

PERINGATAN !!!

*Bismillaahirrahmaanirrahiim
Assalamu'alaikum warahmatullaahi wabarakaatuh*

1. **Skripsi digital ini hanya digunakan sebagai bahan referensi**
2. **Cantumkanlah sumber referensi secara lengkap bila Anda mengutip dari Dokumen ini**
3. **Plagiarisme** dalam bentuk apapun merupakan pelanggaran keras terhadap etika moral penyusunan karya ilmiah
4. **Patuhilah etika penulisan karya ilmiah**

Selamat membaca !!!

Wassalamu'alaikum warahmatullaahi wabarakaatuh



LAPORAN TAHUN TERAKHIR

PENELITIAN FUNDAMENTAL



PENGEMBANGAN MODEL STATISTIKA KEMACETAN KREDIT PERBANKAN

YANG MEMPERTIMBANGKAN PELUANG DAN WAKTU TERJADINYA

KEMACETAN

17 01 16

Tahun ke-2 dari rencana 2 tahun

Tim Peneliti:

Ketua : ABDUL KUDUS, S.SI., M.SI., PH.D. (NIDN 0421036901)

Anggota : Dr. SULIADI, S.SI., M.SI. (NIDN 0416117202)

Anggota : Dr. PUPUNG PURNAMASARI, SE., M.SI. Ak., CA (NIDN 0416046901)

UNIVERSITAS ISLAM BANDUNG

NOVEMBER 2016

HALAMAN PENGESAHAN

Judul : Pengembangan Model Statistika Kemacetan Kredit
Perbankan yang Mempertimbangkan Peluang dan Waktu
Terjadinya Kemacetan

Peneliti/Pelaksana

Nama Lengkap : ABDUL KUDUS S.Si., M.Si., Ph.D.
Perguruan Tinggi : Universitas Islam Bandung
NIDN : 0421036901
Jabatan Fungsional : Lektor
Program Studi : Statistika
Nomor HP : 082118025411
Alamat surel (e-mail) : akudus69@yahoo.com

Anggota (1)

Nama Lengkap : Dr SULIADI S.Si, M.Si
NIDN : 0416117202
Perguruan Tinggi : Universitas Islam Bandung

Anggota (2)

Nama Lengkap : PUPUNG PURNAMASARI S.E., M.Si.
NIDN : 0416046901
Perguruan Tinggi : Universitas Islam Bandung

Institusi Mitra (jika ada)

Nama Institusi Mitra

Alamat

Penanggung Jawab

Tahun Pelaksanaan

Biaya Tahun Berjalan

Biaya Keseluruhan

-

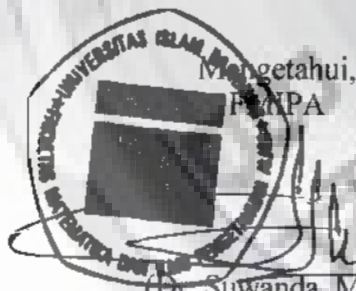
-

: Tahun ke 2 dari rencana 2 tahun

: Rp 50.000.000,00

: Rp 125.000.000,00

17 6189



(Dr. Suwanda, MSi.)
NIP/NIK D.86.0.045

Bandung, 30 - 11 - 2016
Ketua,

(ABDUL KUDUS S.Si., M.Si., Ph.D.)
NIP/NIK D.95.0.223

Menyetujui,
Ketua LPPM



(Prof. Edi Setiadi, SH., M.H.)
NIP/NIK 195911101987031002

RINGKASAN

Penelitian ini bertujuan untuk mengembangkan model-model statistika tentang peluang dan waktu terjadinya kredit macet. Model statistika yang dapat membedakan antara calon debitur yang macet dengan yang lancar dirumuskan dan dibangun. Model tersebut termasuk dalam metode *credit scoring*. Metode ini merupakan suatu alat dan teknik prediksi yang membantu lembaga keuangan dalam pemberian kredit. Tujuan pembuatan *credit scoring* untuk menganalisis dan membuat keputusan yang lebih cepat, tepat dan efisien terhadap penyeleksian nasabah yang akan menerima kredit. Regresi logistik merupakan teknik yang sudah sangat sering digunakan dalam membangun *credit scoring* di sebagian lembaga keuangan. Model regresi logistik akan memodelkan mengenai peluang terjadinya peristiwa “macet”. Di lain pihak, perhatian utama bukan pada berapa peluang kejadian peristiwa “macet”, melainkan juga pada kapan peristiwa itu terjadi. Analisis survival dapat membangun *credit scoring* dengan mempertimbangkan kapan peristiwa “macet” terjadi. Analisis survival fokus pada pemodelan peluang mengenai terjadinya suatu peristiwa yang dipantau sejak dari awal sampai pada munculnya kejadian yang diperhatikan, yaitu “macet”. Model Cox proporsional hazard adalah salah satu model khusus dalam analisis survival yang bertujuan untuk memodelkan data survival dengan variabel penjelas yang mempengaruhi fungsi hazard. Melalui model Cox proporsional hazard dapat dilihat hubungan antara variabel bebas terhadap variabel tak bebas yaitu waktu survival melalui fungsi hazardnya. *Credit scoring* yang dihasilkan oleh model regresi logistik maupun Cox proporsional hazard akan dievaluasi kelayakannya berdasarkan kriteria *mean cost* dengan *cut-off* tertentu. *Cut-off* adalah nilai untuk menentukan bagaimana prediksi dari setiap individu yang diteliti, sedangkan *mean cost* adalah ongkos atau kerugian akibat salah memprediksi. Dimana terdapat dua jenis salah prediksi yaitu salah prediksi jenis I dan salah prediksi jenis II. Salah prediksi jenis I adalah menyatakan prediksinya sebagai “lancar” padahal sebenarnya “macet” dan salah prediksi jenis II adalah menyatakan prediksinya sebagai “macet” padahal “lancar”. Kerugian akibat masing-masing salah prediksi tersebut tentu saja berbeda, dimana salah prediksi jenis I mempunyai *cost* yang lebih besar dibandingkan salah prediksi jenis II. Dengan hasil analisis bahwa nilai *mean cost* model regresi logistik lebih kecil dari pada nilai *mean cost* model Cox *proportional hazard*, artinya model regresi logistik sedikit lebih baik dari pada model Cox *proportional hazard*.

Kajian selanjutnya adalah mengenai masalah estimasi proporsi debitur KPR yang mengalami macet dalam membayar cicilannya. Model *Cure* dengan mengasumsikan data penelitian berdistribusi peluang Log-logistik digunakan untuk melakukan estimasi tersebut. Pengaruh kovariat terhadap variabel respon juga dapat dimodelkan dengan menggunakan model regresi *Cure* Log-logistik dengan estimasi parameter model menggunakan algoritme EM. Hasil perhitungan menunjukkan sebesar 39% taksiran proporsi debitur dinyatakan macet tanpa dipengaruhi oleh kovariat. Sedangkan dengan pengaruh berbagai kovariat memberikan hasil sebesar 83% taksiran proporsi debitur dinyatakan macet dengan pengaruh kovariat numerik dan kovariat kategorik berkategori 0, dan sebesar 46% taksiran proporsi debitur dinyatakan macet dengan pengaruh kovariat numerik dan kovariat kategorik berkategori 1. Hasil pengujian signifikansi parameter regresi *Cure* Log-logistik menunjukkan bahwa kovariat nilai nominal dan kovariat status tempat tinggal berpengaruh terhadap peluang terjadinya peristiwa macet dan kovariat *loan to value* berpengaruh terhadap waktu mencicil debitur.

Kata kunci : *Credit scoring*, regresi logistik, Cox *proportional hazard*, *cut-off* optimal, *mean cost*, distribusi Log-Logistik, model regresi *Cure*-Log-Logistik, Algoritme EM.

PRAKATA

Alhamdulillah rabbil'alamin, puji dan syukur penulis panjatkan kehadirat Allah SWT atas nikmat hidup yang dipersembahkan olehnya kepada penulis, sehingga penulis dapat menyusun laporan akhir penugasan penelitian. Tidak lupa shalawat serta salam yang selalu tercurahkan pada Nabi Muhammad SAW.

Laporan kemajuan penelitian yang berjudul "**Pengembangan Model Statistika Kemacetan Kredit Perbankan yang Mempertimbangkan Peluang dan Waktu Terjadinya Kemacetan**" ini diharapkan dapat menambah pengetahuan bagi para pembacanya.

Penyusunan laporan akhir ini tidak akan bisa selesai tanpa bantuan dan dukungan dari berbagai pihak, oleh karena itu penulis ucapkan terima kasih yang mendalam untuk semua pihak yang telah berkenan memberi bantuan dan dukungan demi proses penyusunan laporan ini. Mereka yang telah membantu semoga mendapat kebaikan yang berlimpah dari Allah SWT, Aamiin.

Penulis menyadari laporan akhir ini masih jauh dari kata sempurna, hal ini dikarenakan keterbatasan pengetahuan, pengalaman, maupun kemampuan yang dimiliki. Oleh karena itu, penulis mengharapkan semua pihak agar dapat memberikan saran-saran penyempurnaan yang berguna bagi penulis khususnya dan juga bagi pembaca pada umumnya.

Akhir kata, penulis selipkan do'a semoga Allah SWT berkenan membalas budi baik serta melimpahkan rahmatnya kepada pihak yang telah memberikan dukungan dan bantuan kepada penulis, *Aamiin Yarabbal'alamin*.

Bandung, 30 November 2016

Penulis

DAFTAR ISI

Halaman

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

BAB 1. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	2
1.3 Urgensi (Keutamaan) Penelitian	3
1.4 Target Luaran	3
BAB 2. TINJAUAN PUSTAKA	4
2.1 State of the Art	4
2.2 Regresi Logistik	4
2.3 Analisis Survival	9
2.4 Model Cox <i>proportional hazard</i>	15
2.5 Taksiran Fungsi Survival dalam Model Cox <i>proportional hazard</i>	19
2.6 Kriteria <i>Mean Cost</i>	20
2.7 Kurva <i>Receiver Operating Characteristic</i> (ROC)	22
2.8. Distribusi Log-logistik	24
2.9 Model <i>Cure</i>	26
2.10 Model <i>Cure</i> Log-logistik	27
2.11 Model Regresi <i>Cure</i> Log-logistik	28
2.12 Penaksiran Parameter Model <i>Cure</i> Log-logistik	30
2.13 Penaksiran Parameter Model Regresi <i>Cure</i> Log-Logistik	33
2.14 Pengujian Signifikansi Parameter Secara Simultan	36
2.15 Pengujian Signifikansi Parameter	36
2.16 Kredit Perbankan dan Kredit Bermasalah	37
2.17 Road map	39

BAB 3. TUJUAN DAN MANFAAT PENELITIAN	40
3.1 Tujuan Penelitian	40
3.2 Manfaat Penelitian	40
 BAB 4. METODE PENELITIAN	 41
4.1 Alur Penelitian	41
4.2 Luaran dan Indikator Capaian	41
 BAB 5. HASIL DAN LUARAN YANG DICAPAI	 43
5.1 Hasil Penelitian	43
5.2 Data Training dan Data Testing	43
5.3 Model Regresi Logistik	43
5.4 Model Cox <i>Proportional Hazard</i>	50
5.5 Perbandingan Nilai <i>Mean Cost</i> Model Regresi Logistik dan Cox <i>Proportional Hazard</i>	58
5.6 Model <i>Cure</i> Log-logistik	58
5.7 Analisis Ketahanan Mencicil Debitur Tanpa Kovariat	59
5.8 Model Regresi <i>Cure</i> Log-logistik	60
5.9 Pengujian Signifikansi Parameter Model Secara Simultan	63
5.10 Pengujian Signifikansi Parameter Model Secara Parsial	64
5.11 Analisis Proporsi Debitur Macet dengan Pengaruh Kovariat.....	65
5.12 Luaran Penelitian.....	67
 BAB 6. KESIMPULAN DAN SARAN	 68
6.1 Kesimpulan	68
6.2 Saran	70
 DAFTAR PUSTAKA	 71
LAMPIRAN	74

DAFTAR TABEL

	Halaman
Tabel 2.1 Matriks cost	21
Tabel 2.2 Tabel Klasifikasi antara y dan \hat{y}	21
Tabel 2.3 Peringkat Kredit Berdasarkan Ketepatan Pembayaran	39
Tabel 5.1 Taksiran Parameter untuk Model Regresi Logistik	44
Tabel 5.2 Nilai Statistik Uji Wald dan P-value Model Regresi Logistik	46
Tabel 5.3 Nilai Taksiran Peluang Kredit Macet ($\hat{\pi}$) untuk Seluruh Debitur yang Termasuk dalam Data Training	47
Tabel 5.4 Proses Mencari Cut-off Optimal Model Regresi Logistik	48
Tabel 5.5 Taksiran Nilai \hat{y} Model Regresi Logistik untuk Seluruh Debitur dalam Data Training	48
Tabel 5.6 Tabel Klasifikasi Pada Data Training	49
Tabel 5.7 Nilai Taksiran Peluang Kredit Macet ($\hat{\pi}$) untuk Seluruh Debitur yang Termasuk dalam Data Testing	49
Tabel 5.8 Taksiran Nilai \hat{y} Model Regresi Logistik untuk Seluruh Debitur dalam Data Testing	50
Tabel 5.9 Tabel Klasifikasi Pada Data Testing	50
Tabel 5.10 Taksiran Parameter untuk Model Cox proportional hazard	51
Tabel 5.11 Nilai Statistik Uji Wald dan P-value Model Cox proportional hazard	53
Tabel 5.12 Nilai Taksiran Peluang Kredit Macet ($1 - \hat{S}(12)$) untuk Seluruh Debitur yang Termasuk dalam Data Training	55
Tabel 5.13 Proses Mencari Cut-off Optimal Model Cox PH	55
Tabel 5.14 Taksiran Nilai $\hat{\delta}$ Model Cox PH untuk Seluruh Debitur dalam Data Training	56
Tabel 5.15 Tabel Klasifikasi Pada Data Training	57
Tabel 5.16 Nilai Taksiran Peluang Kredit Macet ($1 - \hat{S}(12)$) untuk Seluruh Debitur yang Termasuk dalam Data Testing	57
Tabel 5.17 Taksiran Nilai $\hat{\delta}$ Model Cox PH untuk Seluruh Debitur dalam Data Testing	57
Tabel 5.18 Tabel Klasifikasi Pada Data Testing	57
Tabel 5.19 Nilai Mean Cost Model Regresi Logistik dan Cox Proportional Hazard	58
Tabel 5.20 Taksiran Parameter untuk Model Regresi Cure Log-logistik	61
Tabel 5.21 Nilai Statistik Uji dan P-Value Model Regresi Cure Log-logistik	64

