

Klasifikasi *Dynamic Financial Distress* Perusahaan Manufaktur yang Terdaftar di Bursa Efek Indonesia Tahun 2012-2014 Menggunakan Regresi Logistik Biner dan *Classification Analysis & Regression Tree (CART)*

Muniroh dan Agus Suharsono
Jurusan Statistika, Fakultas MIPA, Institut Teknologi Sepuluh Nopember (ITS)
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia
e-mail: agus_s@statistika.its.ac.id dan muniro.soeka@gmail.com

Abstrak—*Financial distress* adalah suatu konsep luas yang terdiri dari beberapa situasi dimana suatu perusahaan menghadapi masalah kesulitan keuangan. Apabila kondisi *financial distress* ini diketahui, diharapkan dapat dilakukan tindakan untuk memperbaiki situasi tersebut sehingga perusahaan tidak akan masuk pada tahap kesulitan yang lebih berat seperti kebangkrutan ataupun likuidasi. Untuk itu dilakukan klasifikasi *dynamic financial distress* dengan menggunakan metode regresi logistik biner dan CART. Tujuan dilakukan penelitian ini yaitu untuk mengetahui faktor-faktor yang diduga berpengaruh signifikan terhadap *financial distress*, serta untuk mengetahui model klasifikasi *financial distress*. Berdasarkan hasil analisis regresi logistik biner, terdapat dua variabel prediktor yang berpengaruh signifikan terhadap kondisi *financial distress* perusahaan manufaktur yaitu variabel rasio likuiditas (X_1) dan rasio aktivitas (X_5), dengan ketepatan klasifikasi sebesar 84,8%. Metode CART menghasilkan pohon klasifikasi maksimum sama dengan pohon klasifikasi optimum, dengan simpul utamanya adalah variabel rasio solvabilitas (X_4). Dengan nilai ketepatan klasifikasi sebesar 96,3%. Perbandingan kedua metode menghasilkan bahwa metode CART memiliki nilai ketepatan klasifikasi yang lebih besar 11,5% dari metode regresi logistik biner.

Kata Kunci— CART, *Financial Distress*, Regresi Logistik Biner.

I. PENDAHULUAN

Krisis keuangan global tak hanya berdampak pada sektor riil, tapi juga sangat memukul sektor finansial. Sektor yang terkena imbasan krisis keuangan global adalah seluruh sektor bidang kehidupan. Namun yang paling tampak gejalanya adalah sektor bidang industri manufaktur. Keadaan industri manufaktur yang saat ini mengalami penurunan kinerja ekspor mengalami defisit neraca perdagangan pada tahun 2011-2013[1]. Beberapa perusahaan manufaktur yang mengalami masalah keuangan mencoba mengatasi masalah tersebut dengan melakukan pinjaman dan penggabungan usaha, atau sebaliknya ada yang menutup usahanya[2].

Financial distress adalah suatu konsep luas yang terdiri dari beberapa situasi dimana suatu perusahaan menghadapi masalah kesulitan keuangan. Istilah umum untuk menggambarkan situasi tersebut adalah kebangkrutan, kegagalan, ketidak mampuan melunasi

hutang, dan *default*[2]. Penelitian sebelumnya pada prediksi *financial distress* kebanyakan hanya berfokus pada model statis untuk memprediksi model, dimana model peramalannya dibentuk dari sampel data pada waktu tertentu saja. Seiring berjalannya waktu, model statis tidak dapat meramalkan kondisi *financial distress* dalam perubahan lingkungan ekonomi maupun perubahan operasional perusahaan dengan baik dan efektif [3].

Regresi logistik merupakan metode klasifikasi dengan pendekatan parametrik, yang memiliki kelebihan yaitu nilai *odds ratio* yang menunjukkan seberapa besar pengaruh variabel prediktor suatu kategori preferensi pada suatu variabel respon[4]. *Classification and Regression Trees (CART)* adalah salah satu metode atau algoritma dari salah satu teknik eksplorasi data yaitu teknik pohon keputusan. CART merupakan metodologi statistik dengan pendekatan nonparametrik yang dikembangkan untuk analisis klasifikasi yang mampu mengatasi keterbatasan asumsi[5]. Oleh karena itu, digunakan pendekatan parametrik dengan metode regresi logistik biner dan pendekatan nonparametrik dengan menggunakan metode CART. Tujuan dari penelitian ini yaitu dapat mengetahui faktor-faktor yang diduga berpengaruh signifikan terhadap *financial distress*, serta dapat mengetahui model klasifikasi *financial distress*, apakah termasuk dalam kondisi *financial distress* atau kondisi *non-financial distress*.

