITR	-1.00E-02	-2.55E-04	1.62E-05	0
DSIR	-2.85E-03	-1.72E-04	1.51E-04	-1.02E-03
ROE	0	0	0	0
NPM	0	0	0	0
OPM	0	0	0	0
DER	6.41E-03	1.83E-04	0	8.50E-06
DAR	0	0	0	0
CR	0	6.33E-04	0	0
QR	1.33E-03	0	0	0
ETD	0	0	0	0
WCA	0	0	0	0
WCS	0	0	0	0
rgdpg	0	0	0	0
bi7drr	0	0	0	0
USD/IDR	4.38E-05	4.93E-05	4.49E-05	-2.96E-05
Inflasi	0	0	0	0
EBITA×EBITA	0	0	0	0
EBITA×DAR	0	0	0	0
EBITA×WCA	0	0	0	0
EBITA/WCS	-9.38E-04	0	0	-9.81E-04
STA×STA	-7.04E-02	-3.15E-03	0	0
STA/ITR	0	0	0	0

STA/WCS	2.05E-04	-5.90E-05	-2.81E-04	-1.30E-04
ITR×DSIR	0	0	0	0
ITR×OPM	-8.02E-01	3.06E-04	1.14E-05	2.31E-05
ITR/NPM	-5.70E-05	1.03E-06	-3.64E-07	-1.28E-06
ITR/QR	4.17E-02	1.57E-02	3.29E-03	2.86E-03
DSIR/STA	1.17E-03	5.29E-04	9.72E-05	4.15E-04
ROE×NPM	0	0	0	0
ROE/QR	-5.41E-03	0	0	0
NPM/ROE	-1.12E-02	0	0	0
OPM/NPM	-5.32E-03	-5.20E-08	0	0
DER×WCS	0	0	0	0
QR/CR	-2.53E-03	0	0	0
ETD/EBITA	-2.30E-03	-5.27E-03	-1.76E-04	-1.03E-04
ETD/QR	-8.55E-02	0	0	0
WCA×WCA	0	0	0	0
WCA/NPM	5.00E-04	9.87E-05	1.63E-04	0
WCS/OPM	1.34E-03	2.61E-04	0	7.06E-04
bi7drr/Inflasi	0	0	0	0

Lampiran 10. Nilai *Gain* Model XGB dengan *Synthetic Features Generation*.

***Size* 0**		***Size* 1**	
Feature	***Gain***	***Feature***	***Gain***
NPM	0.134517	ITR×OPM	0.143577
OPM/NPM	0.113615	EBITA×DAR	0.093106
EBITA	0.109421	ETD	0.086400
ITR×OPM	0.094852	ETD/QR	0.081323
ITR/NPM	0.090509	EBITA	0.076779
WCS/OPM	0.043747	EBITA/WCS	0.062576
WCA	0.034166	ROE	0.049151
ETD	0.033741	STA/WCS	0.044772
ETD/EBITA	0.032888	EBITA×WCA	0.043692
EBITA×EBITA	0.028456	OPM	0.030135

OPM	0.024189	EBITA×EBITA	0.027915
ROE/QR	0.022524	DER	0.024029
DER×WCS	0.019626	DER×WCS	0.021404
EBITA/WCS	0.018949	WCA.NPM	0.018442
EBITA×WCA	0.017398	ETD/EBITA	0.017283
DAR	0.016844	QR	0.016315
STA/WCS	0.015643	DAR	0.014480
STA×STA	0.015503	OPM/NPM	0.013005
DSIR/STA	0.015163	CR	0.012693
EBITA×DAR	0.01305	WCA	0.012368
WCS	0.011602	WCA×WCA	0.009763
WCA/NPM	0.011383	NPM/ROE	0.009275
ETD/QR	0.010485	QR/CR	0.007675
CR	0.008148	DSIR/STA	0.007476
STA	0.007768	ROE×NPM	0.007418
QR/CR	0.007505	STA	0.007219
QR	0.007022	ITR	0.007033
ROE×NPM	0.006953	STA/ITR	0.006720
WCA×WCA	0.006700	WCS	0.006132
DSIR	0.005279	STA×STA	0.005231
ROE	0.005273	rgdpg	0.005187
NPM/ROE	0.004431	bi7drr/Inflasi	0.005069
ITR/QR	0.002969	WCS/OPM	0.003940
DER	0.002394	DSIR	0.003672
STA/ITR	0.001896	USD/IDR	0.003616
USD/IDR	0.001699	ITR/NPM	0.003563
rgdpg	0.001437	ITR/QR	0.003454
ITR	0.001068	bi7drr	0.002863
ITR×DSIR	0.000453	NPM	0.002580
Inflasi	0.000406	ITR×DSIR	0.001474
bi7drr/Inflasi	0.000326	ROE/QR	0.001194
bi7drr	0.000000	Inflasi	0.000000