DAFTAR GAMBAR

Gambar 2.1 Bentuk dari confusion matrix untuk cutpoint c tertentu	22
Gambar 2.2 Kurva ROC	23
Gambar 2.3 Roadmap penelitian	39
Gambar 4.1 Flow chart beserta indikator dan luaran penelitian	42
Gambar 5.1 Kurva Fungsi Ketahanan Marginal	59



DAFTAR LAMPIRAN

CV peneliti	74
Luaran artikel ilmiah dan sertifikat presentasi	86



BAB 1. PENDAHULUAN

1.1 Latar Belakang

Analisis survival adalah metode statistika yang mempelajari pemodelan lamanya masa hidup sampai terjadinya suatu kejadian atau peristiwa. Variabel acak waktu ketahanan hidup (*survival time*), T , merupakan waktu dari awal perlakuan sampai terjadinya kejadian yang pertama kali yang ingin diamati. Kejadian yang dimaksud berupa suatu peristiwa yang ingin diketahui dan mungkin saja belum terjadi pada saat pengumpulan data berakhir sehingga waktu survival-nya tidak teramati secara lengkap. Pada kondisi demikian, pengamatan tersebut dikatakan sebagai pengamatan tersensor. Metode regresi survival adalah metode regresi yang digunakan untuk melihat faktor-faktor yang mempengaruhi peluang terjadinya suatu peristiwa seiring waktu (biasa dikenal dengan nama *covariate*) dengan variabel responnya adalah waktu ketahanan hidup (Cantor, 2003). Salah satu metode regresi survival yang sering digunakan adalah regresi Cox *proportional hazard*.

Analisis survival banyak diterapkan kajian bidang ilmu kesehatan, sedangkan dalam kajian bidang lain dikenal dengan istilah yang berbeda seperti: *event history analysis* dalam bidang sosiologi, analisis reliabilitas atau *failure time analysis* dalam bidang teknik dan industri, dan *duration analysis* atau *transition analysis* dalam bidang ekonomi. Metode tersebut menggunakan pendekatan metode parametrik atau non parametrik.

Dalam kajian bidang ekonomi, salah satu penerapannya adalah dalam perbankan yang salah satu kegiatannya yang dominan adalah penyaluran kredit. Masyarakat yang akan mengajukan kredit harus memenuhi beberapa persyaratan yang sudah ditentukan oleh pihak bank. Pemenuhan terhadap persyaratan tersebut menjadi bahan utama dalam pertimbangan persetujuan kredit yang diajukan. Di samping itu pihak bank juga masih akan melakukan analisis kredit untuk melihat kemampuan dalam pembayaran kredit untuk menghindari kemungkinan terjadinya suatu keadaan dimana seorang debitur tidak mampu membayar lunas kredit bank tepat pada waktunya.

Untuk meminimalisir risiko terjadinya kredit macet, pihak bank menerapkan suatu sistem uji kelayakan kredit yang disebut dengan *credit scoring* atau dikenal juga dengan istilah *credit risk rating*. Prinsip analisis risiko kredit yang digunakan pada metode ini adalah analisis 5C (*The five C's of credit analysis*), yaitu penilaian terhadap *character* (watak), *capacity* (kemampuan), *capital* (modal), *collateral* (jaminan), dan *condition of economy* (kondisi ekonomi). Secara umum, *credit scoring* merupakan suatu metode analisis risiko kredit yang bertujuan untuk menilai dan membedakan risiko kredit (Herli, 2013).

Selain masalah risiko kredit macet, hal lainnya adalah masalah kapan terjadinya kemacetan tersebut. Dalam analisis survival, pemodelan tidak hanya difokuskan pada peluang terjadinya kemacetan akan tetapi juga pada kapan terjadinya kemacetan. Pemodelan ini akan menghasilkan analisis yang lebih tajam dan lebih bermakna bagi pihak perbankan, oleh karena risiko kerugian yang diakibatkan oleh kemacetan yang terjadi pada awal masa cicilan akan lebih besar dari pada kemacetan yang terjadi di fase akhir masa cicilan.

Allen & Rose (2006) menggunakan metode analisis survival dalam pengembangan model *credit scoring*. Model tersebut menggunakan model waktu kegagalan untuk data pinjaman pribadi. Dia menemukan bahwa keputusan diterimanya pengajuan kredit lebih baik dimodelkan dengan menggunakan metode analisis survival dibandingkan dengan regresi multipel. Selain itu dia menyarankan menggunakan analisis survival pada strategi penagihan tunggakan, dimana survival time-nya adalah lamanya waktu yang diperlukan sampai lunas dengan menggunakan berbagai strategi penagihan.

Tong *et al.* (2012) mengembangkan lebih jauh mengenai model *credit scoring*. Dia membandingkan hasil penggunaan model regresi logistik dengan model-model yang dikembangkan dalam analisis survival (*Cox proportional hazards* dan model *mixture cure*), dan hasilnya adalah metode analisis survival lebih unggul.

Pengembangan model dapat dilakukan lebih lanjut dengan mempertimbangkan adanya fenomena debitur macet yang tidak dapat sepenuhnya dimodelkan dengan model survival biasa. Model survival biasa mengasumsikan bahwa semua debitur akan mengalami kejadian yang diperhatikan, yakni semua debitur akan macet. Hal ini sangat tidak sesuai, oleh karena sebagian dari debitur justru diyakini akan dapat melunasi kreditnya. Masalah ini harus dipertimbangkan dalam pengembangan model survival dengan memodifikasi model yang ada agar lebih sesuai dengan fenomena yang ada. Model tersebut termasuk dalam jenis model *cure*.

1.2 Perumusan Masalah

Dalam penelitian akan dicari solusi bagi beberapa masalah berikut:

1. Bagaimana pengembangan model statistika tentang masalah kredit macet yang dikembangkan atas dasar metode-metode yang berasal dari analisis survival? Dua jenis model yang akan dikaji yakni model *survival tree* dan model yang mempertimbangkan keberadaan sebagian debitur yang tidak mengalami macet (model *cure*).
2. Bagaimana perbandingan kinerja dari model yang diusulkan dengan model yang ada?

1.3 Urgensi (Keutamaan) Penelitian

Dalam situasi yang penuh dengan berbagai kemungkinan, maka akan sangat baik jika kemungkinan-kemungkinan tersebut dapat dikuantifikasi. Seperti halnya kemungkinan kerugian yang dihadapi oleh perbankan akibat dari kejadian kredit macet yang seharusnya dapat dikuantifikasi agar kerugiannya dapat diminimalisir. Proses kuantifikasi tersebut dapat dilakukan dengan membangun model statistika yang dicocokkan dengan data empiris yang ada sehingga terbentuk model taksiran yang dapat digunakan untuk mengkuantifikasi kemungkinan-kemungkinan yang akan terjadi pada masa yang akan datang.

Dalam pemodelan kredit macet, selain masalah pengkuantifikasian berapa peluang terjadinya kredit macet juga penting untuk memperkirakan kapan kredit macet tersebut akan terjadi. Antisipasi awal terhadap kemungkinan tersebut akan membantu perbankan terhindar dari kerugian yang lebih besar. Untuk membantu mengantisipasi kerugian tersebut pihak manajemen dapat menggunakan *tools* statistika yang dibangun berdasarkan teori dan dicocokkan dengan data empiris yang ada. Oleh karena itu, akan sangat berguna jika sejak awal pihak bank sudah mempunyai *tools* yang dapat digunakan untuk memilah calon debitur yang baik dan calon debitur yang berpotensi macet.

1.4 Target Luaran

Target luaran penelitian ini adalah model statistika untuk pemodelan kredit macet, yakni 1) model yang dikembangkan dari model regresi logistik, dan 2) model yang dikembangkan dari model regresi survival. Kedua model tersebut akan dibandingkan kinerjanya dengan model yang ada dalam hal kemampuannya untuk memilah calon debitur yang baik dan yang berpotensi macet. Dari dua luaran tersebut dapat disusun artikel ilmiah yang membahas cara untuk memilah dan memilah calon debitur perbankan.

BAB 2. TINJAUAN PUSTAKA

2.1 State of the Art

Sebelum melakukan pembahasan mengenai permasalahan dari penelitian ini, pada bab ini akan diuraikan beberapa teori penunjang yang dapat membantu dalam pelaksanaan penelitian ini. Teori penunjang tersebut adalah: Regresi logistik, analisis *survival*, *model Cox proportional hazard*, kriteria *mean cost*, kurva ROC untuk mencari *cut-off* optimal, serta kredit perbankan dan kredit bermasalah.

2.2 Regresi Logistik

Istilah model linear umum atau generalized linear model (GLM) biasanya merujuk pada model regresi biasa untuk variabel tak bebas kontinu pada variabel bebas kontinu dan/atau kategorik. Istilah GLM merujuk pada kelas model yang lebih besar yang dipopulerkan oleh McCullagh dan Nelder (1989). Dalam model ini, variabel respon y_i diasumsikan mengikuti distribusi dari keluarga eksponensial dengan rata-rata μ_i , yang biasanya diasumsikan sebagai suatu fungsi (seringkali bentuknya nonlinear) dari $\mathbf{x}_i^T \boldsymbol{\beta}$. Beberapa penulis mengatakan bentuknya adalah nonlinear karena μ_i seringkali merupakan fungsi non linear dari kovariat, tetapi McCullagh dan Nelder (1989) mempertimbangkan fungsi tersebut sebagai bentuk yang linear, karena kovariat ini mempengaruhi distribusi dari y_i hanya melalui kombinasi linear dari $\mathbf{x}_i^T \boldsymbol{\beta}$. Seluruh model linear umum akan mempunyai tiga buah komponen. Komponen acak yang mengidentifikasi variabel respon y yang mempunyai distribusi peluang tertentu yang berasal dari keluarga eksponensial. Komponen sistematis yang berisi variabel bebas x yang digunakan sebagai prediktor dalam model. Fungsi penghubung yang menggambarkan hubungan fungsional antara komponen sistematis dengan nilai harapan (rata-rata) dari komponen acak. Model linear umum akan menghubungkan fungsi dari rata-rata tersebut kepada variabel bebas melalui persamaan yang berbentuk linear.

Regresi logistik merupakan metode analisis statistika yang digunakan untuk menganalisis hubungan antara variabel tak bebas yang bersifat biner atau dikotomus dengan satu atau lebih variabel bebas (Hosmer dan Lemeshow, 2000). Pada regresi logistik, variabel tak bebas berskala kategorik. Variabel tak bebas yang dinotasikan dengan y bersifat biner atau dikotomus yang mempunyai dua nilai yaitu 0 dan 1. Dalam keadaan demikian, variabel y mengikuti distribusi Bernoulli untuk setiap observasi tunggal. Fungsi probabilitas untuk setiap observasi diberikan sebagai berikut:

$$f(y) = \pi^y(1 - \pi)^{1-y} \quad y = 0 \text{ dan } 1 \quad \dots(2.1)$$

sehingga diperoleh:

$$\text{Jika } y = 0 \text{ maka } f(y) = \pi^0(1 - \pi)^{1-0} = 1 - \pi$$

$$\text{Jika } y = 1 \text{ maka } f(y) = \pi^1(1 - \pi)^{1-1} = \pi$$

$$E(y_i) = \pi$$

Fungsi regresi logistiknya dapat dituliskan sebagai berikut:

$$\pi(x_i) = \frac{e^{(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})}}{1 + e^{(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})}} \quad \dots(2.2)$$

Jika $\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}$ dengan $p =$ banyak variabel bebas. Maka nilai η_i antara $-\infty$ dan $+\infty$ sehingga nilai π terletak antara 0 dan 1. Hal tersebut menunjukkan bahwa model logistik sebenarnya menggambarkan probabilitas atau risiko dari suatu objek.