II. TINJAUAN PUSTAKA

A. SMOTE

Untuk mengatasi masalah *imbalanced data* dapat menggunakan metode *synthetic minority oversampling technique (SMOTE)* yaitu salah satu metode *oversampling* yang ditemukan oleh Chawla (2003) berupa teknik penambahan jumlah sampel pada kelas minor dengan melakukan replikasi data pada kelas minor secara acak sehingga menghasilkan jumlah data yang sama dengan data pada kelas data mayor. Data yang direplikasi merupakan data yang berasal dari kelas minor. Metode yang digunakan pada algoritma SMOTE adalah *k-nearest neighbors* (ketetanggaan data yang terdekat) yang termasuk dalam kelompok metode

statistik nonparametrik. Metode ini bekerja dengan mengelompokkan data terdekat yang dipilih berdasarkan jarak Euclidean antara kedua data. Penentuan jumlah replikasi yang dilakukan disesuaikan dengan jumlah anggota pada kelas mayor. Jumlah replikasi harus sesuai dengan jumlah k pada *nearest neighbour*, jika jumlah replikasi sebanyak n maka jumlah k sebanyak $n-1$.

Misalkan terdapat dua struktur data dengan p dimensi yaitu $x^T = [x_1, x_2, \dots, x_p]$ dan $y^T = [y_1, y_2, \dots, y_p]$ maka jarak Euclidean $d(x,y)$ yang dihasilkan antara kedua data ditunjukkan pada persamaan sebagai berikut.

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2} \quad (1)$$

“Synthetic” atau replikasi data dilakukan dengan menggunakan persamaan sebagai berikut.

$$x_{syn} = x_i + (x_{knn} - x_i) \times \tau \quad (2)$$

dengan

x_{syn} = data hasil replikasi.

x_i = data yang akan direplikasi

x_{knn} = data yang memiliki jarak terdekat dari data yang akan direplikasi.

τ = bilangan random 0 sampai 1

B. Regresi Logistik Biner

Bentuk umum model regresi logistiknya adalah sebagai berikut.

$$\pi(x) = \frac{e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}}{1 + e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}} \quad (3)$$

1) Estimasi Parameter

Fungsi *Maximum Likelihood Estimation* (MLE) yang diperoleh adalah:

$$l(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (4)$$

Keterangan:

i : 1, 2, ..., p

y_i : pengamatan pada variabel respon ke- i

π_i : peluang untuk variabel respon ke- i

2) Pengujian Parameter

Hipotesis yang digunakan untuk pengujian parameter serentak adalah sebagai berikut.

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

$$H_1 : \text{minimal terdapat satu } \beta_i \neq 0; \quad i = 1, 2, \dots, p$$

Statistik uji yang digunakan adalah uji G sebagai berikut

$$G = -2 \ln \frac{\binom{n_1}{n} \binom{n_0}{n}}{\sum_{i=1}^n \hat{\pi}_i^{y_i} (1 - \hat{\pi}_i)^{(1-y_i)}} \quad (5)$$

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

Pengujian parameter secara parsial menggunakan statistik uji Wald. Hipotesis yang akan diuji adalah:

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0; \quad i = 1, 2, \dots, p$$

Statistik uji yang dipakai adalah statistik W yaitu:

$$W_i = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \quad (6)$$

Statistik uji W tersebut, yang juga disebut sebagai Statistik uji Wald, mengikuti distribusi normal sehingga H_0 ditolak jika $|W| > Z_{\alpha/2}$ [6].

C. Classification Analysis and Regression Tree (CART)

CART merupakan metodologi statistik nonparametrik yang dikembangkan untuk topik analisis klasifikasi, baik untuk variabel respon kategorik maupun kontinu. Langkah-langkah penerapan Algoritma CART adalah sebagai berikut.

1) Pembentukan pohon klasifikasi

Proses pembentukan pohon klasifikasi terdiri atas 3 tahapan, yaitu

a. Pemilihan Pemilah (Classifier)

Pemilihan pemilah tergantung pada jenis pohon atau tergantung pada jenis variabel responnya. Mengukur tingkat keheterogenan suatu kelas dari suatu simpul tertentu dalam pohon klasifikasi disebut *impurity measure* $i(t)$. Ukuran ini akan membantu menemukan fungsi pemilah yang optimal. Fungsi keheterogenan $i(t)$ adalah sebagai berikut.

i. Indeks Gini: $i(t) = \sum_{i \neq j} P(i|t)P(j|t)$

ii. Indeks Informasi: $i(t) = -\sum_j P(j|t) \log[P(j|t)]$

iii. Indeks Twoing: $i(t) = \frac{P_L P_R}{4} [\sum |P(j|t_L) - P(j|t_R)|]^2$

iv. Indeks Entropi: $i(t) = N_j(t) \log[P(j|t)]$

$P(j|t)$ adalah peluang j pada simpul t . *Goodness of Split* $\phi(s,t)$ didefinisikan sebagai penurunan keheterogenan.