***Size* 2**		***Size* 3**	
Feature	***Gain***	***Feature***	***Gain***
EBITA	0.122963	ETD/QR	0.082258
ITR×OPM	0.103068	EBITA×DAR	0.073179
ROE	0.102191	OPM	0.071052

ETD	0.080231	EBITA×EBITA	0.064723
EBITA×DAR	0.078683	EBITA/WCS	0.055163
ETD/QR	0.062421	EBITA×WCA	0.046176
WCS	0.048873	OPM/NPM	0.043919
EBITA×WCA	0.046037	NPM	0.041066
ROE/QR	0.037289	QR	0.030859
DAR	0.029163	WCA/NPM	0.029585
WCA	0.029095	QR/CR	0.028716
STA	0.028950	ITR/NPM	0.027127
EBITA×EBITA	0.024178	WCS	0.026475
DER×WCS	0.019958	ITR×OPM	0.023864
ETD/EBITA	0.019930	ROE	0.023842
NPM	0.019473	ROE×NPM	0.023222
NPM/ROE	0.016184	STA/ITR	0.021007
WCA/NPM	0.015933	EBITA	0.020734
ITR/NPM	0.014274	STA×STA	0.019939
WCA×WCA	0.011749	STA	0.019282
CR	0.010482	WCA×WCA	0.018776
QR/CR	0.009866	ROE/QR	0.018434
DER	0.008652	DAR	0.018099
QR	0.006741	NPM/ROE	0.017891
OPM/NPM	0.006425	STA/WCS	0.017078
USD/IDR	0.006140	WCS/OPM	0.016703
OPM	0.005949	DER×WCS	0.016485
STA×STA	0.004935	ETD	0.015890
ITR/QR	0.004801	bi7drr	0.011321
rgdpg	0.004388	DER	0.010914
EBITA/WCS	0.004200	DSIR/STA	0.010666
STA/ITR	0.003875	rgdpg	0.010653
Inflasi	0.002657	CR	0.010099
STA/WCS	0.002484	ETD/EBITA	0.009037
bi7drr/Inflasi	0.002200	WCA	0.007241
WCS/OPM	0.002013	bi7drr/Inflasi	0.004551
ROE×NPM	0.001636	ITR	0.003769
DSIR	0.001064	DSI	0.003186
ITR×DSIR	0.000847	ITR/QR	0.002935
ITR	0.000000	USD/IDR	0.002933
bi7drr	0.000000	ITR×DSIR	0.000777
DSIR/STA	0.000000	Inflasi	0.000374

Lampiran 11. Hasil Akurasi Model Penelitian (Savera, 2020)

Model	***Size***	**AUC (%)**		***Accuracy* (%)**	
		Train	***Test***	***Train***	***Test***
SVM	0	94,12	92,00	94,23	91,95
	1	93,68	82,49	84,83	84,60
	2	82,63	78,55	83,37	80,00
	3	83,47	59,22	82,76	62,30
GEVR	0	97,60	97,08	93,67	92,41
	1	89,47	92,80	85,86	86,67
	2	87,21	87,82	84,14	82,07
	3	84,94	82,88	82,33	79,54
LOGIT	0	89,49	89,85	93,29	92,18
	1	74,17	82,09	84,55	86,21
	2	70,85	77,98	83,60	80,46
	3	66,92	55,45	81,38	72,64
Lasso SVM	0	88,32	88,43	91,79	90,80
	1	63,48	63,19	82,62	76,78
	2	60,23	59,98	79,85	75,63
	3	57,59	56,38	78,02	73,33
Elastic-Net SVM	0	92,55	93,24	90,66	92,18
	1	81,61	84,78	84,48	85,98
	2	80,72	76,75	79,54	74,94
	3	77,68	70,16	78,53	71,03
Backward GEVR	0	97,54	97,05	93,54	93,56
	1	88,86	92,63	85,31	86,67
	2	86,99	89,13	83,98	82,99
	3	84,15	85,00	82,41	80,92
Lasso LOGIT	0	89,39	90,61	93,52	92,64
	1	72,05	81,28	84,00	85,98
	2	67,53	71,41	82,30	80,69
	3	60,60	61,99	79,05	76,32
Elastic-Net LOGIT	0	89,69	89,28	92,64	90,61
	1	71,86	83,02	83,86	86,90
	2	69,27	75,98	82,99	81,61
	3	64,60	61,17	80,69	76,09

Stepwise LOGIT	0	89,19	90,12	93,10	92,87
	1	75,17	80,52	85,17	85,52
	2	70,81	76,15	83,37	81,84
	3	66,11	56,81	80,69	73,33