Dimana fungsi $\pi(x_i)$ di atas berbentuk non linear sehingga untuk membuatnya menjadi fungsi linear harus dilakukan transformasi logit sebagai berikut:

$$g(x_i) = \ln\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \quad \dots(2.3)$$

2.2.1 Pendugaan Parameter Model Regresi Logistik

Pendugaan parameter yang digunakan regresi logistik adalah metode kemungkinan maksimum (*maximum likelihood*, ML). Metode kemungkinan maksimum adalah metode

penaksiran yang banyak digunakan untuk menaksir model logistik, baik untuk data dikelompokkan dan data yang tidak dikelompokkan (*ungrouped data*). Prinsip dasar dari metode kemungkinan maksimum adalah memilih suatu penaksir parameter sedemikian rupa sehingga dapat memaksimumkan fungsi peluang yang diamati. Jika x_i dan y_i adalah pasangan variabel bebas dan tak bebas pada pengamatan ke $-i$ dan diasumsikan bahwa setiap pasangan pengamatan saling bebas dengan pasangan pengamatan lainnya, $i = 1, 2, \dots, n$, maka fungsi probabilitas untuk setiap pasangan adalah sebagai berikut:

$$f(y_i) = \pi_i^{y_i}(1 - \pi_i)^{1-y_i} \quad ; y_i = 0 \text{ dan } 1 \quad \dots(2.4)$$

Jika peluang sukses dimodelkan dengan variabel bebas x_i mengikuti model persamaan regresi logistik, fungsi *likelihood* untuk distribusi Bernoulli adalah:

$$\begin{aligned} l(\beta) &= \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \\ &= \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \\ &= \prod_{i=1}^n \left(\frac{\pi_i}{1 - \pi_i} \right)^{y_i} (1 - \pi_i) \end{aligned}$$

dengan persamaan logistik $\pi(x_i) = \frac{e^{(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})}}{1 + e^{(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})}} = \frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}}$

$$1 - \pi_i = \frac{e^{(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})}}{1 + e^{(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})}} = 1 - \frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}} = (1 + e^{x_i^T \beta})^{-1}$$

Dari persamaan diatas maka diperoleh $\frac{\pi_i}{1 - \pi_i} = e^{x_i^T \beta}$. Sehingga fungsi *likelihood* dapat diperoleh menjadi:

$$l(\beta) = \prod_{i=1}^n e^{x_i^T \beta y_i} (1 + e^{x_i^T \beta})^{-1} \quad \dots(2.5)$$

Setelah fungsi *likelihood* didapat, langkah selanjutnya yaitu memperoleh nilai log-*likelihood* yang dapat dinyatakan sebagai berikut:

$$L(\beta) = \ln l(\beta)$$

$$L(\beta) = \ln \left[\prod_{i=1}^n e^{x_i^T \beta y_i} (1 + e^{x_i^T \beta})^{-1} \right]$$

$$L(\beta) = \sum_{i=1}^n [x_i^T \beta y_i - \ln(1 + e^{x_i^T \beta})] \quad \dots(2.6)$$

Untuk mendapatkan nilai penaksiran koefisien regresi logistik ($\hat{\beta}$) dilakukan dengan penurunan $L(\beta)$ terhadap β dan disamakan dengan 0 sebanyak p buah. Turunan pertama dari $x_i^T \beta$ terhadap β_j adalah x_{ij} , sehingga:

$$\begin{aligned} \frac{\partial L(\beta)}{\partial \beta_j} &= \sum_{i=1}^n y_i x_{ij} - \sum_{i=1}^n \left(\frac{1}{1 + x_i^T \beta} \right) e^{x_i^T \beta} x_{ij} \\ &= \sum_{i=1}^n [y_i x_{ij} - (\pi_i) x_{ij}] \\ &= \sum_{i=1}^n (y_i - \mu_i) x_{ij} \quad \dots(2.7) \end{aligned}$$

dimana $\mu_i = E(y_i) = \pi_i$.

Persamaan diatas merupakan sistem persamaan non linier, maka langkah selanjutnya untuk melakukan penyelesaian penaksiran parameter digunakan iterasi *Newton Raphson*. Langkah-langkah yang ditempuh adalah sebagai berikut:

1. Tentukan nilai awal taksiran parameter dalam model, $\beta^{(0)}$.
2. Tentukan vektor skor statistik ($U(\beta)$) dan matriks informasi ($I(\beta)$). Dalam hal ini, vektor statistik merupakan turunan pertama dari fungsi *log-likelihood* terhadap parameter-parameternya atau $\frac{\partial \log L(\beta)}{\partial \beta} = U(\beta)$. Sedangkan matriks informasi turunan kedua dari fungsi *log-likelihood* terhadap parameter-parameter atau $-\frac{\partial^2 L(\beta)}{\partial \beta^2} = I(\beta)$.
3. Untuk iterasi ke- m hitung:

$$\underline{\beta}^{(m+1)} = \underline{\beta}^{(m)} - [I(\beta^{(m)})]^{-1} U(\beta^{(m)}), \text{ untuk } m = 0, 1, 2, \dots \quad \dots(2.29)$$

4. Apabila selisih mutlak antara $\underline{\beta}^{(m)}$ dengan $\underline{\beta}^{(m+1)}$ mendekati nol maka proses iterasi berhenti. Tetapi jika $|\underline{\beta}^{(m)} - \underline{\beta}^{(m+1)}| > 0,0001$. Maka ulangi langkah ke-2 dan langkah ke-3 sampai memperoleh $|\underline{\beta}^{(m)} - \underline{\beta}^{(m+1)}| \leq 0,0001$.

2.2.2 Pengujian Parameter Model Regresi Logistik

Pengujian parameter model dilakukan untuk memeriksa signifikansi variabel bebas yang ada dalam model. Untuk mengetahui pengaruh variabel bebas terhadap variabel tidak bebas secara serentak di dalam model regresi logistik maka digunakan statistik uji G yang merupakan uji rasio kemungkinan (*likelihood ratio test*) dengan hipotesis :

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

$$H_1 : \text{paling tidak terdapat satu } \beta_j \neq 0 \quad ; j = 1, 2, \dots, p$$

Untuk menguji hipotesis digunakan statistik uji G :

$$G = -2 \ln \frac{\binom{n_1}{n_1} \binom{n_0}{n_0}}{\prod_{i=1}^n \hat{\pi}_i^{y_i} (1-\hat{\pi}_i)^{(1-y_i)}} = -2 \ln \left[\frac{L_0}{L_p} \right] \quad \dots(2.9)$$

dengan :

$$n_1 = \sum y_i$$

$$n_0 = \sum (1 - y_i)$$

$$\hat{\pi}_i = \frac{e^{(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})}}{1 + e^{(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})}}$$

L_0 = *likelihood* tanpa variabel bebas

L_p = *likelihood* dengan variabel bebas

Kriteria uji :

Tolak H_0 apabila $G > \chi^2_{(1-\alpha, p)}$ atau $p\text{-value} \leq \alpha$.

Dengan α : taraf arti

Selanjutnya dilakukan pengujian parameter secara parsial dengan menggunakan uji *Wald* dengan hipotesis sebagai berikut :

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0 \quad ; j = 1, 2, \dots, p$$

Statistik uji Wald didefinisikan (Hosmer dan Lemeshow, 2000) sebagai :

$$W = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad \dots(2.10)$$

Dengan $\hat{\beta}_j$ merupakan penaksir dari β_j sedangkan $SE(\hat{\beta}_j)$ adalah penaksir standar error dari $\hat{\beta}_j$. Statistik uji Wald akan mengikuti distribusi normal baku jika H_0 benar sehingga kriteria uji Wald yaitu tolak H_0 apabila $|W| > Z_{\alpha/2}$ atau p-value $\leq \alpha$. Sedangkan dalam *software* SPSS, *output* yang dikeluarkan adalah W^2 berdistribusi chi-kuadrat dengan derajat bebas 1 sehingga kriteria uji Wald yaitu tolak H_0 apabila $|W| > \chi^2_{(\alpha,1)}$ atau p-value $\leq \alpha$.

2.3 Analisis Survival

Analisis *survival* adalah metode-metode analisis statistika tentang lamanya ketahanan hidup sampai terjadinya suatu peristiwa. Analisis *survival* telah menjadi alat penting untuk menganalisis data waktu antar kejadian (*time to event data*) atau menganalisis data yang berhubungan dengan waktu, mulai dari *time origin* sampai terjadinya suatu peristiwa khusus. Kejadian khusus (*failure event*) tersebut dapat berupa kegagalan, kematian, kambuhnya suatu penyakit, respon dari suatu percobaan, atau peristiwa lain yang dipilih sesuai dengan kepentingan peneliti. Peristiwa khusus tersebut dapat berupa kejadian positif seperti kelahiran, kelulusan sekolah, kesembuhan dari suatu penyakit (Kleinbaum & Klein, 2005).

Analisis *survival* banyak diterapkan dalam bidang biologi, kedokteran, kesehatan umum seperti daya hidup pasien kanker paru-paru, sosiologi, teknik, seperti menganalisis masa hidup lampu pijar, ekonomi, demografi, dan epidemiologi (Collett, 2003).

Data *survival* merupakan data tentang pengamatan jangka waktu dari awal pengamatan sampai terjadinya suatu peristiwa. Waktu *survival* dapat didefinisikan sebagai waktu dari awal pengamatan hingga terjadinya peristiwa gagal, dapat dalam hari bulan, maupun tahun. Waktu awal (*time origin* atau *start-point*) yaitu waktu pada saat terjadinya kejadian awal, seperti waktu seorang divonis menderita kanker, waktu pemberian perlakuan dan lain-lain. Waktu kegagalan

(*failure time* atau *end-point*) yaitu waktu pada saat terjadinya kejadian akhir seperti kematian, kejadian dan lain-lain (Collett, 2003).

Waktu *survival* mengukur lamanya waktu sampai terjadinya peristiwa yang diperhatikan seperti kegagalan, macet, kematian, kambuh, perkembangan pemberian obat terhadap suatu penyakit, dan sebagainya (Purnama, 2011). Lamanya waktu ini merupakan variabel acak yang diasumsikan mengikuti distribusi tertentu. Distribusi dari waktu *survival* biasanya digambarkan oleh dua fungsi, yaitu fungsi *survival* dan fungsi *hazard* yang mempunyai hubungan satu sama lain.

Penentuan waktu *survival*, ada tiga faktor yang dibutuhkan.

- a. Waktu awal pencatatan (*time origin* atau *start-point*) harus didefinisikan dengan tepat pada setiap individu, misalkan awal mula pengamatan berupa tanggal perawatan pasien.
- b. Waktu akhir pencatatan (*failure time* atau *end-point*) didefinisikan jelas untuk mengetahui status tersensor atau tidak tersensor, meninggal atau sembuh seorang pasien.
- c. Satuan pengukuran sebagai batas dari waktu kejadian dari awal sampai akhir kejadian, misalnya satuan tahunan, bulanan, harian, mingguan, harian. Data tersensor merupakan data yang tidak bisa diamati secara utuh, karena adanya individu yang hilang ataupun dengan alasan lain, sehingga tidak dapat diambil datanya sampai akhir pengamatan. Dengan kata lain, pada akhir pengamatan individu tersebut belum mengalami peristiwa tertentu. Jika berada dalam keadaan sebaliknya maka data tersebut disebut data tidak tersensor (Lee & Wang, 2003).

Dalam mendapatkan data *survival* sering dijumpai suatu individu tidak mengalami kejadian sampai batas waktu pengamatan. Biasanya untuk mendapatkan data *survival* yang lengkap sampai semua individu mengalami kejadian membutuhkan waktu yang lama sehingga pengamatan yang dilakukan tidak efektif dan mengakibatkan biaya yang dikeluarkan sangat

banyak. Untuk mengatasi hal tersebut maka perlu dilakukan penyensoran data. Konsep penyensoran inilah yang membedakan antara analisis *survival* dengan ilmu-ilmu statistika yang lainnya (Kleinbaum & Klein, 2005).

Menurut Klein & Moeschberger (2003) dalam analisis *survival* terdapat empat jenis penyensoran.

a. Penyensoran kanan (*right censoring*)

Penyensoran terjadi jika objek pengamatan atau individu yang diamati masih tetap hidup pada saat waktu yang telah ditentukan. Dengan kata lain individu tersebut belum mengalami kejadian sampai akhir periode pengamatan, sedangkan waktu awal dari objek pengamatan dapat diamati secara penuh. Sebagai contoh, seorang pasien kanker diamati dari awal perawatan sampai akhir perawatan ternyata pasien tersebut masih hidup. Kemudian pasien melanjutkan perawatan di luar negeri sehingga tidak bisa diamati lagi (*lost to follow up*). Pasien ini memiliki waktu *survival* setidaknya beberapa waktu. Sehingga waktu pengamatan individu tersebut dikatakan penyensoran kanan.

b. Penyensoran kiri (*left censoring*)

Penyensoran kiri terjadi jika semua informasi yang diinginkan diketahui dari seseorang individu telah diperoleh pada awal pengamatan. Dengan kata lain pada saat waktu awal pengamatan individu tidak teramati pada awal pengamatan sementara kejadian dapat diamati secara penuh sebelum penelitian berakhir. Sebagai contoh, dalam sebuah penelitian untuk menentukan sebaran pengguna ganja di kalangan anak laki-laki di sebuah sekolah. Dengan mengajukan pertanyaan “kapan pertama kali anda menggunakan ganja?”. Ternyata terdapat beberapa anak menjawab “saya pernah menggunakannya, tetapi saya tidak tahu tepatnya kapan pertama kali menggunakannya”, pada kasus ini anak tersebut mengalami penyensoran kiri.

c. **Penyensoran selang (*interval censoring*)**

Penyensoran selang terjadi jika informasi yang dibutuhkan telah dapat diketahui pada kejadian peristiwa di dalam selang pengamatan atau penyensoran yang waktu daya tahannya berada dalam suatu selang tertentu. Sebagai contoh, beberapa tikus yang diberikan karsinogen pada makanannya, dilakukan studi selama 10 bulan kepada 10 tikus dan penelitian dilakukan setiap akhir tahun, jika 2 dari 8 tikus tewas karena kanker pada bulan ke-5 dan ke-7, maka dua tikus tersebut mengalami penyensoran selang.

d. **Penyensoran acak (*random censoring*)**

Penyensoran acak terjadi jika individu yang diamati meninggal atau mengalami kejadian karena sebab yang lain, bukan disebabkan dari tujuan utama penelitian. Sebagai contoh, 10 tikus yang diberikan zat karsinogen pada makanannya. Pada saat pengamatan ada 1 dari 10 tikus tersebut meninggal karena terjepit (tewas bukan karena penelitian utama) bukan karena terkena kanker, maka tikus tersebut mengalami pensensoran acak.

Penyensoran-penyensoran di atas disebabkan oleh beberapa hal antara lain: (Kleinbaum & Klein, 2005).

- a. *Loss to follow up*, objek menghilang selama masa pengamatan terjadi apabila individu pindah atau menolak untuk berpartisipasi.
- b. Individu tidak mengalami kejadian gagal (*failure event*) sebelum pengamatan berakhir.
- c. Individu terpaksa dihentikan dari pengamatan karena kematian (jika kematian bukan *failure event*) atau disebabkan alasan lain.