$$\phi(s,t) = \Delta i(s,t) = i(t) - P_L i(t_L) - P_R i(t_R) \quad (7)$$

Pemilah yang menghasilkan nilai $\Delta i(s,t)$, lebih tinggi merupakan pemilah yang lebih baik t_L dan t_R , merupakan partisi dari simpul t menjadi dua himpunan bagian saling lepas dimana P_L dan P_R adalah proporsi masing-masing peluang simpul. Karena $t_L \cup t_R = t$ maka nilai $\Delta i(s,t)$ merepresentasikan perubahan dari keheterogenan dalam simpul t yang semata-mata disebabkan oleh pemilah s . Jika simpul yang diperoleh merupakan kelas yang tidak homogen, prosedur yang sama diulangi sampai pohon klasifikasi menjadi suatu konfigurasi tertentu, dan memenuhi.

$$\Delta i(s^*, t_i) = \max_{s \in S} \Delta i(s, t_i) \quad (8)$$

b. Penentuan Simpul Terminal

Suatu simpul t akan menjadi simpul terminal atau tidak, akan dipilah kembali bila pada simpul t tidak terdapat penurunan keheterogenan dengan adanya batasan minimum n seperti hanya terdapat satu pengamatan pada tiap simpul anak.

c. Penandaan Label Kelas

Penandaan label kelas pada simpul terminal dilakukan berdasarkan aturan jumlah terbanyak, yaitu

$$P(j_0 | t) = \max_j P(j | t) = \max_j \frac{N_j(t)}{N(t)} \quad (9)$$

Dengan $P(j_0 | t)$ adalah proporsi kelas j pada simpul t , $N_j(t)$ adalah jumlah pengamatan kelas j pada simpul t dan $N(t)$ adalah jumlah pengamatan pada simpul t . Label kelas simpul terminal t adalah j_0 yang memberi nilai dugaan kesalahan pengklasifikasian simpul t terbesar.

Proses pembentukan pohon klasifikasi berhenti saat terdapat hanya satu pengamatan dalam tiap simpul anak atau adanya batasan minimum n , semua pengamatan dalam tiap simpul anak identik, dan adanya batasan jumlah level/kedalaman pohon maksimal.

2) *Pemangkasan pohon klasifikasi*

Pemangkasan dilakukan dengan jalan memangkas bagian pohon yang kurang penting sehingga didapatkan pohon optimal. Ukuran pemangkasan yang digunakan untuk memperoleh ukuran pohon yang layak adalah Cost complexity minimum. Sub pohon dari pohon terbesar T_{max} ($T < T_{max}$) ukuran *cost complexity* yaitu

$$R_{\alpha}(t) = R(T) + \alpha |T| \tag{10}$$

dimana

$R(T)$ = Resubstitution Estimate (Proporsi kesalahan pada sub pohon)

α = kompleksitas parameter (complexity parameter)

\bar{T} = ukuran banyaknya simpul terminal pohon T

$R_{\alpha}(t)$ merupakan kombinasi linear biaya dan kompleksitas pohon yang dibentuk dengan menambahkan cost penalty bagi kompleksitas terhadap biaya kesalahan klasifikasi pohon. *Cost complexity pruning* menentukan suatu pohon bagian $T(\alpha)$ yang meminimumkan pada seluruh pohon bagian. Atau untuk setiap nilai α , dicari pohon bagian $T(\alpha) < T_{max}$ yang meminimumkan $R_{\alpha}(t)$ yaitu

$$R_{\alpha}(T(\alpha)) = \min_{T < T_{max}} R_{\alpha}(T) \tag{11}$$

Jika $R(T)$ digunakan sebagai kriteria penentuan pohon optimal maka akan cenderung pohon terbesar adalah T_1 , sebab semakin besar pohon, maka semakin kecil nilai $R(T)$ nya.

3) *Penentuan Pohon Klasifikasi Optimal*

Pohon klasifikasi yang berukuran besar memberikan nilai penduga pengganti paling kecil, sehingga pohon ini cenderung dipilih untuk menduga nilai respon. Ukuran pohon yang besar akan menyebabkan nilai kompleksitas yang tinggi karena struktur data yang digambarkan cenderung kompleks, sehingga perlu dipilih pohon optimal yang berukuran sederhana tetapi memberikan nilai penduga pengganti cukup kecil.

III. METODOLOGI PENELITIAN

A. *Sumber Data*

Data yang digunakan dalam penelitian ini adalah data sekunder yang berasal dari laporan tahunan (annual report) perusahaan tahun 2012-2014. Populasi dalam penelitian ini adalah perusahaan go public yang tercatat di Bursa Efek Indonesia (BEI) dan sampel penelitian adalah perusahaan manufaktur go public yang tercatat di BEI. Data diperoleh dari BEI, Saham OK, dan ICMD (Indonesia Capital Market Directory).