Lampiran 12. *Syntax Generalized Extreme Value Regression*

```
library(data.table)
library(caret)
library(ROCR)
library(bgeva)
#Memasukkan data
data_train <- read.csv('C:/Users/asus/Documents/Tugas
Akhir/Data_Training/Training_size3.csv')
data_test <-  read.csv('C:/Users/asus/Documents/Tugas
Akhir/Data_Testing/Testing_size3.csv')
#Membagi data
X_train <- data_train[,-19]
Y_train <- data_train[,19]
X_test <- data_test[,-19]
Y_test <- data_test[,19]
#Memisahkan data berdasarkan variabel
EBITA = X_train[,1]
STA = X_train[,2]
ITR = X_train[,3]
DSI = X_train[,4]
ROE = X_train[,5]
NPM = X_train[,6]
OPM = X_train[,7]
DER = X_train[,8]
DAR = X_train[,9]
CR = X_train[,10]
QR = X_train[,11]
ETD = X_train[,12]
WCA = X_train[,13]
WCS = X_train[,14]
rgdpg = X_train[,15]
bi7drr = X_train[,16]
USD.IDR = X_train[,17]
```

```
Inflasi = X_train[,18]
FD=factor(Y_train)
EBITA.EBITA.2= X_train[,53]
EBITA.DAR.2= X_train[,61]
EBITA.WCA.2= X_train[,65]
EBITA.WCS.3= X_train[,83]
STA.STA.2= X_train[,121]
STA.ITR.3= X_train[,139]
STA.WCS.3= X_train[,150]
ITR.DSI.2= X_train[,188]
ITR.OPM.2= X_train[,191]
ITR.NPM.3= X_train[,207]
ITR.QR.3= X_train[,212]
DSI.STA.3= X_train[,267]
ROE.NPM.2= X_train[,314]
ROE.QR.3= X_train[,336]
NPM.ROE.3= X_train[,390]
OPM.NPM.3= X_train[,448]
DER.WCS.2= X_train[,493]
QR.CR.3= X_train[,660]
ETD.EBITA.3= X_train[,698]
ETD.QR.3= X_train[,708]
WCA.WCA.2= X_train[,737]
WCA.NPM.3= X_train[,748]
WCS.OPM.3= X_train[,792]
bi7drr.Inflasi.3= X_train[,882]
#Menggabungkan data
datatrain <-
data.frame(FD,EBITA,STA,ITR,DSI,ROE,NPM,OPM,DER,DAR,C
R,QR,ETD,WCA,WCS,rgdpg,bi7drr,USD.IDR,Inflasi,
EBITA.EBITA.2,EBITA.DAR.2,EBITA.WCA.2,EBITA.WCS.3,ST
A.STA.2,STA.ITR.3,STA.WCS.3,ITR.DSI.2,
ITR.OPM.2,ITR.NPM.3,ITR.QR.3,DSI.STA.3,ROE.NPM.2,ROE.Q
R.3,NPM.ROE.3,OPM.NPM.3,DER.WCS.2,
QR.CR.3,ETD.EBITA.3,ETD.QR.3,WCA.WCA.2,WCA.NPM.3,WC
S.OPM.3,bi7drr.Inflasi.3)
#Mengubah data karakter menjadi numerik
datatrain[] <- lapply(datatrain, function(x)
as.numeric(as.character(x)))
```

```
#Mengisi missing value dengan nilai 0
datatrain[is.na(datatrain)] <- 0
#membentuk model
out =
bgeva(FD~EBITA+STA+ITR+DSI+ROE+NPM+OPM+DER+DAR
+CR+QR+ETD+WCA+WCS+rgdpg+bi7drr+USD.IDR+Inflasi+EBI
TA.EBITA.2+EBITA.DAR.2+EBITA.WCA.2+EBITA.WCS.3+STA
.STA.2+STA.ITR.3+STA.WCS.3+ITR.DSI.2+ITR.OPM.2+ITR.NP
M.3+ITR.QR.3+DSI.STA.3+ROE.NPM.2+ROE.QR.3+NPM.ROE.3
+OPM.NPM.3+DER.WCS.2+
QR.CR.3+ETD.EBITA.3+ETD.QR.3+WCA.WCA.2+WCA.NPM.3+
WCS.OPM.3+bi7drr.Inflasi.3,
          data = datatrain, pr.tol= 1e-5, aut.sp=FALSE, 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(datatrain, select=-FD), 0.1)
pred.train1 = hasil_train1$kelas
pred.train = factor(pred.train1)
pred.train
```

```
#Mengkategorikan kelas
Y_train2 <- factor(Y_train, labels = c("no", "yes"))
#membentuk confusion matriks
table(
  Res = Y_train2,
  XB = pred.train
)
#Menampilkan estimasi parameter model
options(max.print = 99999)
summary(out,s.meth="svd",sig.lev = 0.05)
#Memisahkan data testing
EBITA_t = X_test[,1]
STA_t = X_test[,2]
ITR_t= X_test[,3]
DSI_t= X_test[,4]
ROE_t= X_test[,5]
NPM_t= X_test[,6]
OPM_t= X_test[,7]
DER_t= X_test[,8]
DAR_t= X_test[,9]
CR_t= X_test[,10]
QR_t= X_test[,11]
ETD_t= X_test[,12]
WCA_t= X_test[,13]
WCS_t= X_test[,14]
Rgdpg_t= X_test[,15]
bi7drr_t= X_test[,16]
USD_IDR = X_test[,17]
Inflasi_t= X_test[,18]
EBITA_EBITA_2= X_test[,53]
EBITA_DAR_2= X_test[,61]
EBITA_WCA_2= X_test[,65]
EBITA_WCS_3= X_test[,83]
STA_STA_2= X_test[,121]
STA_ITR_3= X_test[,139]
STA_WCS_3= X_test[,150]
ITR_DSI_2= X_test[,188]
ITR_OPM_2= X_test[,191]
ITR_NPM_3= X_test[,207]
```

```
ITR_QR_3= X_test[,212]
DSI_STA_3= X_test[,267]
ROE_NPM_2= X_test[,314]
ROE_QR_3= X_test[,336]
NPM_ROE_3= X_test[,390]
OPM_NPM_3= X_test[,448]
DER_WCS_2= X_test[,493]
QR_CR_3= X_test[,660]
ETD_EBITA_3= X_test[,698]
ETD_QR_3= X_test[,708]
WCA_WCA_2= X_test[,737]
WCA_NPM_3= X_test[,748]
WCS_OPM_3= X_test[,792]
bi7drr_Inflasi_3= X_test[,882]
FDt = Y_test
#Menggabungkan data testing
datatest <-
data.frame(FDt,EBITA_t,STA_t,ITR_t,DSI_t,ROE_t,NPM_t,OPM_t,
DER_t,DAR_t,CR_t,QR_t,ETD_t,WCA_t,WCS_t,Rgdpg_t,bi7drr_t,
USD_IDR,Inflasi_t,
EBITA_EBITA_2,EBITA_DAR_2,EBITA_WCA_2,EBITA_WCS_
3,STA_STA_2,STA_ITR_3,STA_WCS_3,ITR_DSI_2,ITR_OPM_2,
ITR_NPM_3,ITR_QR_3,DSI_STA_3,
ROE_NPM_2,ROE_QR_3,NPM_ROE_3,OPM_NPM_3,DER_WCS
_2,QR_CR_3,ETD_EBITA_3,ETD_QR_3,WCA_WCA_2,WCA_NP
M_3,WCS_OPM_3,bi7drr_Inflasi_3)
#Mengubah karakter menjadi numerik
datatest[] <- lapply(datatest, function(x) as.numeric(as.character(x)))
#mengecheck missing value
sum(is.na(datatest))
hasil_test1 = peluang(parameter,subset(datatest, select=-FDt),0.1)
pred.test1 = hasil_test1$kelas
pred.test = factor(pred.test1)
pred.test
Y_test2 <- factor(Y_test, labels = c("no", "yes"))
table(
  Res = Y_test2,
  XB = pred.test
)
```