2.3.1 Fungsi *Survival*

Fungsi kepadatan peluang adalah peluang suatu individu mati atau gagal dalam interval waktu t sampai $t + \Delta t$. Fungsi kepadatan peluang dinotasikan dengan $f(t)$ dan dirumuskan dengan

$$f(t) = \lim_{\Delta t \rightarrow 0} \left[\frac{P(t < T < (t + \Delta t))}{\Delta t} \right] = \lim_{\Delta t \rightarrow 0} \left[\frac{P(F(t + \Delta t) - F(t))}{\Delta t} \right] \quad \dots(2.11)$$

Misalkan T adalah variabel acak bukan negatif pada interval $[0, \infty)$ yang menunjukkan waktu hidup pada suatu populasi dan $f(t)$ merupakan fungsi kepadatan peluang dari variabel acak T maka fungsi distribusi kumulatif $F(t)$ adalah (Lawless, 1982)

$$\begin{aligned} F(t) &= P(T \leq t) \\ &= \int_0^t f(x) dx \end{aligned} \quad \dots(2.12)$$

Dari persamaan (2.12), diperoleh

$$f(t) = \frac{d(F(t))}{dt} = F'(t) \quad \dots(2.13)$$

Jika T merupakan variabel acak tidak negatif pada interval $[0, \infty)$ yang waktu individu sampai mengalami kejadian pada populasi, $f(t)$ merupakan fungsi kepadatan peluang dari t maka peluang suatu individu tidak mengalami kejadian sampai waktu t dinyatakan dengan fungsi *survival* $S(t)$ (Lawless, 2007). Dalam penelitian ini, fungsi $S(t)$ tersebut menyatakan peluang seorang debitur tetap bertahan melakukan cicilan atas pinjamannya (tidak macet) pada waktu t .

$$\begin{aligned} S(t) &= P(\text{individu bertahan pada waktu } t) \\ S(t) &= P(T \geq t) \\ &= \int_t^{\infty} f(x) dx \end{aligned} \quad \dots(2.14)$$

Dari definisi fungsi distribusi kumulatif dari T , fungsi *survival* dapat dinyatakan sebagai berikut.

$$\begin{aligned} S(t) &= P(T \geq t) \\ &= 1 - P(\text{individu gagal sebelum waktu } t) \end{aligned}$$

$$= 1 - P(T \leq t)$$

$$= 1 - F(t)$$

$$F(t) = 1 - S(t)$$

$$\frac{d(F(t))}{dt} = \frac{d(1 - S(t))}{dt}$$

$$f(t) = -\frac{d(S(t))}{dt} = -S'(t) \quad \dots(2.15)$$

Hubungan kepadatan peluang, fungsi distribusi kumulatif dari T dengan fungsi *survival* yaitu

$$f(t) = F'(t) = -S'(t) \quad \dots(2.16)$$

2.3.2 Fungsi Hazard

Misalkan T variabel acak non negatif pada interval $[0, \infty)$ yang menunjukkan waktu individu sampai mengalami kejadian pada suatu populasi, maka peluang bahwa individu mengalami kejadian pada interval $(t, t + \Delta t)$ dinyatakan dengan fungsi *hazard* $h(t)$ (Lawless, 2007). Sehingga fungsi hazard dalam penelitian ini adalah peluang debitur yang belum mengalami macet sampai dengan t bulan akan mengalami macet dalam waktu dekat.

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

$$= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t, T \geq t)}{\Delta t \cdot P(T \geq t)}$$

$$= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t \cdot S(t)}$$

$$= \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t \cdot S(t)}$$

$$= \frac{1}{S(t)} \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t}$$

$$\begin{aligned}
 &= \frac{F'(t)}{S(t)} \\
 &= \frac{f(t)}{S(t)} \quad \dots(2.17)
 \end{aligned}$$

2.3.3 Hazard Kumulatif

Dari hasil substitusi persamaan (2.15) dan (2.17) diperoleh sebagai berikut (Lawless, 2007)

$$h(t) = \frac{-S'(t)}{S(t)} = \frac{d}{dt} \log S(t) \quad \dots(2.18)$$

Berdasarkan persamaan (2.19) diperoleh (Lawless, 2007)

$$\log(S(t)) \Big|_0^t = - \int_0^t h(x) dx \quad \dots(2.19)$$

karena $S(0) = 1$ sehingga (Lawless, 2007)

$$S(t) = \exp \left[- \int_0^t h(x) dx \right] \quad \dots(2.20)$$

Dari persamaan (2.21) didapatkan fungsi *hazard* maka fungsi kumulatif *hazard* dinyatakan dengan $H(t)$ (Lawless, 2007).

$$H(t) = \int_0^t h(x) dx \quad \dots(2.21)$$

Selain itu persamaan (2.21) dapat dituliskan (Lawless, 2007).

$$S(t) = \exp[-H(t)] \quad \dots(2.22)$$

2.4 Model Cox proportional hazard

Untuk memodelkan data survival dengan variabel penjelas yang mempengaruhi fungsi hazard adalah model hazard proporsional yang diusulkan oleh Cox dan juga dikenal sebagai regresi Cox. Kelebihan Cox adalah tidak harus memiliki fungsi dari distribusi parametrik. Asumsi pemodelan hanya memvalidasi asumsi bahwa fungsi hazard harus proporsional setiap

waktu. Asumsi proporsional pada model dapat diketahui melalui plot $-\log [-\log S(t)]$ terhadap waktu survival (t) untuk setiap kategori yang ada dalam p variabel penjelas yang membentuk pola yang sejajar pada level yang berbeda-beda.

Misal hazard tergantung pada nilai-nilai x_1, x_2, \dots, x_p dari p variabel penjelas, X_1, X_2, \dots, X_p . Nilai-nilai dari variabel penjelas dalam model hazard proporsional dinyatakan dalam bentuk vektor \mathbf{x} , sehingga $\mathbf{x} = (x_1, x_2, \dots, x_p)'$. Misalkan $h_0(t)$ adalah fungsi hazard untuk individu yang semua variabel penjelas vektor \mathbf{x} mempunyai nilai nol, maka fungsi $h_0(t)$ disebut *baseline hazard function*.

Model hazard proporsional umum adalah sebagai berikut

$$h_i(t) = \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) h_0(t) \quad \dots(2.23)$$

dapat pula dinyatakan dalam bentuk persamaan log,

$$\log \left\{ \frac{h_i(t)}{h_0(t)} \right\} = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \quad \dots(2.24)$$

dengan $h_i(t)$: Fungsi hazard untuk individu ke- i

$h_0(t)$: Fungsi hazard dasar (baseline)

$\beta_1, \beta_2, \dots, \beta_p$: Koefisien regresi

$x_{i1}, x_{i2}, \dots, x_{ip}$: Nilai variabel untuk individu ke- i

2.4.1 Pendugaan Parameter Model *Cox proportional hazard*

Untuk menentukan model terbaik diperlukan taksiran koefisien variabel X_1, X_2, \dots, X_p yaitu $\beta_1, \beta_2, \dots, \beta_p$. Koefisien β dalam model *Cox proportional hazard* dapat ditaksir menggunakan metode *Maksimum Likelihood*. Apabila terdapat n individu, di antaranya terdapat r individu yang tidak tersensor dan $n - r$ individu yang tersensor maka urutan waktu r waktu kegagalan dinotasikan oleh $t_{(1)} < t_{(2)} < \dots < t_{(r)}$, sehingga $t_{(q)}$ adalah urutan waktu kegagalan ke- q . Menurut Cox (1972) fungsi *likelihood* untuk model *hazard* proporsional adalah

$$L(\beta) = \prod_{q=1}^r \frac{\exp(x_q^T \beta)}{\sum_{i \in R(t_q)} \exp(x_i^T \beta)} \quad \dots(2.25)$$

$x_{(q)}$ adalah vektor variabel dari individu yang gagal pada saat ke- q dengan waktu t_q . Notasi $R(t_q)$ adalah seluruh individu yang memiliki resiko gagal pada waktu ke- q . Jika terdapat n waktu *survival* yang diobservasi, dinotasikan oleh t_1, t_2, \dots, t_n dan δ_i adalah *value indicator* maka fungsi likelihoodnya dinyatakan dalam fungsi parsial *likelihood* sebagai berikut

$$L(\beta) = \prod_{i=1}^n \left[\frac{\exp(x_i^T \beta)}{\sum_{i \in R(t_q)} \exp(x_i^T \beta)} \right]^{\delta_i} \quad \dots(2.26)$$

dengan $\delta_i = \begin{cases} 0, & \text{individu yang tersensor} \\ 1, & \text{individu tidak tersensor} \end{cases}$

Fungsi *log likelihood* yang bersesuaian adalah

$$\log L(\beta) = \sum_{i=1}^n \delta_i \{ x_i^T \beta - \log \sum_{i \in R(t_q)} \exp(x_i^T \beta) \} \quad \dots(2.27)$$

Untuk mendapatkan nilai penaksiran $\hat{\beta}$ dilakukan dengan penurunan $L(\beta)$ terhadap β dan disamakan dengan 0. Turunan pertama dari $x_i^T \beta$ terhadap β_j adalah x_{ij} , sehingga:

$$\begin{aligned} \frac{\partial \ln L(\beta)}{\partial \beta_j} &= \frac{\partial (\sum_{i=1}^n \delta_i \{ (\sum_{j=1}^p \beta_j x_{ij}) - \ln (\sum_{i \in R(t_q)} \exp(\sum_{j=1}^p \beta_j x_{ij})) \})}{\partial \beta_j} \\ &= \sum_{i=1}^n \delta_i \left[\sum_{j=1}^p x_{ij} \frac{\sum_{i \in R(t_q)} \exp(\sum_{j=1}^p \beta_j x_{ij}) \sum_{j=1}^p x_{ij}}{\sum_{i \in R(t_q)} \exp(\sum_{j=1}^p \beta_j x_{ij})} \right] \quad \dots(2.28) \end{aligned}$$

Persamaan diatas merupakan sistem persamaan non linier, maka langkah selanjutnya estimasi koefisien $\hat{\beta}$ diselesaikan dengan metode numerik melalui penyelesaian *Newton Raphson*. Langkah-langkah yang ditempuh adalah sebagai berikut:

1. Tentukan nilai awal taksiran parameter dalam model, $\beta^{(0)}$.
2. Tentukan vektor skor statistik ($U(\beta)$) dan matriks informasi ($I(\beta)$). Dalam hal ini, vektor statistik merupakan turunan pertama dari fungsi *log-likelihood* terhadap

parameter-parameternya atau $\frac{\partial \log L(\beta)}{\partial \beta} = U(\beta)$. Sedangkan matriks informasi turunan kedua dari fungsi *log-likelihood* terhadap parameter-parameter atau $-\frac{\partial^2 L(\beta)}{\partial \beta^2} = I(\beta)$.

3. Untuk iterasi ke- m hitung:

$$\underline{\beta}^{(m+1)} = \underline{\beta}^{(m)} - [I(\beta^{(m)})]^{-1} U(\beta^{(m)}), \text{ untuk } m = 0, 1, 2, \dots \quad \dots(2.29)$$

4. Apabila selisih mutlak antara $\underline{\beta}^{(m)}$ dengan $\underline{\beta}^{(m+1)}$ mendekati nol maka proses iterasi berhenti. Tetapi jika $|\underline{\beta}^{(m)} - \underline{\beta}^{(m+1)}| > 0,0001$. Maka ulangi langkah ke-2 dan langkah ke-3 sampai memperoleh $|\underline{\beta}^{(m)} - \underline{\beta}^{(m+1)}| \leq 0,0001$.

2.4.2 Pengujian Parameter Model *Cox proportional hazard*

Pengujian parameter model dilakukan untuk memeriksa signifikansi variabel bebas yang ada dalam model. Untuk mengetahui pengaruh variabel bebas terhadap variabel tidak bebas secara serentak di dalam model *Cox proportional hazard* maka digunakan statistik uji G yang merupakan uji rasio kemungkinan (*likelihood ratio test*) dengan hipotesis :

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

$$H_1 : \text{paling tidak terdapat satu } \beta_j \neq 0 \quad ; j = 1, 2, \dots, p$$

Untuk menguji hipotesis digunakan statistik uji G :

$$G = -2 \ln \left[\frac{L_0}{L_p} \right] \quad \dots(2.30)$$

dengan :

L_0 = *likelihood* tanpa variabel bebas

L_p = *likelihood* dengan variabel bebas

Kriteria uji :

Tolak H_0 apabila $G > \chi^2_{(1-\alpha;p)}$ atau $p\text{-value} \leq \alpha$.

Dengan α : taraf arti

Selanjutnya dilakukan pengujian parameter secara parsial dengan menggunakan uji *Wald* dengan hipotesis sebagai berikut :

$$H_0 : \beta_j = 0$$

$$H_0 : \beta_j \neq 0 \quad ; j = 1, 2, \dots, p$$

Statistik uji *Wald* didefinisikan (Hosmer dan Lemeshow, 2000) sebagai :

$$W = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad \dots(2.31)$$

Dengan $\hat{\beta}_j$ merupakan penaksir dari β_j sedangkan $SE(\hat{\beta}_j)$ adalah penaksir standar error dari $\hat{\beta}_j$. Statistik uji *Wald* akan mengikuti distribusi normal baku jika H_0 benar sehingga kriteria uji *Wald* yaitu tolak H_0 apabila $|W| > Z_{\alpha/2}$ atau p-value $\leq \alpha$. Sedangkan dalam *software* SPSS, *output* yang dikeluarkan adalah W^2 berdistribusi chi-kuadrat dengan derajat bebas 1 sehingga kriteria uji *Wald* yaitu tolak H_0 apabila $|W| > \chi^2_{(\alpha,1)}$ atau p-value $\leq \alpha$.

2.5 Taksiran Fungsi *Survival* dalam Model *Cox proportional hazard*

Pada model *Cox proportional hazard* terdapat p variabel X_1, X_2, \dots, X_p dan taksiran koefisien dari variabel tersebut adalah $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p$ maka taksiran fungsi *hazard* ke- i adalah

$$\hat{h}_i(t) = \exp(x_i^T \hat{\beta}) \hat{h}_0(t) \quad \dots(2.32)$$

Besaran kontribusi terhadap fungsi *survival* dalam interval (t_{q-1}, t_q) adalah

$$s_q = \exp \left[- \int_{t_{q-1}}^{t_q} h_0(u) du \right] \quad \dots(2.33)$$

$$\hat{s}_q = \left[1 - \frac{\exp(x_i^T \hat{\beta})}{\sum_{l \in R(t_q)} \exp(x_l^T \hat{\beta})} \right]^{\exp(x_i^T \hat{\beta})} \quad \dots(2.34)$$

dimana s_q adalah peluang dasar dalam interval (t_{q-1}, t_q) .