Variabel yang digunakan pada penelitian ini disajikan pada Tabel 1.

TABEL 1. VARIABEL PENELITIAN

Variabel	Nama Variabel	Pengukuran	Skala
x_1	Rasio Likuiditas	Working Capital dibagi Total Assets	Rasio
x_2	Rasio Profitabilitas	Retained Earning dibagi total Assets	Rasio
x_3	Rasio Rentabilitas	Earning Before Interest and Taxes dibagi Total Assets	Rasio
x_4	Rasio Solvabilitas	Market Value Equity dibagi Value Of Total Debt	Rasio
x_5	Rasio Aktivitas	Sales dibagi Total Assets	Rasio
x_6	ROI	(Total penjualan-Investasi) dibagi Investasi	Rasio
x_7	ROE	Profit Before Income Tax dibagi Total Equity	Rasio
Y	<i>Financial distress</i>	1 jika <i>financial distress</i> , 0 jika non <i>financial distress</i>	Nominal

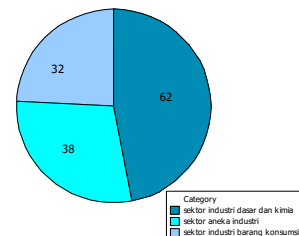
B. *Tahapan Analisis Data*

Berikut ini adalah tahapan yang dilakukan dalam analisis data pada penelitian kali ini:

- Melakukan pengumpulan data sekunder, yaitu data laporan keuangan perusahaan manufaktur *go public* yang terdaftar di BEI tahun 2012-2014.
- Melakukan klasifikasi awal terhadap data untuk kriteria *financial distress* dan *non financial distress*.
- Membagi data menjadi data *training* dan data *testing*.
- Melakukan analisis statistika deskriptif untuk mengetahui karakteristik data pada masing-masing kelas, *financial distress* dan *non financial distress*.
- Melakukan analisis SMOTE untuk *inbalance data*. Tahapannya adalah sebagai berikut:
 - Menentukan data minor dan mayor.
 - Mereplikasi setiap data minor dengan mencari nilai *knearest neighbour*.
 - Menghitung sintetik data.
- Melakukan analisis regresi logistik biner. Tahapannya adalah sebagai berikut:
 - Melakukan uji asumsi independensi.
 - Estimasi Parameter
 - Melakukan pengujian parameter secara serentak menggunakan statistik uji-G.
 - Melakukan pengujian parameter secara parsial dengan statistik uji-Wald.
 - Memodelkan kondisi keuangan perusahaan berdasarkan variabel prediktor yang berpengaruh.
 - Melakukan uji ketepatan klasifikasi.
 - Interpretasi hasil
- Melakukan analisis CART. Tahapannya adalah sebagai berikut.
 - Membentuk pohon klasifikasi dengan menentukan pemilihan pemilah, penentuan simpul terminal, dan penandaan label kelas.
 - Melakukan pemangkasan pohon klasifikasi.
 - Menentukan pohon klasifikasi optimal.
- Membandingkan hasil dari analisis regresi logistik biner dan metode CART dengan melihat peluang kesalahan klasifikasi.
- Melakukan kesimpulan dan saran.

IV. HASIL DAN PEMBAHASAN

A. *Statistika Deskriptif*



Gambar 1. Pie Chart Per Sektor Manufaktur

Berdasarkan Gambar 1 diatas dapat diketahui bahwa sektor industri manufaktur terdiri dari 3 sektor yaitu industri dasar dan kimia sebanyak 62 perusahaan, sektor aneka industri sebanyak 38 perusahaan dan sektor industri barang konsumsi sebanyak 32 perusahaan.

B. *Preprocessing Imbalance Data*

Metode yang digunakan untuk *preprocessing imbalanced* data adalah *Synthetic Minority Oversampling Technique* (SMOTE) yang merupakan teknik pengambilan sampel untuk meningkatkan jumlah data

pada kelas minoritas dengan cara mereplikasi jumlah data pada kelas minoritas secara acak sehingga jumlahnya sama dengan data pada kelas mayoritas.

TABEL 2. DISTRIBUSI DATA SIMULASI SEBELUM DAN SETELAH SMOTE TAHUN 2013-2014

Kelas mayor	Kelas minor	Replikasi	Kelas mayor	Kelas minor baru
109(83%)	23 (17%)	4	109(49%)	115(51%)

TABEL 3. DISTRIBUSI DATA SIMULASI SEBELUM DAN SETELAH SMOTE TAHUN 2012-2013

Kelas mayor	Kelas minor	Replikasi	Kelas mayor	Kelas minor baru
110(83%)	22 (17%)	4	110(50%)	110(50%)

Tabel 2 dan 3 menunjukkan hasil distribusi data simulasi SMOTE dimana jumlah data kelas 1 yang semula berjumlah 23 maka setelah direplikasi sebanyak 4 kali akan menjadi 115 data untuk data tahun 2013-2014 dan jumlah data kelas 1 yang semula berjumlah 22 maka setelah direplikasi sebanyak 4 kali akan menjadi 110 data untuk data tahun 2012-2013.