Lampiran 13. *Syntax Logistic Regression*

```
library(data.table)
library(caret)

data_train <- read.csv('C:/Users/asus/Documents/Tugas
Akhir/Data_Training/New/Data Training_Size3.csv')
data_test <-  read.csv('C:/Users/asus/Documents/Tugas
Akhir/Data_Testing/News/Data Testing_Size3.csv')
X_train <- data_train[,-19]
Y_train <- as.factor(data_train[,19])
X_test <- data_test[,-19]
Y_test <- as.factor(data_test[,19])

datatrain=data.frame(Y_train,X_train)
datatest=data.frame(Y_test,X_test)

datatrain[] <- lapply(datatrain, function(x)
as.numeric(as.character(x)))
datatest[] <- lapply(datatest, function(x) as.numeric(as.character(x)))

datatrain[is.na(datatrain)] <- 0
sum(is.na(datatrain))

model=glm(Y_train~.,data=datatrain,family=binomial(link='logit'),co
ntrol = list(epsilon=1e-6, maxit=10000))
OR=exp(coef(model))

pred.train=round(predict(model,subset(datatrain,select=-Y_train),
                    type="response"))
pred.test=round(predict(model,subset(datatest,select=-Y_test),
                   type="response"))

cm.train=table(actual=Y_train,prediction=pred.train)
cm.train
cm.test=table(actual=Y_test,prediction=pred.test)
cm.test

options(max.print = 99999)
```

```
summary(model)
ModelStep=step(model)
summary(ModelStep)
ORstep=exp(coef(ModelStep))
pred.train.step=round(predict(ModelStep,subset(datatrain,select = -
Y_train),type="response"))
pred.test.step=round(predict(ModelStep,subset(datatest,select = -
Y_test),type="response"))
cm.train.step=table(actual=Y_train,prediction=pred.train.step)
cm.train.step
cm.test.step=table(actual=Y_test,prediction=pred.test.step)
cm.test.step
```