Sedangkan peluang dasar dari fungsi *survival* dalam waktu t_q adalah

$$S_0(t_q) = \prod_{j=0}^{q-1} s_j \quad \dots(2.35)$$

dimana s_j adalah besaran kontribusi dari interval ke- j terhadap nilai fungsi *survival* dasar.

$$\hat{S}_0(t) = \prod_{t_q < t} \hat{s}_q = \prod_{t_q < t} \left[1 - \frac{\exp(x_i^T \hat{\beta})}{\sum_{i \in R(t_q)} \exp(x_i^T \hat{\beta})} \right]^{\exp(x_i^T \hat{\beta})} \quad \dots(2.36)$$

$$\hat{S}_0(t) = 0 \text{ jika } t > t_n$$

Fungsi *survival* dengan vektor kovariat x adalah

$$S(t_i; x) = [S_0(t_i)]^{\exp(x_i^T \beta)} = \left(\prod_{j=0}^{q-1} s_j \right)^{\exp(x_i^T \beta)} \quad \dots(2.37)$$

$$\hat{S}(t; x_i) = [\hat{S}_0(t)]^{\exp(x_i^T \hat{\beta})} \quad \dots(2.38)$$

Dalam penelitian ini, fungsi $\hat{S}(t; x_i)$ tersebut menyatakan peluang seorang debitur tetap bertahan melakukan cicilan atas pinjamannya (tidak macet) pada waktu t . Sedangkan $1 - \hat{S}(t; x_i)$ menyatakan peluang seorang debitur mengalami macet.

2.6 Kriteria *Mean Cost*

Mean cost merupakan risiko dari perbankan. Pengertian *mean cost* adalah ongkos atau kerugian akibat salah memprediksi. Makin kecil nilai *mean cost* maka model akan semakin bagus. Sebelum mengetahui nilai *mean cost*, akan ditetapkan nilai *cut-off* yang optimal. *Cut-off* adalah nilai untuk menentukan bagaimana prediksi dari setiap individu yang diteliti. Nilai *cut-off* biasanya ditentukan oleh masing-masing kebijakan bank yaitu sebesar 0,5. Tetapi agar prediksi tepat, maka harus ditetapkan nilai *cut-off* optimal.

Jika peluang debitur mengalami macet \geq *cut-off* optimal maka $\hat{y} = 1$ dan

Jika peluang debitur mengalami macet $<$ *cut-off* optimal maka $\hat{y} = 0$

Dengan *matriks cost*:

Tabel 2.1 Matriks cost

		Prediksi (\hat{y})	
		Lancar (0)	Macet (1)
Aktual (y)	Lancar (0)	0	1
	Macet (1)	20	0

Matriks cost diatas menunjukkan besar biaya yang digunakan atas kesalahan melakukan prediksi:

- Seseorang yang diprediksi dengan benar menimbulkan biaya 0
- Seseorang yang aktualnya lancar diprediksi macet menimbulkan biaya 1 (salah prediksi jenis II)
- Seseorang yang aktualnya macet diprediksi lancar menimbulkan biaya 20 (salah prediksi jenis I)

Dengan tabel kontingensi antara y (aktual) dan \hat{y} (prediksi):

Tabel 2.2 Tabel Klasifikasi antara y dan \hat{y}

		\hat{y}	
		0	1
y	0	f_{00}	f_{01}
	1	f_{10}	f_{11}

dimana f_{00} : banyaknya individu yang diprediksi dengan benar

f_{01} : banyaknya individu yang aktualnya lancar diprediksi macet

f_{10} : banyaknya individu yang aktualnya macet diprediksi lancar

f_{11} : banyaknya individu yang diprediksi dengan benar

$$\text{Maka Mean cost} = \frac{(0 \times f_{00}) + (1 \times f_{01}) + (20 \times f_{10}) + (0 \times f_{11})}{n}$$

$$= \frac{f_{01} + 20f_{10}}{n}$$

...(2.39)

2.7 Kurva Receiver Operating Characteristic (ROC)

Kurva ROC adalah plot kombinasi nilai sensitivitas dengan nilai 1-spesifisitas dengan berbagai *cut-off* yang mungkin. Suatu model dikatakan sempurna jika memprediksi semua debitur baik yang diwakili oleh garis horizontal atau tingkat positif benar. Sebuah titik di dalam kurva ROC ditentukan dengan terlebih dahulu membentuk tabel klasifikasi dengan *cut-off* tertentu. Adapun penentuan *cut-off*-nya dilakukan berdasarkan nilai peluang macet ($\hat{\pi}$). Setelah nilai peluang macet ($\hat{\pi}$) diperoleh, kemudian hitung nilai sensitivitas dan spesifisitas terlebih dahulu. Hasil tabulasi silang disebut *confusion matrix*.

Bentuk dari *confusion matrix* diperlihatkan pada Gambar 2.1

		Prediksi (\hat{y})		Jumlah Total Kolom
		+	-	
		(Positif)	(Negatif)	
Aktual (y)	+	TP	FN	P
	-	FP	TN	N

Gambar 2.1 Bentuk dari *confusion matrix* untuk *cutpoint* c tertentu

Dengan :

Benar Positif (*True Positives* (TP)) : Yang diprediksi positif dan sebenarnya positif

Salah Positif (*False Positives* (FP)) : Yang diprediksi positif dan sebenarnya negatif

Salah Negatif (*False Negatives* (FN)) : Yang diprediksi negatif dan sebenarnya positif

Benar Negatif (*True Negatives* (TN)) : Yang diprediksi negatif dan sebenarnya negatif

P : Jumlah untuk total kolom positif

N : Jumlah untuk total kolom negatif

Beberapa parameter pengukur kinerja ditunjukkan dengan Persamaan (2.40) sampai dengan Persamaan (2.46).

$$FP\ rate = \frac{FP}{N} \quad \dots(2.40)$$

$$TP\ rate = \frac{TP}{P} = Recall \quad \dots(2.41)$$

$$Presisi = \frac{TP}{TP+FP} \quad \dots(2.42)$$

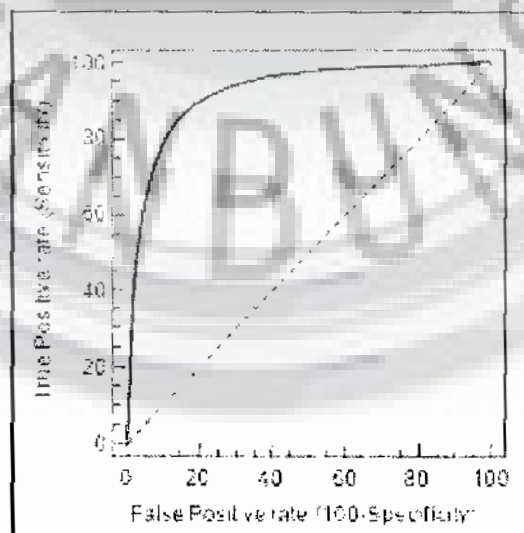
$$Akurasi = \frac{TP+TN}{P+N} \quad \dots(2.43)$$

$$Sensitivitas = Recall \quad \dots(2.44)$$

$$Spesifisitas = \frac{TN}{FP+TN} = \frac{TN}{N} = 1 - FP\ rate \quad \dots(2.45)$$

$$1 - Spesifisitas = 1 - (1 - FP\ rate) = FP\ rate \quad \dots(2.46)$$

Kurva *Receiver Operating Characteristic* (ROC) dibentuk berdasar pasangan koordinat 1-spesifisitas dan sensitivitas untuk berbagai nilai *cut-off*, seperti pada Gambar 2.2.



Gambar 2.2. Kurva ROC

Hasil semua pasangan sensitivitas dan spesifisitas untuk berbagai nilai *c* dapat digambarkan dalam plot. Plot ini sangat bermanfaat untuk menentukan *cut-off* optimal yang

akan digunakan untuk memprediksi kelas-kelas dari pengamatan, yakni dicari nilai $\hat{\pi}$ dimana selisih antara sensitivitas dan 1-spesifisitas adalah maksimum.

2.8. Distribusi Log-logistik

Distribusi Log-logistik merupakan salah satu distribusi peluang kontinu untuk variabel random *non-negatif* (Athoillah & Sudarno, 2012). Distribusi ini digunakan dalam analisis tahan hidup sebagai model parametrik, misalnya untuk meneliti waktu ketahanan mencicil debitur sebuah bank. Peubah acak T dikatakan mengikuti distribusi atau menyebar Log-logistik dengan parameter μ dan σ adalah jika fungsi densitas peluangnya sebagai berikut (Xian, 2012) :

$$f(t) = \frac{\sigma^{-1} e^{-\frac{\log t - \mu}{\sigma}}}{\left(1 + e^{-\frac{\log t - \mu}{\sigma}}\right)^2} \quad \dots(2.47)$$

Dimana $\theta = \frac{1}{\sigma}$ adalah parameter bentuk, dan $\phi = \frac{-\mu}{\sigma}$ adalah parameter lokasi. Momen ke- k untuk distribusi log-logistik (Klugman dkk, 2004) di atas adalah

$$E[T^k] = e^{k\mu} \Gamma(1+k\sigma) \Gamma(1-k\sigma) \quad ; \quad -\frac{1}{\sigma} < k < \frac{1}{\sigma} \quad \dots(2.48)$$

Fungsi distribusinya adalah

$$F(t) = \int_0^t \frac{\sigma^{-1} e^{-\frac{\log y - \mu}{\sigma}}}{\left(1 + e^{-\frac{\log y - \mu}{\sigma}}\right)^2} dy$$

Misal $u = 1 + e^{-\frac{\log t - \mu}{\sigma}}$

$$du = e^{-\frac{\log t - \mu}{\sigma}} \sigma^{-1} dt$$

$$dt = \frac{du}{e^{-\frac{\log t - \mu}{\sigma}} \sigma^{-1}}$$

maka ,

$$F(t) = \int_0^t \frac{\sigma^{-1} e^{-\frac{\log t - \mu}{\sigma}}}{\left(1 + e^{-\frac{\log t - \mu}{\sigma}}\right)^2} dy$$

$$= \int_0^t \frac{\sigma^{-1} e^{-\frac{\log t - \mu}{\sigma}}}{u^2} \frac{du}{\sigma^{-1} e^{-\frac{\log t - \mu}{\sigma}}}$$

$$= \int_0^t u^{-2} du$$

$$= -\left(1 + e^{-\frac{\log t - \mu}{\sigma}}\right)^{-1} \Big|_0^t$$

$$= -\frac{1}{1 + e^{-\frac{\log t - \mu}{\sigma}}} + 1$$

$$= 1 - \frac{1}{1 + e^{-\frac{\log t - \mu}{\sigma}}}$$

$$= \frac{1 + e^{-\frac{\log t - \mu}{\sigma}} - 1}{1 + e^{-\frac{\log t - \mu}{\sigma}}}$$

Jadi,

$$F(t) = \frac{e^{-\frac{\log t - \mu}{\sigma}}}{1 + e^{-\frac{\log t - \mu}{\sigma}}} \dots (2.49)$$

Fungsi survival dari distribusi Log-logistik didefinisikan sebagai berikut :

$$S(t) = 1 - F(t)$$

$$= 1 - \frac{e^{-\frac{\log t - \mu}{\sigma}}}{1 + e^{-\frac{\log t - \mu}{\sigma}}}$$

$$= \frac{1 + e^{-\frac{\log t - \mu}{\sigma}} - e^{-\frac{\log t - \mu}{\sigma}}}{1 + e^{-\frac{\log t - \mu}{\sigma}}}$$

$$S(t) = \frac{1}{1 + e^{\frac{\log t - \mu}{\sigma}}} \quad \dots(2.50)$$

Sedangkan fungsi hazard dari distribusi Log-logistik didefinisikan sebagai berikut :

$$h(t) = \frac{f(t)}{S(t)}$$

$$= \frac{\sigma^{-1} e^{\frac{\log t - \mu}{\sigma}}}{\left(1 + e^{\frac{\log t - \mu}{\sigma}}\right)^2} \cdot \frac{1}{1 + e^{\frac{\log t - \mu}{\sigma}}}$$

$$h(t) = \frac{\sigma^{-1} e^{\frac{\log t - \mu}{\sigma}}}{1 + e^{\frac{\log t - \mu}{\sigma}}} \quad \dots(2.51)$$

2.9 Model Cure

Model Cure adalah salah satu model khusus dalam analisis survival yang termasuk ke dalam metode parametrik, bertujuan untuk mengestimasi seberapa besar proporsi individu yang tidak pernah mengalami peristiwa yang diperhatikan (*survival time*), dalam jangka waktu tertentu yang cukup panjang (Abreu & Rocha, 2006). Model Cure berasumsi bahwa terdapat proporsi subyek yang tidak pernah mengalami peristiwa yang diperhatikan sehingga kurva survivalnya tidak akan mencapai nol (Cantor, 2003).

Pada umumnya, model dalam analisis survival tidak memperhatikan atau memperhitungkan beberapa individu dalam populasi yang tidak mengalami peristiwa. Dalam analisis survival, fungsi survival yang telah dibahas sebelumnya didasarkan pada distribusi yang biasa dimana $F(t) \rightarrow 1$ seiring $t \rightarrow \infty$, ini berarti $S(t) \rightarrow 0$ seiring $t \rightarrow \infty$. Dalam pemodelan ketahanan mencicil kredit bank, model tersebut dapat dikatakan lebih realistis apabila memiliki peluang tidak nol untuk peristiwa yang tidak terjadi (*cure*).