C. Regresi Logistik Biner

Regresi logistik biner dilakukan untuk mengetahui faktor-faktor apa saja yang mempengaruhi kondisi *financial distress* pada perusahaan manufaktur yang tercatat di BEI . Pertama-tama dilihat nilai korelasi antar variabel prediktor apakah memiliki nilai korelasi yang tinggi antar variabel prediktor yang kemudian dilihat nilai VIF untuk menduga apakah terjadi kasus multikolinearitas. Nilai korelasi antar variabel prediktor disajikan dalam Tabel 4.

TABEL 4. TABEL KORELASI ANTAR VARIABEL PREDIKTOR TAHUN 2012-2013

	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
X ₁						
X ₂	-0.692					
	0.000					
X ₃	0.151	0.139				
	0.026	0.039				
X ₄	-0.597	0.171	-0.244			
	0.000	0.011	0.000			
X ₅	0.011	-0.127	0.120	-0.020		
	0.875	0.061	0.076	0.770		
X ₆	0.170	0.088	0.742	-0.204	-0.110	
	0.012	0.195	0.000	0.002	0.102	
X ₇	0.154	-0.004	0.666	-0.091	-0.074	0.896
	0.022	0.958	0.000	0.178	0.276	0.000

Ket: rasio likuiditas (X₁), rasio profitabilitas (X₂), rasio rentabilitas (X₃)
Rasio solvabilitas(X₄),rasio aktivitas (X₅),ROI (X₆), ROE (X₇)

Berdasarkan Tabel 4 dapat diketahui bahwa terdapat korelasi yang tinggi antar variable, hal tersebut dapat mengindikasikan adanya kasus multikolinearitas dan dapat menyebabkan pengujian parameter tidak signifikan. Untuk itu dilakukan deteksi multikolinearitas untuk mengetahui apakah terjadi kasus multikolinearitas dengan melihat nilai VIF yang ditunjukkan pada Tabel 5.

TABEL 5. NILAI VIF MASING-MASING VARIABEL PREDIKTOR

Variabel Prediktor	VIF
Rasio Likuiditas (X ₁)	3.844
Rasio Profitabilitas (X ₂)	2.783
Rasio Rentabilitas (X ₃)	2.643
Rasio Solvabilitas (X ₄)	1.972
Rasio Aktivitas (X ₅)	1.175
ROI (X ₆)	7.046
ROE (X ₇)	5.634

Berdasarkan Tabel 5 dapat diketahui bahwa VIF seluruh variabel prediktor sudah kurang dari 10, artinya tidak terjadi indikasi kasus multikolinearitas. Selanjutnya dilakukan pengujian parameter secara serentak dan parsial.

1) Pengujian Parameter

Hipotesis

H₀ : β₁ = β₂= . . . = β_p=0

H₁ : minimal ada satu β₁= β₂= . . . = β_p ≠ 0

Daerah kritis

Tolak H₀ jika α < 0.05

Statistik Uji

TABEL 6. UJI SERENTAK TAHUN 2012-2013

	Chi-square	df	Sig.
Step	12.947	1	0.000
Step 2	23.259	2	0.000
Model	23.259	2	0.000

Berdasarkan Tabel 6 didapatkan nilai P-value pada model sebesar 0,000, jika dibandingkan dengan nilai α sebesar 0,05 maka diperoleh keputusan tolak H₀ yang artinya terdapat minimal satu variabel prediktor yang berpengaruh terhadap kondisi *financial distress* perusahaan manufaktur.

Hipotesis

H₀ : β_i = β_i = 0

H₁ : β_i ≠ 0 ; i = 1, 2, ..., p

Daerah kritis

Tolak H₀ jika α < 0.05

Statistik Uji

TABEL 7. UJI PARSIAL TAHUN 2012-2013

	B	S.E.	Wald	df	Sig.	Exp(B)	
X ₁	-1.925	0.609	9.999	1	0.002	0.146	
Step 2 ^b	X ₅	-1.842	0.586	9.893	1	0.002	0.158
	Constant	0.925	0.659	1.972	1	0.160	2.522

Ket: rasio likuiditas (X₁), rasio aktivitas (X₅)

$$\pi(x) = \frac{\exp(0,925 - 1,925x_1 - 1,842x_5)}{1 + \exp(0,925 - 1,925x_1 - 1,842x_5)}$$

Berdasarkan Tabel 7 telah dilakuakn uji signifikansi parameter dengan menggunakan subset dan diperoleh model terbaik dimana variabel yang berpengaruh signifikan terhadap kondisi finansial distress perusahaan manufaktur adalah variabel rasio likuiditas (X₁) dan rasio aktivitas (X₅). Hal ini dapat dilihat dari p-value yang kurang dari α = 5%.