Lampiran 14. *Syntax Lasso Logistic Regression*

```
library(glmnet)

data_train <- read.csv('C:/Users/asus/Documents/Tugas
Akhir/Data_Training/New/Data Training_Size3.csv')
X_train <- data_train[,-19]
Y_train <- as.factor(data_train[,19])

datatrain=data.frame(Y_train,X_train)
datatrain[] <- lapply(datatrain, function(x)
as.numeric(as.character(x)))

datatrain[is.na(datatrain)] <- 0

Xtrain <- data.matrix(X_train)
mode(Xtrain) <- "numeric"

data_test <-  read.csv('C:/Users/asus/Documents/Tugas
Akhir/Data_Testing/News/Data Testing_Size3.csv')
X_test <- data_test[,-19]
Y_test <- as.factor(data_test[,19])
Xtest <- data.matrix(X_test)

datatest=data.frame(Y_test,X_test)
datatest[] <- lapply(datatest, function(x) as.numeric(as.character(x)))
```

```
datatest[is.na(datatest)] <- 0

sum(is.na(datatrain))
sum(is.na(Xtest))

cv.lasso=cv.glmnet(Xtrain,Y_train,alpha=1, family= "binomial",
                   type.measure ="auc",maxit=10000)

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=Y_train,predict=lasso_pred.train>0.5)
cm.train

lasso_pred.test <-
round(predict(cv.lasso,newx=Xtest,s=cv.lasso$lambda.min,
type="response"))
cm.test=table(actual=Y_test,predict=lasso_pred.test>0.5)
cm.test
```

Lampiran 15. *Syntax Elastic-Net Logistic Regression*

```
library(glmnet)

data_train <- read.csv('C:/Users/asus/Documents/Tugas
Akhir/Data_Training/New/Data Training_Size3.csv')
X_train <- data_train[,-19]
Y_train <- as.factor(data_train[,19])

datatrain=data.frame(Y_train,X_train)
datatrain[] <- lapply(datatrain, function(x)
as.numeric(as.character(x)))

datatrain[is.na(datatrain)] <- 0
```

```
Xtrain <- data.matrix(X_train)
mode(Xtrain) <- "numeric"

data_test <-  read.csv('C:/Users/asus/Documents/Tugas
Akhir/Data_Testing/News/Data Testing_Size3.csv')
X_test <- data_test[,-19]
Y_test <- as.factor(data_test[,19])
Xtest <- data.matrix(X_test)

datatest=data.frame(Y_test,X_test)
datatest[] <- lapply(datatest, function(x) as.numeric(as.character(x)))

datatest[is.na(datatest)] <- 0
sum(is.na(datatrain))
sum(is.na(Xtest))

cv.elastic_net_0.1 <- cv.glmnet(Xtrain,Y_train,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), Y_train)
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,Y_train,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), Y_train)
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,Y_train,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), Y_train)
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,Y_train,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), Y_train)
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,Y_train,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), Y_train)
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,Y_train,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), Y_train)
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,Y_train,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), Y_train)
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,Y_train,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), Y_train)
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,Y_train,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), Y_train)
auc.cv.elastic_net_0.9 <- performance(predtr_cv.elastic_net_0.9,
```

```
                                measure = "auc")
auc.cv.elastic_net_0.9@y.values[1]

#regression coefficients Size 0
cv.elastic_net_0.7[["lambda.min"]]
coef(cv.elastic_net_0.7,s=cv.elastic_net_0.7$lambda.min)
predtest.cv.elastic_net_0.7 <- round(predict(cv.elastic_net_0.7,newx=
Xtest,s=cv.elastic_net_0.7$lambda.min, type="response"))
predtest_cv.elastic_net_0.7 <-
  prediction(as.numeric(predtest.cv.elastic_net_0.7), Y_test)
auctest.cv.elastic_net_0.7 <-
  performance(predtest_cv.elastic_net_0.7, measure = "auc")
auctest.cv.elastic_net_0.7@y.values[1]
cm.train=table(actual=Y_train,predict=predtr.cv.elastic_net_0.7>0.5)
cm.train
cm.test=table(actual=Y_test,predict=predtest.cv.elastic_net_0.7>0.5)
cm.test

#regression coefficients Size 1
cv.elastic_net_0.2[["lambda.min"]]
coef(cv.elastic_net_0.2,s=cv.elastic_net_0.2$lambda.min)
predtest.cv.elastic_net_0.2 <- round(predict(cv.elastic_net_0.2,newx=
Xtest,s=cv.elastic_net_0.2$lambda.min, type="response"))
predtest_cv.elastic_net_0.2 <-
  prediction(as.numeric(predtest.cv.elastic_net_0.2), Y_test)
auctest.cv.elastic_net_0.2 <-
  performance(predtest_cv.elastic_net_0.2, measure = "auc")
auctest.cv.elastic_net_0.2@y.values[1]
cm.train=table(actual=Y_train,predict=predtr.cv.elastic_net_0.2>0.5)
cm.train
cm.test=table(actual=Y_test,predict=predtest.cv.elastic_net_0.2>0.5)
cm.test

#regression coefficients Size 2
cv.elastic_net_0.1[["lambda.min"]]
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), Y_test)
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=Y_train,predict=predtr.cv.elastic_net_0.1>0.5)
cm.train
cm.test=table(actual=Y_test,predict=predtest.cv.elastic_net_0.1>0.5)
cm.test

#regression coefficients Size 3
cv.elastic_net_0.7[["lambda.min"]]
coef(cv.elastic_net_0.7,s=cv.elastic_net_0.7$lambda.min)
predtest.cv.elastic_net_0.7 <- round(predict(cv.elastic_net_0.7,newx=
Xtest,s=cv.elastic_net_0.7$lambda.min, type="response"))
predtest_cv.elastic_net_0.7 <-
  prediction(as.numeric(predtest.cv.elastic_net_0.7), Y_test)
auctest.cv.elastic_net_0.7 <-
  performance(predtest_cv.elastic_net_0.7, measure = "auc")
auctest.cv.elastic_net_0.7@y.values[1]
cm.train=table(actual=Y_train,predict=predtr.cv.elastic_net_0.7>0.5)
cm.train
cm.test=table(actual=Y_test,predict=predtest.cv.elastic_net_0.7>0.5)
cm.test
```