Terdapat dua kelas utama dalam model Cure, yaitu model campuran (*mixture*) dan model bukan campuran (*nonmixture*). Model Cure Mixture secara eksplisit memodelkan kelangsungan hidup yang terdiri dari 2 jenis subpopulasi individu yakni populasi individu yang mengalami peristiwa yang diperhatikan (*macet*) dan populasi individu yang tidak mengalami peristiwa yang diperhatikan (*tidak macet*). Biasanya, peluang seorang individu yang tidak mengalami macet dimodelkan dengan regresi logistik.

Misalkan π adalah proporsi untuk subpopulasi yang mengalami macet (*uncure*) atau mengalami peristiwa yang diperhatikan dan $1-\pi$ adalah proporsi subpopulasi yang tidak mengalami macet (*cure*), dimana U merupakan indikator yang menunjukkan seorang individu dapat mengalami macet ($U = 1$) dan ($U = 0$) menunjukkan seorang individu yang tidak mengalami macet, dan T adalah variabel acak *non-negatif* yang menunjukan waktu kegagalan yang didefinisikan hanya jika $U = 1$. Model Cure Mixture didefinisikan sebagai berikut (Corbiere & Joly, 2006) :

$$S(t|x,z) = \pi(z)S(t|U=1,x) + 1 - \pi(z) \quad \dots(2.52)$$

Dimana $S(t|x,z)$ adalah peluang bersyarat dari fungsi survival terhadap T untuk seluruh populasi, $S(t|U=1,x) = P(T>t|U=1,x)$ adalah peluang bersyarat dari fungsi survival untuk individu yang mengalami macet (*survival time*) yang mempunyai vektor kovariat $x = (x_1, \dots, x_p)'$, dan $\pi(z) = P(U=1|z)$ merupakan peluang dari peristiwa macet atau peristiwa yang diperhatikan itu sendiri sehingga mempunyai vektor kovariat $z = (z_1, \dots, z_q)'$, yang mana kedua kovariat tersebut bisa saja sama.

2.10 Model Cure Log-logistik

Model Cure Log-logistik memodelkan proporsi individu yang tidak mengalami peristiwa atau peristiwa yang diperhatikan (*tidak macet*), berdasarkan model distribusi peluang

Log-logistik. Dengan fungsi survival $S(t | U = 1) = \left[1 + \exp\left(\frac{\log t - \mu}{\sigma}\right) \right]^{-1}$ model Cure Log-logistik dapat didefinisikan sebagai berikut (Corbiere & Joly, 2006) :

$$S(t) = \pi \left[1 + \exp\left(\frac{\log t - \mu}{\sigma}\right) \right]^{-1} + 1 - \pi \quad \dots(2.52)$$

Fungsi peluang bersyarat dari fungsi survival terhadap T ketika sebagian populasi mengalami peristiwa yang diperhatikan (macet) dapat didefinisikan sebagai fungsi survival distribusi peluang log-logistik, sedangkan untuk sebagian populasi yang tidak mengalami peristiwa yang diperhatikan (tidak macet) yang mana $U = 0$, memiliki fungsi peluang bersyarat dari fungsi survival yang sama dengan 1.

2.11 Model Regresi Cure Log-logistik

Apabila terdapat beberapa kovariat dimana perubahan harga beberapa kovariat tersebut menyebabkan perubahan terhadap waktu bertahan hidup, dimungkinkan adanya suatu hubungan fungsional atau saling mempengaruhi antar kovariat terhadap waktu bertahan hidup tersebut. Hubungan fungsional dimodelkan dengan suatu model yang dinamakan model regresi. Model regresi Cure Log-logistik memodelkan hubungan fungsional antar kovariat dengan waktu bertahan hidup dimana terdapat sekelompok individu yang tidak mengalami peristiwa yang diperhatikan berdasarkan model distribusi peluang Log-logistik. Dalam model regresi Cure Log-logistik, vektor kovariat $x = (x_1, \dots, x_p)'$ mempengaruhi parameter μ , dimana (Corbiere & Joly, 2006) :

$$\mu = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \dots + \gamma_p x_p$$

$$\mu = \gamma' x \quad \dots(2.53)$$

γ mewakili vektor dari parameter regresi yang belum diketahui. Ketika x secara multipikatif mempengaruhi μ , maka x dapat meningkatkan atau bahkan menurunkan waktu

mencivil debitur. Model regresi Cure Log-logistik dimana kovariat x dapat mempengaruhi waktu mencivil debitur dapat didefinisikan sebagai berikut (Corbiere & Joly, 2006) :

$$S(t | x) = \pi \left[1 + \exp \left(\frac{\log t - \mu}{\sigma} \right) \right]^{-1} + 1 - \pi$$

Dengan $\mu = \gamma' x$, maka

$$S(t | x) = \pi \left[\frac{1}{1 + e^{\frac{\log t - \gamma' x}{\sigma}}} \right] + 1 - \pi \quad \dots(2.54)$$

Sedangkan pengaruh dari z dalam peluang $\pi(z)$ dapat dimodelkan dengan model regresi biner dengan logit sebagai berikut (Corbiere & Joly, 2006) :

$$\log \text{it}(\pi(z)) = \beta_0 + \beta_1 z_1 + \dots + \beta_q z_q = \beta' z$$

$$\log \left(\frac{\pi(z)}{1 - \pi(z)} \right) = \beta' z$$

$$\pi(z) = \frac{1}{1 + e^{-\beta' z}}$$

... (2.55)

Dimana,

Maka, model regresi Cure Log-logistik dimana kovariat x dapat mempengaruhi μ dan z dapat mempengaruhi $\pi(z)$ dapat didefinisikan sebagai berikut (Corbiere & Joly, 2006) :

$$S(t | x, z) = \left(\frac{1}{1 + e^{-\beta' z}} \right) \left(\frac{1}{1 + e^{\frac{\log t - \gamma' x}{\sigma}}} \right) + \left(1 - \frac{1}{1 + e^{-\beta' z}} \right) \quad \dots(2.56)$$

Dimana β_0 adalah intersep dan β adalah vektor dari parameter regresi yang berhubungan dengan z .

2.12 Penaksiran Parameter Model Cure Log-logistik

Parameter μ , σ dan π pada model Cure Log-logistik (pers 2.16) merupakan parameter yang belum diketahui dan akan ditaksir menggunakan metode kemungkinan maksimum. Fungsi kemungkinan adalah peluang bersama dari data pengamatan yang dianggap sebagai fungsi dari parameter yang tidak diketahui dalam asumsi model. Pendugaan parameter dengan metode kemungkinan maksimum adalah nilai ketika fungsi kemungkinan mencapai maksimum. Secara umum, fungsi kemungkinan untuk model Cure Log-logistik adalah sebagai berikut :

$$L(\mu, \sigma, \pi) = \prod_{i=1}^n \{(\pi_i) f(t_i | U = 1)\}^{\delta_i} \{(1 - \pi_i) + \pi_i S(t_i | U = 1)\}^{1 - \delta_i} \quad \dots (2.57)$$

Pada definisi sebelumnya, U merupakan variabel acak ketika individu mengalami mengalami macet jika ($U = 1$), dan tidak mengalami macet ketika ($U = 0$). Jika dikaitkan dengan indikator penyensoran δ , dimana jika $\delta_i = 1$ maka pasti $u_i = 1$. Artinya jika waktu mencicil seorang debitur teramati lengkap sampai terjadi peristiwa macet, maka debitur tersebut pasti mengalami peristiwa macet. Tetapi jika $\delta_i = 0$ maka nilai u_i akan mempunyai dua kemungkinan, yaitu $u_i = 1$ atau $u_i = 0$. Artinya jika waktu mencicil seorang debitur tidak teramati lengkap (tersensor), maka debitur tersebut mungkin mengalami macet atau mungkin lancar, sehingga dapat dikatakan pengamatan u_i tidak teramati (*missing*).

Ketika u_i merupakan nilai dari variabel acak U_i , maka fungsi log-kemungkinan yang lengkap adalah gabungan dari dua komponen yang saling bebas, yaitu l_f yang hanya bergantung pada π , dan l_s yang hanya bergantung pada μ dan σ (Corbiere & Joly, 2006) :

$$l_f(\pi; u) = \log \prod_{i=1}^n \pi^{u_i} (1 - \pi)^{1 - u_i} \quad \dots (2.58)$$

dan

$$l_s(\mu, \sigma; u) = \log \prod_{i=1}^n h(t_i | U = 1)^{\delta_i u_i} S(t_i | U = 1)^{u_i} \quad \dots (2.59)$$

dimana $S(t|U=1)$ merupakan fungsi peluang bersyarat dari fungsi survival terhadap T ketika sebagian populasi mengalami peristiwa yang diperhatikan (macet) dengan fungsi survival pada persamaan 2.13 :

$$S(t, |U = 1) = \frac{1}{1 + e^{\frac{\log t - \mu}{\sigma}}}$$

Dan $h(t|U=1)$ merupakan fungsi peluang bersyarat dari fungsi hazard terhadap T ketika sebagian populasi mengalami peristiwa yang diperhatikan (macet) dengan fungsi hazard pada pers 2.14 :

$$h(t | U = 1) = \frac{\sigma^{-1} e^{\frac{\log t - \mu}{\sigma}}}{1 + e^{\frac{\log t - \mu}{\sigma}}}$$

Untuk memaksimalkan fungsi log-kemungkinan di atas digunakan algoritme EM (*Expectation Maximization*) yang merupakan salah satu metode optimasi dalam memaksimalkan fungsi kemungkinan yang mengandung data tidak terobservasi (*missing*).

Tahapan algoritme EM adalah sebagai berikut :

- 1) Tentukan nilai awal untuk parameter $\mu^{(0)}$, $\sigma^{(0)}$, dan $\pi^{(0)}$
- 2) Tahap Ekspektasi dalam iterasi ke-0, yaitu menghitung nilai ekspektasi dari variabel acak tidak teramati (u_i) ketika $\delta = 0$.

$$\begin{aligned} y_i^{(0)} &= E\{u_i | \mu^{(0)}, \sigma^{(0)}, \pi^{(0)}\} \\ &= \delta_i + (1 - \delta_i) \frac{\pi^{(0)} S^{(0)}(t | U = 1)}{1 - \pi^{(0)} + \pi^{(0)} S^{(0)}(t | U = 1)} \quad \dots(2.60) \end{aligned}$$

- 3) Tahap Maksimisasi, yaitu memaksimalkan fungsi kemungkinan pada persamaan 2.58 dan 2.59 dengan mensubstitusikan nilai u_i dengan nilai ekspektasinya (y_i) menggunakan metode Newton-Raphson maka akan diperoleh nilai-nilai $\mu^{(1)}$, $\sigma^{(1)}$, dan $\pi^{(1)}$.

- 4) Secara umum pada iterasi ke- r , ganti $\mu^{(r)}$, $\sigma^{(r)}$, dan $\pi^{(r)}$ dengan $\mu^{(r+1)}$, $\sigma^{(r+1)}$, dan $\pi^{(r+1)}$ pada iterasi selanjutnya, kemudian kembali melakukan tahap 2 dan 3 secara berulang sampai diperoleh penaksir parameter yang konvergen $|\mu^{(r)} - \mu^{(r+1)}| \leq \varepsilon$, $|\sigma^{(r)} - \sigma^{(r+1)}| \leq \varepsilon$, dan $|\pi^{(r)} - \pi^{(r+1)}| \leq \varepsilon$ dengan $\varepsilon = 10^{-5}$

Pada tahap ke-3, maksimisasi dilakukan dengan menggunakan metode Newton-Raphson yang merupakan suatu metode untuk menyelesaikan sistem persamaan nonlinear (Agresti, 1990). Metode Newton-Raphson dapat dikembangkan dari perluasan deret Taylor, yang dinyatakan sebagai :

$$l(\mu) = l(\mu^{(r)}) + (\mu - \mu^{(r)})l'(\mu^{(r)}) + \frac{(\mu - \mu^{(r)})^2}{2!} l''(\mu^{(r)}) + \dots \text{ untuk } r = 0, 1, 2, \dots \quad \dots(2.61)$$

Jika μ terdiri dari $\mu_1, \mu_2, \dots, \mu_i$ dan $l(\mu_1, \mu_2, \dots, \mu_i)$ dapat ditulis sebagai berikut :

$$l(\mu) = T(\mu) = \begin{bmatrix} \frac{\partial l}{\partial \mu_1} \\ \frac{\partial l}{\partial \mu_2} \\ \vdots \\ \frac{\partial l}{\partial \mu^{(r)}} \end{bmatrix} \quad \quad \quad l''(\mu) = H(\mu) = \begin{bmatrix} \frac{\partial^2 l}{\partial \mu_1^2} & \frac{\partial^2 l}{\partial \mu_1 \partial \mu_2} & \dots & \frac{\partial^2 l}{\partial \mu_1 \partial \mu_i} \\ \frac{\partial^2 l}{\partial \mu_2 \partial \mu_1} & \frac{\partial^2 l}{\partial \mu_2^2} & \dots & \frac{\partial^2 l}{\partial \mu_2 \partial \mu_i} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l}{\partial \mu_i \partial \mu_1} & \frac{\partial^2 l}{\partial \mu_i \partial \mu_2} & \dots & \frac{\partial^2 l}{\partial \mu_i^2} \end{bmatrix}$$

Maka persamaan 2.61 dapat ditulis dengan :

$$l(\mu) = l(\mu^{(r)}) + (\mu - \mu^{(r)})T(\mu^{(r)}) + \frac{(\mu - \mu^{(r)})^2}{2!} H(\mu^{(r)}) \quad \dots(2.62)$$

Turunan dari persamaan 2.62 terhadap μ adalah :

$$T(\mu) = T(\mu^{(r)}) + (\mu - \mu^{(r)})H(\mu^{(r)})$$

Jika $T(\mu) = 0$, maka akan diperoleh :

$$T(\mu^{(r)}) + (\mu - \mu^{(r)})H(\mu^{(r)}) = 0$$

Pendekatan dari $\mu^{(r)}$ adalah $\mu^{(r+1)}$, maka persamaan di atas dapat ditulis :

$$T(\mu^{(r)}) + (\mu^{(r+1)} - \mu^{(r)})H(\mu^{(r)}) = 0 \quad \dots(2.63)$$

Dengan menyelesaikan persamaan 2.63, maka dapat diperoleh suatu iterasi sebagai berikut :

$$\mu^{(r+1)} = \mu^{(r)} - T(\mu^{(r)})H(\mu^{(r)})^{-1} \quad \dots(2.64)$$

Untuk parameter σ
$$\sigma^{(r+1)} = \sigma^{(r)} - T(\sigma^{(r)})H(\sigma^{(r)})^{-1} \quad \dots(2.64)$$

Dan untuk parameter π
$$\pi^{(r+1)} = \pi^{(r)} - T(\pi^{(r)})H(\pi^{(r)})^{-1} \quad \dots(2.64)$$

Dengan $\mu^{(r)}$, $\sigma^{(r)}$, dan $\pi^{(r)}$ menyatakan nilai μ , σ , dan π pada iterasi ke- r dan nilai taksiran awal dinyatakan dengan $\mu^{(0)}$, $\sigma^{(0)}$, dan $\pi^{(0)}$, proses iterasi akan berhenti ketika nilai telah konvergen.