2) Ketepatan Klasifikasi

TABEL 8. NILAI VIF MASING-MASING VARIABEL PREDIKTOR

Observed	Predicted						
	Data Traininig			Data Testing			
	y	Percentage	Correct	y	Percentage	Correct	
Step 2 Y	0	108	2	98.2	107	2	98.2
	1	18	4	18.2	19	4	17.4
Overall Percentage			84.8	84.1			

Berdasarkan Tabel 8 dapat diperoleh informasi kebenaran pengklasifikasian (*accuracy rate*) pada data *training* keseluruhan adalah 84,8% dan proporsi benar (*accuracy rate*) pada data *testing* keseluruhan adalah 84,1%.

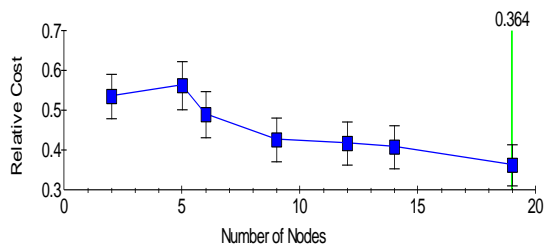
D. Klasifikasi Financial Distress dengan Metode CART

Sesuai pada algoritma yang telah dijelaskan pada bab dua, maka tahapan-tahapan yang dilakukan yaitu pembentukan pohon klasifikasi, dilanjutkan dengan pemangkasan pohon klasifikasi, dan terakhir yaitu penentuan pohon klasifikasi optimum. Variabel penting dalam pembentukan pohon klasifikasi maksimum disajikan pada Tabel 9.

Tabel 9 Variabel Penting Pembentukan Pohon Klasifikasi Maksimum

Variable	Score
Rasio Likuiditas (X_1)	100.00
Rasio Rentabilitas (X_3)	98.66
Rasio Solvabilitas (X_4)	94.49
ROI (X_6)	92.43
Rasio Aktivitas (X_5)	72.26
Rasio Profitabilitas (X_2)	49.91
ROE (X_7)	49.74

Berdasarkan skor variabel yang dihasilkan diketahui bahwa variabel X_1 mempunyai skor variabel tertinggi yaitu 100,00. Sehingga dapat dikatakan bahwa variabel rasio likuiditas (X_1) merupakan faktor terpenting dalam mengklasifikasikan finansial distress perusahaan manufaktur. Selain itu, terdapat beberapa variabel lain yang juga berpengaruh besar dalam pembentukan pohon klasifikasi yaitu rasio rentabilitas (X_3), rasio solvabilitas (X_4), ROI (X_6) dan rasio aktivitas (X_5).



Gambar 2. Plot *Relative Cost* dan Banyaknya Simpul Terminal

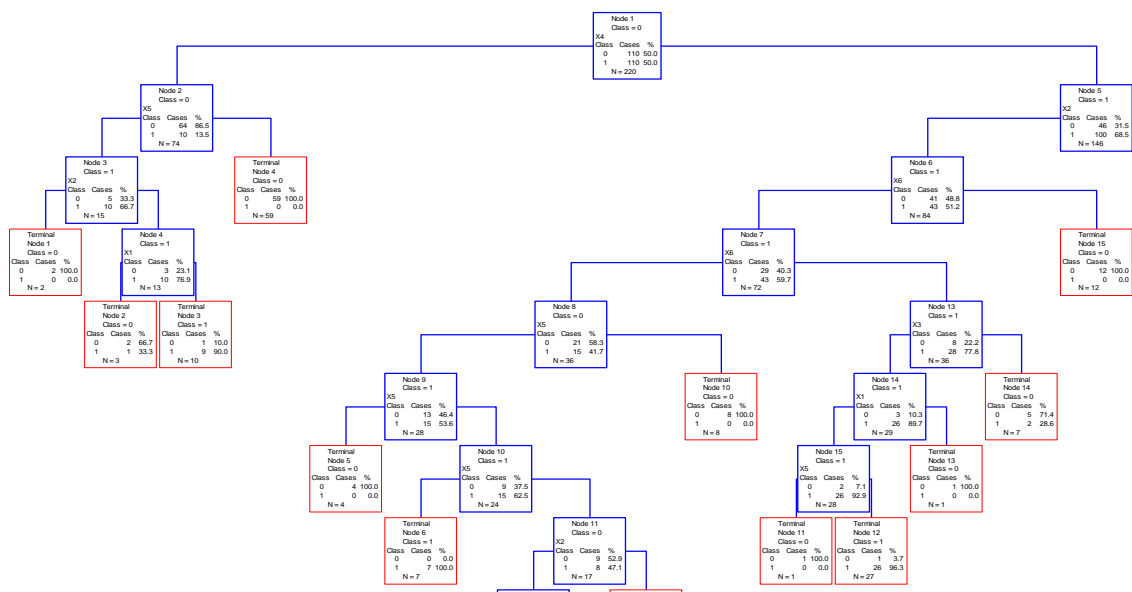
Jika dilihat pada *plot relative cost* pohon klasifikasi dengan simpul terminal sebanyak 19 simpul sudah dapat dikatakan optimal ditunjukkan dengan nilai *relative cost* yang dihasilkan sebesar 0,364, ditandai dengan garis berwarna hijau. Sedangkan, untuk biaya kesalahan (*cross-validation relative cost*) menghasilkan nilai yang paling minimum yakni sebesar $0,364 \pm 0,064$ atau