Lampiran 16. *Syntax Support Vector Machine*

```
library(e1071)
library(data.table)
library(caret)
library(kernlab)

data_train <- read.csv('C:/Users/asus/Documents/Tugas
Akhir/Data_Training/New/Data Training_Size3.csv')
X_train <- data_train[,-19]
Y_train <- as.factor(data_train[,19])

datatrain=data.frame(Y_train,X_train)
datatrain[] <- lapply(datatrain, function(x)
as.numeric(as.character(x)))
```

```
datatrain[is.na(datatrain)] <- 0

Xtrain <- data.matrix(X_train)
mode(Xtrain) <- "numeric"
data_test <-  read.csv('C:/Users/asus/Documents/Tugas
Akhir/Data_Testing/News/Data Testing_Size3.csv')
X_test <- data_test[,-19]
Y_test <- as.factor(data_test[,19])
Xtest <- data.matrix(X_test)

datatest=data.frame(Y_test,X_test)
datatest[] <- lapply(datatest, function(x) as.numeric(as.character(x)))
datatest[is.na(datatest)] <- 0

svmfit=svm(Y_train ~ ., data = datatrain)
print(svmfit)
predtr=round(predict(svmfit,datatrain[,-1]))
predtest=round(predict(svmfit,datatest[,-1]))

cm.train=table(actual=Y_train,prediction=predtr)
cm.train
cm.test=table(actual=Y_test,prediction=predtest)
cm.test
```