2.13 Penaksiran Parameter Model Regresi Cure Log-Logistik

Parameter β , γ , dan σ pada model regresi Cure Log-logistik (pers 2.20) merupakan parameter yang belum diketahui dan akan ditaksir menggunakan metode kemungkinan maksimum. Fungsi kemungkinan adalah peluang bersama dari data pengamatan yang dianggap sebagai fungsi dari parameter yang tidak diketahui dalam asumsi model. Pendugaan parameter dengan metode kemungkinan maksimum adalah nilai ketika fungsi kemungkinan maksimum.

Berbeda dengan model Cure Log-logistik, pada model regresi terdapat koefisien regresi yang akan ditaksir, yaitu koefisien regresi γ yang menjelaskan pengaruh kovariat x terhadap waktu mencencil debitur, dan koefisien regresi β yang menjelaskan pengaruh kovariat z terhadap peluang dari peristiwa atau peristiwa yang diperhatikan (macet). Fungsi kemungkinan didefinisikan sebagai berikut :

$$L(\beta, \gamma, \sigma) = \prod_{i=1}^n \{(\pi_i(z_i))f(t_i | U=1, x_i)\}^{\delta_i} \{(1-\pi_i(z_i)) + \pi_i(z_i)S(t_i | U=1, x_i)\}^{1-\delta_i} \quad \dots(2.65)$$

Sama seperti pada penaksiran model Cure Log-logistik jika $\delta_i = 1$ maka pasti $u_i = 1$. Artinya jika waktu mencuil seorang debitur teramati lengkap sampai terjadi peristiwa macet, maka debitur tersebut pasti mengalami peristiwa macet. Tetapi jika $\delta_i = 0$ maka nilai u_i akan mempunyai dua kemungkinan, yaitu $u_i = 1$ atau $u_i = 0$. Artinya jika waktu mencuil seorang debitur tidak teramati lengkap (tersensor), maka debitur tersebut mungkin mengalami macet atau mungkin lancar, sehingga dapat dikatakan pengamatan u_i tidak teramati (*missing*).

Ketika u_i merupakan nilai dari variabel acak U_i , maka fungsi log-kemungkinan yang lengkap adalah gabungan dari dua komponen yang saling bebas, yaitu l_l yang hanya bergantung pada β , dan l_s yang hanya bergantung pada γ dan σ (Corbie & Joly, 2006) :

$$l_l(\beta; u) = \log \prod_{i=1}^n \pi(z_i)^{u_i} (1 - \pi(z_i))^{1-u_i} \quad \dots(2.66)$$

dan

$$l_s(\gamma, \sigma; u) = \log \prod_{i=1}^n h(t | U = 1, x_i)^{\delta_i u_i} S(t | U = 1, x_i)^{u_i} \quad \dots(2.67)$$

$S(t|U=1)$ merupakan fungsi peluang bersyarat dari fungsi survival terhadap T ketika sebagian populasi mengalami peristiwa yang diperhatikan (macet).

$$S(t, | U = 1) = \frac{1}{1 + e^{\frac{\log t - \gamma x}{\sigma}}}$$

Dan $h(t|U=1)$ merupakan fungsi peluang bersyarat dari fungsi hazard terhadap T ketika sebagian populasi mengalami peristiwa yang diperhatikan (macet).

$$h(t | U = 1) = \frac{\sigma^{-1} e^{\frac{\log t - \gamma x}{\sigma}}}{1 + e^{\frac{\log t - \gamma x}{\sigma}}}$$

Dan,

$$\pi_i(z_i) = \frac{1}{1 + e^{\beta z_i}}$$

Sama seperti penaksiran model Cure Log-Logistik, untuk memaksimumkan fungsi log-kemungkinan di atas digunakan algoritme EM (*Expectation Maximization*) yang merupakan salah satu metode optimasi dalam memaksimumkan fungsi kemungkinan yang mengandung data tidak terobservasi (*missing*). Tahapan algoritme EM adalah sebagai berikut :

- 1) Tentukan nilai awal untuk parameter $\beta^{(0)}$, $\gamma^{(0)}$, dan $\sigma^{(0)}$
- 2) Tahap Ekspektasi dalam iterasi ke-0, yaitu menghitung nilai ekspektasi dari variabel acak tidak teramati (u_i) ketika $\delta = 0$.

$$\begin{aligned} y_i^{(0)} &= E\{u_i | \beta^{(0)}, \gamma^{(0)}, \sigma^{(0)}\} \\ &= \delta_i + (1 - \delta_i) \frac{\pi^{(0)}(z_i) S^{(0)}(t | U = 1, x_i)}{1 - \pi^{(0)}(z_i) + \pi^{(0)}(z_i) S^{(0)}(t | U = 1, x_i)} \end{aligned} \quad \dots(2.68)$$

- 3) Tahap Maksimisasi, yaitu memaksimumkan fungsi kemungkinan pada persamaan 2.66 dan 2.67 dengan mensubstitusikan nilai u_i dengan nilai ekspektasinya (y_i), menggunakan metode Newton-Raphson secara umum pada iterasi ke- r diperoleh :

$$\beta^{(r+1)} = \beta^{(r)} - T(\beta^{(r)})H(\beta^{(r)})^{-1} \quad \dots(2.69)$$

$$\gamma^{(r+1)} = \gamma^{(r)} - T(\gamma^{(r)})H(\gamma^{(r)})^{-1} \quad \dots(2.70)$$

Dan

$$\sigma^{(r+1)} = \sigma^{(r)} - T(\sigma^{(r)})H(\sigma^{(r)})^{-1} \quad \dots(2.71)$$

- 4) Ganti $\beta^{(r)}$, $\gamma^{(r)}$, dan $\sigma^{(r)}$ dengan $\beta^{(r+1)}$, $\gamma^{(r+1)}$, dan $\sigma^{(r+1)}$ pada iterasi selanjutnya, kemudian kembali melakukan tahap 2 dan 3 secara berulang sampai diperoleh penaksir parameter yang konvergen $|\beta^{(r)} - \beta^{(r+1)}| \leq \varepsilon$, $|\gamma^{(r)} - \gamma^{(r+1)}| \leq \varepsilon$, dan $|\sigma^{(r)} - \sigma^{(r+1)}| \leq \varepsilon$ dengan $\varepsilon = 10^{-5}$.

2.14 Pengujian Signifikansi Parameter Secara Simultan

Pengujian signifikansi parameter secara simultan bertujuan untuk memeriksa kemaknaan koefisien secara serentak. Pengujian signifikansi parameter regresi Cure Log-logistik secara simultan adalah dengan menggunakan uji rasio kemungkinan (*Likelihood Ratio Test*) dengan hipotesis sebagai berikut :

$$H_0 : \gamma_1 = \gamma_2 = \dots = \gamma_{12} = \beta_1 = \beta_2 = \dots = \beta_{12} = 0$$

H_1 : Paling sedikit terdapat satu γ atau satu $\beta \neq 0$

Statistik uji :

$$G = -2 \ln \left[\frac{\text{Fungsi kemungkinan tan pa kovariat}}{\text{Fungsi kemungkinan dengan kovariat}} \right]$$

$$G = \left(2 \left(\log \prod_{i=1}^n \pi(z_i)^{n_i} (1 - \pi(z_i))^{1-n_i} \right) + 2 \left(\log \prod_{i=1}^n h(t | U = 1, x_i)^{\delta_i n_i} S(t | U = 1, x_i)^{n_i} \right) \right) - \left(2 \left(\log \prod_{i=1}^n \pi^n (1 - \pi)^{1-n} \right) + 2 \left(\log \prod_{i=1}^n h(t | U = 1)^{\delta_i n_i} S(t | U = 1)^{n_i} \right) \right)$$

Kriteria uji : H_0 ditolak jika statistik uji $G > \chi^2_{(\alpha, p)}$.

2.15 Pengujian Signifikansi Parameter

Pengujian signifikansi parameter model bertujuan untuk mengetahui apakah suatu kovariat mempengaruhi respon atau tidak. Pengujian signifikansi parameter model regresi Cure Log-logistik menggunakan uji *t* student dengan hipotesis sebagai berikut :

$$H_0 : \gamma_i = \beta_i = 0$$

$H_1 : \gamma_i \neq 0 ; \beta_i \neq 0$

Statistik uji untuk parameter γ :

$$t = \frac{\hat{\gamma}}{SE(\hat{\gamma})}$$

Statistik uji untuk parameter β :

$$t = \frac{\hat{\beta}}{SE(\hat{\beta})}$$

Kriteria uji : H_0 ditolak jika statistik uji $t > t_{(\alpha, n-1)}$.

2.16 Kredit Perbankan dan Kredit Bermasalah

Menurut Triandaru (2006), kredit merupakan pemberian fasilitas pinjaman kepada nasabah, baik berupa fasilitas pinjaman tunai (*cash loan*) maupun pinjaman nontunai (*non-cash loan*). Pemberian kredit, dalam pengertian sebagai *cash loan* merupakan salah satu bentuk usaha yang dilakukan oleh sebuah bank. Berdasarkan UU Nomor 10 tahun 1998 tentang perubahan atas UU Nomor 7 tahun 1992 tentang perbankan, yang dimaksud dengan kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan atau kesepakatan pinjam-meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam untuk melunasi utangnya setelah jangka waktu tertentu dengan pemberian bunga. Dalam memberikan kredit atau pembiayaan, bank umum wajib mempunyai keyakinan berdasarkan analisis yang mendalam atas itikad dan kemampuan serta kesanggupan nasabah debitur untuk melunasi utangnya atau mengembalikan pembiayaan yang dimaksud sesuai dengan perjanjian yang disepakati. Mengingat hal tersebut, adanya prinsip kehati-hatian dalam pengelolaan bank serta adanya risiko yang selalu melekat dalam penyaluran dana, maka sebelum kredit atau pembiayaan disalurkan bank selalu ingin mengetahui segala sesuatu tentang kemampuan dan kemauan nasabah debiturnya untuk mengembalikan dana yang telah diberikan oleh bank. Hal-hal yang selalu ingin diketahui bank sebelum menyalurkan dananya dalam bentuk kedit maupun pembiayaan antara lain :

1. Perijinan dan legalitas kegiatan atau usaha nasabah secara yuridis. Terhentinya kegiatan usaha nasabah akan menyebabkan hilang atau berkurangnya kemampuan nasabah untuk

mengembalikan dana yang telah diterima, sehingga kredit atau pembiayaan menjadi bermasalah.

2. Karakter nasabah, dapat diukur oleh beberapa indikator. Indikator tersebut antara lain adalah profesi. Penampilan, lingkungan sosial, pengalaman, dan tindakan atau perilaku masa lalu.
3. Pengalaman dan manajemen nasabah, pengalaman yang tidak sesuai dengan bidang kegiatan yang akan dijalankan akan mengurangi kinerja usaha nasabah.
4. Keuangan nasabah, sehat tidak sehatnya suatu kegiatan atau usaha nasabah sangat bergantung pada keadaan keuangan nasabah tersebut.
5. Kemampuan teknis nasabah, menyangkut faktor-faktor yang dapat mendukung kelancaran kegiatan atau usaha nasabah secara teknis.
6. Sosial, keberadaan kegiatan atau usaha yang dibiayai oleh bank sedikit banyak akan membawa dampak tertentu terhadap masyarakat. Pihak bank harus ekstra hati-hati apabila dampak yang ditimbulkan adalah sesuatu yang bertentangan dengan masyarakat karena dapat berdampak pula pada kelancaran pembayaran kredit debitur.

Dalam pelepasan kartu kredit selalu terkandung risiko yang akan ditanggung oleh bank. Risiko-risiko yang berkaitan dengan pelepasan kredit nasabah dapat dikelompokkan ke dalam 4 kelompok yaitu *Credit Risk*, *Liquidity Risk*, *Price Risk*, dan *Prepayment Risk* (Prasetya, 2006).

Manajemen risiko kredit dilakukan agar bank dapat memperkirakan besarnya kerugian yang akan dialami dan untuk melindungi modal bank serta memaksimalkan *risk return trade off*. Kerugian potensial meliputi kerugian yang dapat diperkirakan sebelumnya (*Expected Loss*) sehingga tidak mempengaruhi kesehatan bank karena telah tertutup oleh provisi yang dikenakan kepada debitur dan penyisihan penghapusan aktiva produktif dan kerugian yang tidak diperkirakan (*Unexpected Loss*) atau tingkat kerugian yang berada di atas rata-rata sehingga dapat langsung mengkonsumsi modal, yang mengakibatkan menurunkan kesehatan bank.

Dalam Prasetya (2006) kredit bermasalah dikelompokkan menjadi tiga yaitu kredit kurang lancar, kredit diragukan, dan kredit macet.