bekisar antara 0,400 sampai 0,300 dengan nilai kompleksitas sebesar 0,000.

Variabel rasio solvabilitas (X_4) memilah simpul utama (simpul 1) menjadi simpul kiri dan simpul kanan. Simpul akan memilah kekiri jika variabel rasio solvabilitas (X_4) $\leq 1,020$ (simpul 2) dan kekanan jika sebaliknya (simpul 5). Total sebanyak 74 data $\leq 1,020$ menjadi simpul kiri (simpul 2) dan 146 $> 1,020$ menjadi anggota simpul kanan (simpul 5). Simpul 2 yang beranggotakan 74 perusahaan manufaktur dengan nilai rasio solvabilitas $\leq 1,020$, selanjutnya dipilah menjadi simpul baru kiri dan kanan menurut rasio aktivitas (X_5). Jika nilai rasio aktivitas $\leq 0,599$ maka akan dipilah menjadi anggota simpul kiri baru (simpul 3), tetapi jika nilai rasio aktivitas $> 0,599$ maka perusahaan manufaktur tersebut akan dipilah menjadi simpul kanan baru (simpul terminal 4). Diantara 74 perusahaan manufaktur yang menjadi anggota simpul 2, diperoleh hasil bahwa terdapat 15 perusahaan manufaktur menjadi anggota simpul 3 dengan karakteristik nilai rasio solvabilitas (X_4) $\leq 1,020$ dan nilai rasio aktivitas $\leq 0,599$. Sisanya sebanyak 59 perusahaan manufaktur yang menjadi anggota simpul terminal 4 dengan karakteristik nilai rasio solvabilitas (X_4) $\leq 1,020$ dan nilai rasio aktivitas $> 0,599$.

Simpul 3 yang beranggotakan 15 perusahaan manufaktur dengan karakteristik nilai rasio solvabilitas (X_4) $\leq 1,020$ dan nilai rasio aktivitas $\leq 0,599$, selanjutnya dipilah menjadi simpul baru kiri dan kanan menurut rasio profitabilitas (X_2). Jika nilai rasio profitabilitas (X_2) $\leq 0,155$, maka perusahaan manufaktur tersebut akan dipilah ke simpul baru kiri (simpul terminal 1). Sedangkan jika nilai rasio profitabilitas (X_2) $> 0,155$, maka akan dipilah ke simpul kanan baru (simpul 4), seperti yang disajikan pada Gambar 3.

Karakteristik pada masing-masing kelas pada variabel respon dapat diketahui dengan menelusuri pohon klasifikasi optimal yang sudah terbentuk berdasarkan aturan pemilahan indeks gini. Karakteristik yang kuat didapatkan pada simpul terminal yang memiliki persentase kelas terbesar (100%) pada pelabelan kelas. Berikut adalah karakteristik masing-masing kelas yang disajikan pada Tabel 10.



Gambar 3. Potongan Struktur Pohon Klasifikasi Optimal

TABEL 10. KARAKTERISTIK KELAS PERUSAHAAN MANUFAKTUR MENURUT PRESENTASE KELAS TERBESAR

Kelas	Karakteristik
0 (perusahaan manufaktur tidak mengalami financial distress)	Rasio aktivitas $\leq 0,599$, dan rasio solvabilitas $\leq 1,020$
(1) Perusahaan Manufaktur Mengalami <i>Financial Distress</i>	Rasio likuiditas $\leq 0,634$, rasio profitabilitas $\leq 0,155$, rasio aktivitas $\leq 0,599$, dan rasio solvabilitas $\leq 1,020$

Berikut adalah perbandingan hasil ketepatan klasifikasi pohon optimal yang ditunjukkan pada Tabel 11.