Lampiran 17. *Syntax Lasso Support Vector Machine*

```
library(penalizedSVM)
library(e1071)
library(sparseSVM)
library(data.table)
library(caret)

data_train <- read.csv('C:/Users/asus/Documents/Tugas
Akhir/Data_Training/New/Data Training_Size3.csv')
X_train <- data_train[,-19]
Y_train <- as.factor(data_train[,19])

datatrain=data.frame(Y_train,X_train)
```

```
datatrain[] <- lapply(datatrain, function(x)
as.numeric(as.character(x)))

datatrain[is.na(datatrain)] <- 0
Xtrain <- data.matrix(X_train)
mode(Xtrain) <- "numeric"

data_test <-  read.csv('C:/Users/asus/Documents/Tugas
Akhir/Data_Testing/News/Data Testing_Size3.csv')
X_test <- data_test[,-19]
Y_test <- as.factor(data_test[,19])
Xtest <- data.matrix(X_test)

datatest=data.frame(Y_test,X_test)
datatest[] <- lapply(datatest, function(x) as.numeric(as.character(x)))
datatest[is.na(datatest)] <- 0

sum(is.na(datatrain))
sum(is.na(Xtest))
Lasso=sparseSVM(Xtrain, Y_train, 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)
Lasso

#SIZE 0
pred.train=as.numeric(predict(Lasso,Xtrain,lambda=
                        0.003071531,type="class"))
(misclasstrain <- table(predict =pred.train , actual = Y_train))

pred.test=as.numeric(predict(Lasso,Xtest,lambda=
                        0.003071531,type="class"))
(misclasstest <- table(predict =pred.test , actual = Y_test))

#SIZE 1
pred.train=as.numeric(predict(Lasso,Xtrain,lambda=
                        0.002598451,type="class"))
```

```
(misclasstrain <- table(predict =pred.train , actual = Y_train))
pred.test=as.numeric(predict(Lasso,Xtest,lambda=
                        0.002598451,type="class"))
(misclasstest <- table(predict =pred.test , actual = Y_test))

#SIZE 2
pred.train=as.numeric(predict(Lasso,Xtrain,lambda=
                         0.002458197,type="class"))
(misclasstrain <- table(predict =pred.train , actual = Y_train))

pred.test=as.numeric(predict(Lasso,Xtest,lambda=
                        0.002458197,type="class"))
(misclasstest <- table(predict =pred.test , actual = Y_test))

#SIZE 3
pred.train=as.numeric(predict(Lasso,Xtrain,lambda=
                         0.002295990,type="class"))
(misclasstrain <- table(predict =pred.train , actual = Y_train))
pred.test=as.numeric(predict(Lasso,Xtest,lambda=
                        0.002295990,type="class"))
(misclasstest <- table(predict =pred.test , actual = Y_test))
```

Lampiran 18. *Syntax Elastic-Net Support Vector Machine*

```
library(sparseSVM)
library(penalizedSVM)
library(e1071)

data_train <- read.csv('C:/Users/asus/Documents/Tugas
Akhir/Data_Training/New/Data Training_Size1.csv')
X_train <- data_train[,-19]
Y_train <- as.numeric(data_train[,19])

datatrain=data.frame(Y_train,X_train)
datatrain[] <- lapply(datatrain, function(x)
as.numeric(as.character(x)))

datatrain[is.na(datatrain)] <- 0
```

```
Xtrain <- data.matrix(X_train)
mode(Xtrain) <- "numeric"

data_test <-  read.csv('C:/Users/asus/Documents/Tugas
Akhir/Data_Testing/News/Data Testing_Size1.csv')
X_test <- data_test[,-19]
Y_test <- as.numeric(data_test[,19])
Xtest <- data.matrix(X_test)

datatest=data.frame(Y_test,X_test)
datatest[] <- lapply(datatest, function(x) as.numeric(as.character(x)))
datatest[is.na(datatest)] <- 0

Y_train[Y_train==0] = -1
Y_test[Y_test==0] = -1

lambda1.scad <- c (seq(0.01 ,0.05, .01),  seq(0.1,0.5, 0.2), 1 )

fit.elasticnet=svmfs(x=Xtrain,y=Y_train,
                fs.method="DrHSVM",
                grid.search = "discrete", lambda1.set=lambda1.scad,
                parms.coding = "none", show="none",
                maxIter = 1000, maxevals=1000, inner.val.method =
"cv", cross.inner= 5,
                verbose=FALSE)

#coefficient
fit.elasticnet$model$b
fit.elasticnet$model$w

pred.train=predict(fit.elasticnet,newdata=Xtrain,
newdata.labels=Y_train)
pred.train$tab
pred.test=predict(fit.elasticnet, newdata=Xtest,
              newdata.labels=Y_test)
pred.test$tab
```