TABEL 11. PERBANDINGAN KETEPATAN KLASIFIKASI POHON OPTIMAL SEBELUM DAN SESUDAH *PRE-PROCESSING* DENGAN SMOTE

Pohon Klasifikasi Optimal	Ketepatan Klasifikasi (%)	
	<i>Training</i>	<i>Testing</i>
Sebelum SMOTE	93,3	61,3
Sesudah SMOTE	96,3	82,1

Berdasarkan Tabel 11 diatas, akurasi klasifikasi *financial distress* perusahaan manufaktur setelah dilakukan *pre-processing* dengan SMOTE lebih tinggi, sehingga dapat dikatakan bahwa pohon klasifikasi optimal yang dihasilkan dengan jumlah simpul terminal sebanyak 19 simpul sudah layak, sehingga pohon klasifikasi tersebut dapat digunakan untuk mengklasifikasikan *financial distress* perusahaan manufaktur dengan menggunakan data yang baru. Perbandingan Hasil Klasifikasi Antara Metode Regresi Logistik Biner dan CART

Menurut Hosmer dan Lemeshow (2000) salah satu ukuran kebaikan model adalah jika memiliki peluang kesalahan klasifikasi yang minimal dan ketepatan prediksi dari model yang maksimal. Hasil perbandingan klasifikasi antara metode regresi logistik biner dan CART disajikan pada Tabel 12.

TABEL 12. PERBANDINGAN KETEPATAN KLASIFIKASI

Metode	Ketepatan Klasifikasi (%)	
	<i>Testing</i>	<i>Training</i>
RegLog Biner	84,8	84,1
CART	96,3	82,1

Berdasarkan Tabel 12 untuk data *training* diperoleh ketepatan klasifikasi analisis regresi logistik sebesar 84,8% dan ketepatan klasifikasi metode CART sebesar 96,3%. Hal ini menunjukkan bahwa metode CART memiliki nilai ketepatan klasifikasi yang lebih besar 11,5% dari metode CART. Akan tetapi dapat dikatakan bahwa kedua metode tersebut sudah cukup baik dalam memprediksi variabel respon dalam kasus ini kondisi finansial perusahaan manufaktur.

Perbedaan tingkat ketepatan prediksi dapat disebabkan oleh oleh perbedaan hasil pengklasifikasian. Dalam regresi logistik variabel prediktor yang terpilih sebagai variabel yang berpengaruh terhadap variabel respon adalah rasio likuiditas (X_1) dan rasio aktivitas (X_5). Sedangkan pada metode CART variabel yang terpilih adalah rasio profitabilitas (X_2), rasio rentabilitas (X_4), rasio solvabilitas (X_4), rasio aktivitas (X_5), dan ROI (X_6). Variabel rasio aktivitas (X_5) merupakan variabel yang konsisten berpengaruh pada analisis regresi logistik maupun metode CART.

V. KESIMPULAN DAN SARAN

Hasil perbandingan kedua metode yaitu untuk data *Testing* diperoleh ketepatan klasifikasi analisis regresi logistik sebesar 84,8% dan ketepatan klasifikasi metode CART sebesar 96,3%. Hal ini menunjukkan bahwa metode CART memiliki nilai ketepatan klasifikasi yang lebih besar 11,5% dari metode regresi logistik biner. Akan tetapi dapat dikatakan bahwa kedua metode tersebut sudah cukup baik dalam memprediksi variabel respon dalam kasus ini kondisi finansial perusahaan manufaktur.

Saran untuk peneliti selanjutnya, penambahan variabel prediktor dengan skala nominal sangat disarankan jika memakai metode yang sama dengan alasan kemudahan dalam membaca output untuk karakteristik kelas.

VI. DAFTAR PUSTAKA

- [1] Kompasiana. 2013. http://www.kompasiana.com/virginiafladian/dampak-krisis-global-terhadap-ketenagakerjaan-diindonesia_55282627f17e61a01d8b4603. Diakses Tanggal 21 November 2015.
- [2] Ardhinita, Nur A. 2009. "Manfaat Rasio Keuangan Untuk Memprediksi Kondisi *Financial distress* Pada Perusahaan Manufaktur Yang Terdaftar Di Bursa Efek Indonesia" Universitas Muhammadiyah Surakarta, Surakarta.
- [3] Sun, Jie, and Li, Hui, 2011. "Dynamic Financial Distress Prediction Using Instance Selection for The Disposal of Concept Drift", 2011
- [4] Breiman L, Friedman J, Olshen R, dan Stone C. 1993. "Classification and Regression Trees". Chapman Hall : New York - London.
- [5] W. Hosmer and S. Lemeshow. "Applied Logistic Regression". Wiley, New York, 1989
- [6] Agresti, A. (1990). "Categorical Data Analysis". New York: John Wiley dan Sons, Inc.