Lampiran 19. *Syntax Extreme Gradient Boosting*

```
library(xgboost)
library(knitr)
library(data.table)
library(tidyverse)
library(caret)
data_train <- read.csv('C:/Users/asus/Documents/Tugas
Akhir/Data_Training/New/Data Training_Size3.csv')
data_test <-  read.csv('C:/Users/asus/Documents/Tugas
Akhir/Data_Testing/News/Data Testing_Size3.csv')

data_train[] <- lapply(data_train, function(x)
as.numeric(as.character(x)))
data_train[is.na(data_train)] <- 0

data_test[] <- lapply(data_test, function(x)
as.numeric(as.character(x)))
data_test[is.na(data_test)] <- 0


X_train <- data_train[,-19]
Y_train <- data_train[,19]
X_test <- data_test[,-19]
Y_test <- data_test[,19]
output_vector = data_train[,19]==1

Xtrain <- data.matrix(X_train)
mode(Xtrain) <- "numeric"
Xtest <- data.matrix(X_test)
mode(Xtest) <- "numeric"

Y_train2 <- factor(Y_train, labels = c("yes", "no"))
Y_test2 <- factor(Y_test, labels = c("no", "yes"))

#Mencari parameter terbaik
model.xgb <- train(
  X_train %>% as.matrix(),
  Y_train2 %>% as.matrix(),
  tuneGrid = expand_grid(
```

```
    nrounds = c(50, 100),
    max_depth = c(1,2,3,4,5),
    eta = c(0.01, 0.05, 0.1),
    gamma = c(0,0.01,0.05,0.1),
    min_child_weight = c(1,3,5),
    colsample_bytree = c(0.3,0.6,0.8, 1),
    subsample = c(0.8,0.9,1)
  ),
  trControl = trainControl(
    method = "cv",5,
    verboseIter = TRUE
  ),
  method = "xgbTree",
  verbose = TRUE
)

table(
  Res = Y_train2,
  XB = predict(model.xgb, X_train)
)

table(
  Res = Y_test2,
  XB = predict(model.xgb, X_test)
)

sparse.matrix.train= sparse.model.matrix(FD~.-1, data = data_train)
sparse.matrix.test=  sparse.model.matrix(FD~.-1, data = data_test)

xgb.model.best=   xgboost(data= Xtrain,
                          label= output_vector,
                          eval.metric= 'logloss',
                          objective= "reg:logistic",
                          max.depth= 3,
                          eta= 0.1,
                          nthread = 5,
                          subsample= 1,
                          colsample_bytree= 0.6,
                          lambda= 1,
```

```
                    gamma= 0.05,
                    min_child_weight= 1,
                    nround= 500
)
mean(xgb.model.best$evaluation_log$train_logloss)


Y_train2 <- factor(Y_train, labels = c("no", "yes"))
pred.train=predict(xgb.model.best, newdata=Xtrain,
              newdata.labels=Y_train2)

cm.train=table(actual=Y_train2,predict=pred.train>0.5)
cm.train

Y_test2 <- factor(Y_test, labels = c("no", "yes"))

pred.test=predict(xgb.model.best, newdata=Xtest,
               newdata.labels=Y_test2)

cm.test=table(actual=Y_test2,predict=pred.test>0.5)
cm.test

options(max.print = 99999)
importance = xgb.importance(feature_names =
sparse.matrix.train@Dimnames[[2]],
                     model = xgb.model.best)  #Grab all important
features
importance
```

Lampiran 20. Surat Keterangan Data

SURAT PERNYATAAN

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

Nama : Muhammad Adlansyah Muda

NRP : 06211740000074

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

1. Sumber : Tugas Akhir Rizki Nanda Savera (Statistika ITS, 2020)
   Keterangan : Prediksi *Financial Distress* pada Perusahaan Sektor Industri di Indonesia menggunakan Metode *Support Vector Machine*, *Generalized Extreme Value Regression*, dan *Logistic Regression* dengan Seleksi Variabel

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

Mengetahui
Pembimbing Tugas Akhir

Surabaya, 06 Februari 2021

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

(Muhammad Adlansyah Muda)
NRP. 06211740000074

*(coret yang tidak perlu)

BIODATA PENULIS

Penulis pada penelitian ini ialah Muhammad Adlansyah Muda atau akrab dipanggil Adlan merupakan mahasiswa Statistika angkatan 2017. Pendidikan yang telah ditempuh oleh penulis adalah SDN 013 Tampan Pekanbaru, SMPN 8 Pekanbaru, SMAN 4 Pekanbaru, dan kemudian menempuh pendidikan di departemen Statistika Institut Teknologi Sepuluh Nopember melalui jalur SBMPTN 2017. Penulis yang memiliki hobi traveling dan bermain game ini selama menjadi mahasiswa cukup aktif dalam kepanitiaan, organisasi, dan mengikuti perlombaan. Kepanitiaan pertama penulis ialah menjadi sie media dan informasi pada kegiatan Pekan Raya Statistika (PRS) 2018, kemudian menjadi fasilitator subdivisi konsumsi pada kegiatan Generasi Integralistik (GERIGI) ITS 2018, dan terakhir menjadi sie sponsorship pada kegiatan Pekan Raya Statistika (PRS) 2019. Organisasi pertama penulis ialah menjadi staff departemen sosial Badan Eksekutif Mahasiswa (BEM) Fakultas Matematika, Komputasi, dan Sains Data (FMKSD) 2018/2019 dan yang terakhir menjadi wakil ketua divisi pers Himpunan Mahasiswa Statistika ITS (HIMASTA-ITS) 2019/2020. Pada bidang akademik, penulis dengan rekan tim berhasil menjadi finalis dalam lomba karya tulis ilmiah nasional yang diselenggarakan oleh Labma Scientific Fair (LSF) UII 2019 dan kemudian penulis bersama rekan tim juga pernah menjadi semifinalis dalam kompetisi essay statistika dalam acara Statistics Festival UGM 2020. Jika ingin memberikan saran, kritik, dan diskusi lebih lanjut dapat menghubungi melalui email: adlansyahmuda@gmail.com.