Tabel 2.3 Peringkat Kredit Berdasarkan Ketepatan Pembayaran

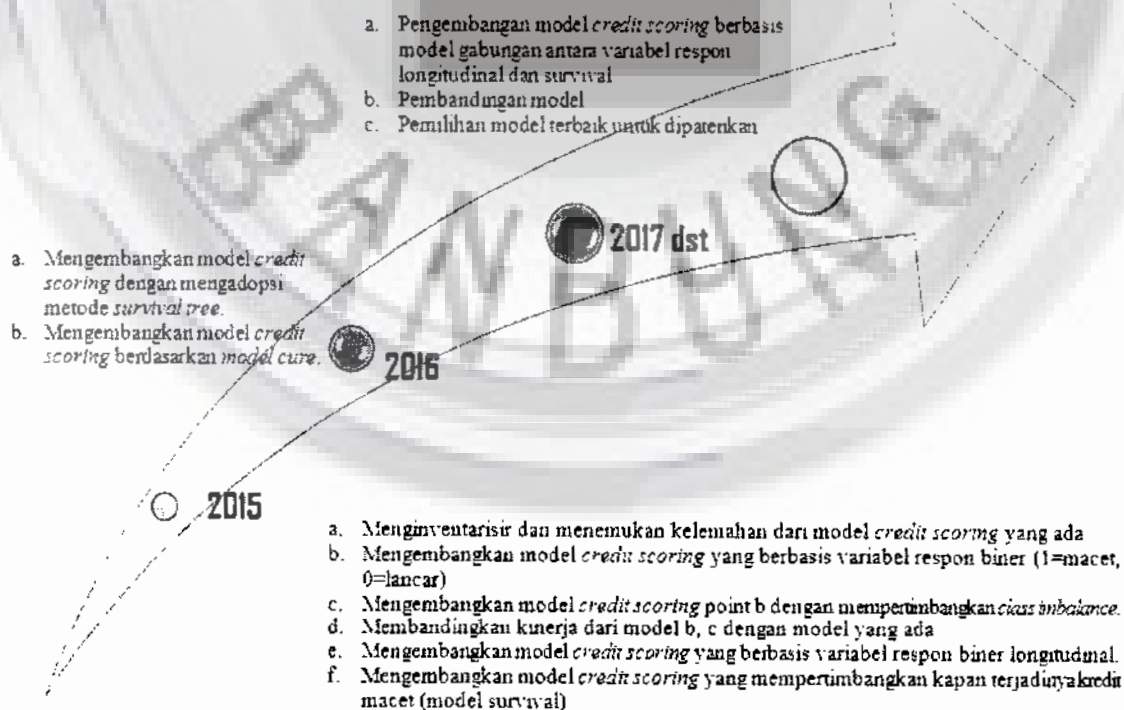
Kategori	Peringkat	Hari Tunggakan
Kurang Lancar (KL)	III	9-120 hari
Diragukan (D)	IV	121-180 hari
Macet (M)	V	>180 hari

Sumber : Rangkuman SK Direksi BI No.7/3/DPNP 31 Januari 2005

Secara singkat, penyebab dari kredit bermasalah adalah faktor internal dan faktor eksternal. Faktor eksternal meliputi kondisi luar yang berdampak pada penghasilan debitur misalnya kebijakan pemerintah yang menaikkan harga kebutuhan pokok sehingga berdampak terhadap penghasilan debitur. Sedangkan faktor internal disebabkan oleh kebijakan bank itu sendiri misalnya sistem prosedur bank dan jaminan yang tidak *marketable*.

2.17 Road map

Penelitian ini merupakan tahap awal untuk menemukan *tools* yang mempunyai kinerja sangat baik dalam mengidentifikasi calon debitur macet. Gambar 2.3 menunjukkan road map yang digunakan dalam penelitian ini.



Gambar 2.3 Roadmap penelitian

BAB 3. TUJUAN DAN MANFAAT PENELITIAN

3.1 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Untuk mengembangkan model statistika tentang masalah kredit macet dengan mengambil dasar pengembangan dari metode analisis regresi logistik dan analisis survival.
2. Untuk membandingkan model yang diusulkan dengan model yang ada.

3.2 Manfaat Penelitian

Penelitian ini diharapkan dapat membantu pihak perbankan untuk menghindarkan diri atas kerugian yang lebih besar. Untuk membantu mengantisipasi kerugian tersebut pihak manajemen dapat menggunakan *tools* statistika yang dibangun berdasarkan teori dan dicocokkan dengan data empiris yang ada. Oleh karena itu, akan sangat berguna jika sejak awal pihak bank sudah mempunyai *tools* yang dapat digunakan untuk memilah calon debitur yang baik dan calon debitur yang berpotensi macet.

BAB 4. METODE PENELITIAN

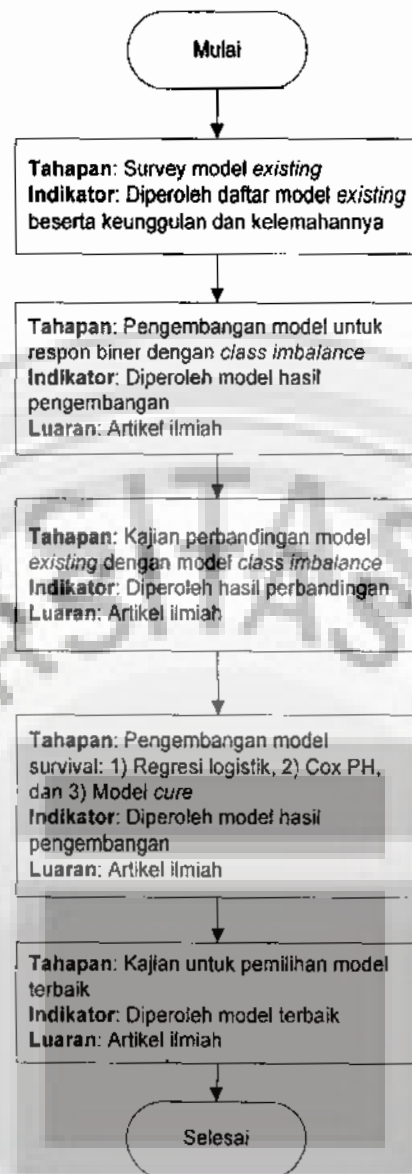
4.1 Alur Penelitian

Pengembangan model statistika tentang kemacetan kredit perbankan akan dilakukan dalam tahapan sebagai berikut:

1. Melakukan tinjauan pustaka dan survey tentang model *existing* yang digunakan di dunia perbankan.
2. Mengembangkan model *credit scoring* yang berbasis pada variabel respon biner dengan mempertimbangkan keseimbangan antara persentase yang macet dan lancar (*class imbalance*). Adanya ketidakseimbangan antara persentase yang macet dan lancar akan menyebabkan berkurangnya kinerja model-model yang berbasis pada variabel respon biner biasa (model regresi logistik, probit dll.)
3. Melakukan kajian perbandingan kinerja model yang dikembangkan dengan model *existing* dengan menggunakan data riil dan simulasi.
4. Mengembangkan model *credit scoring* yang mempertimbangkan kapan terjadinya kemacetan. Fenomena kredit yang dimodelkan dengan model ini adalah yang mempunyai jangka waktu tertentu untuk melunasi kreditnya. Sehingga selain masalah peluang terjadinya kemacetan, juga masalah kapan kemacetan itu terjadi. Model yang akan dikembangkan mengadopsi model-model yang ada dalam pustaka analisis survival, seperti model *Cox proportional hazards*, dan model *cure*.
5. Pemilihan model terbaik yang mampu memisahkan antara calon debitur yang lancar dengan yang berpotensi mengalami kemacetan.

4.2 Luaran dan Indikator Capaian

Tahapan penelitian akan dimonitor dan dievaluasi berdasarkan indikator-indikator pencapaiannya sebagaimana ditampilkan dalam Gambar 4.1.



Gambar 4.1. Flow chart beserta indikator dan luaran penelitian

BAB 5. HASIL DAN LUARAN YANG DICAPAI

5.1 Hasil Penelitian

Pada bab ini, akan ditampilkan hasil-hasil dari penelitian sejauh ini. Langkah pertama adalah membagi data menjadi data *training* dan data *testing*, pemodelan model regresi logistik, pemodelan *Cox proportional hazard*, membentuk tabel klasifikasi, menghitung *mean cost* dan dilanjutkan dengan menarik kesimpulan.

5.2 Data Training dan Data Testing

Tahapan pertama adalah membagi data menjadi dua kelompok yaitu data *training* sebesar 50% dan data *testing* sebesar 50%.

$$N_{\text{training}} = \frac{50}{100} \times 504 = 252$$

$$N_{\text{testing}} = \frac{50}{100} \times 504 = 252$$

Adapun pemilihan data *training* dilakukan dengan menggunakan *random sample* pada *software R i386 2.15.1*.

5.3 Model Regresi Logistik

5.3.1 Taksiran Parameter Model Regresi Logistik

Model regresi logistik yang memodelkan hubungan antara variabel tak bebas yang bersifat biner atau dikotomis dengan satu atau lebih variabel bebas dicocokkan terhadap data *training* menghasilkan taksiran model sebagaimana ditampilkan pada Tabel 5.1.

Tabel 5.1 Taksiran Parameter untuk Model Regresi Logistik

Variabel	Parameter	Taksiran	Galat Baku
	β_0	-7,888	2,526
Nilai nominal pinjaman	β_1	0,003	0,001
Debt Ratio	β_2	-0,077	0,078
Lama angsuran	β_3	-0,021	0,027
Loan to value	β_4	-0,014	0,016
Lama masa kerja	β_5	-0,014	0,044
Usia debitur	β_6	0,138	0,047
Jumlah tanggungan	β_7	0,663	0,608
Pekerjaan debitur	β_8	-1,039	0,782
Pendidikan terakhir (Universitas)	β_{91}	-0,467	0,673
Pendidikan terakhir (SMA/lebih rendah)	β_{92}	-0,168	0,696
Status tempat tinggal (Milik saudara/teman)	β_{101}	1,456	0,724
Status tempat tinggal (Kontrak)	β_{102}	1,468	0,682
Status pernikahan	β_{11}	0,389	0,583
Jenis kelamin	β_{12}	1,297	0,657

Berdasarkan hasil perhitungan yang tersaji pada Tabel 5.1, diperoleh nilai-nilai taksiran parameter untuk model regresi logistik. Model ini menunjukkan hubungan variabel bebas terhadap peluang terjadinya event, yaitu "macet". Nilai taksiran parameter β_0 sampai β_{12} untuk intersep, variabel nilai nominal pinjaman, *debt ratio*, lama angsuran, *loan to value*, lama masa kerja, usia debitur, jumlah tanggungan, pekerjaan debitur, pendidikan terakhir (universitas), pendidikan terakhir (SMA/lebih rendah), status tempat tinggal (milik saudara/teman), status tempat tinggal (kontrak), status pernikahan dan kovariat jenis kelamin berturut-turut adalah, -7,888 (dengan galat baku 2,526), 0,003 (dengan galat baku 0,001), -0,077 (dengan galat baku 0,078), -0,021 (dengan galat baku 0,027), -0,014 (dengan galat baku 0,016), -0,014 (dengan galat baku 0,044), 0,138 (dengan galat baku 0,047), 0,663 (dengan galat baku 0,608), -1,039 (dengan galat baku 0,782), -0,467 (dengan galat baku 0,673), -0,168 (dengan galat baku 0,696), 1,456 (dengan galat baku 0,724), 1,468 (dengan galat baku 0,682), 0,389 (dengan galat baku 0,583), 1,297 (dengan galat baku 0,657).

Dengan demikian, model taksiran logit peluang kredit macet adalah sebagai berikut:

$$\begin{aligned}\hat{g}(x_i) = & -7,888 + 0,003x_{i1} - 0,077x_{i2} - 0,021x_{i3} - 0,014x_{i4} - 0,014x_{i5} + 0,138x_{i6} \\ & + 0,663x_{i7} - 1,039x_{i8} - 0,467x_{i9} - 0,168x_{i10} + 1,456x_{i11} + 1,468x_{i12} \\ & + 0,389x_{i13} + 1,297x_{i14}\end{aligned}$$

5.3.2 Pengujian Signifikansi Parameter Model Secara Simultan

Pada bagian ini dilakukan pengujian signifikansi parameter secara simultan yang bertujuan untuk memeriksa kemaknaan variabel secara serentak. Hipotesis yang digunakan adalah :

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_{12} = 0$$

$$H_1 : \text{paling tidak terdapat satu } \beta \neq 0$$

Statistik Uji:

$$\begin{aligned}G &= -2 \ln L_0 - (-2 \ln L_p) \\ &= 167,329 - 119,250 \\ &= 48,079\end{aligned}$$

Dari hasil $G = 48,079$ diperoleh nilai $p\text{-value}$ sebesar $1,28 \times 10^{-5}$. Dengan taraf arti $\alpha = 5\%$ maka nilai $p\text{-value} < \alpha$ sehingga H_0 ditolak. Artinya, minimal ada satu variabel bebas berpengaruh terhadap peluang terjadinya macet debitur KPR Bank ABC, Tbk.

5.3.3 Pengujian Signifikansi Parameter Model Secara Parsial

Pengujian signifikansi parameter model regresi logistik menggunakan uji *Wald* untuk mengetahui variabel mana yang mempengaruhi peluang kredit macet debitur secara signifikan.

Tabel 5.2 Nilai Statistik Uji *Wald* dan *P-value* Model Regresi Logistik

Variabel	Parameter	Wald	P-value
	β_0	9,752	0,002*
Nilai nominal pinjaman	β_1	16,002	0,000*
Debt Ratio	β_2	0,969	0,325
Lama angsuran	β_3	0,625	0,429
Loan to value	β_4	0,730	0,393
Lama masa kerja	β_5	0,103	0,748
Usia debitur	β_6	8,718	0,003*
Jumlah tanggungan	β_7	1,190	0,275
Pekerjaan debitur	β_8	1,766	0,184
Pendidikan terakhir (Universitas)	β_{91}	0,482	0,488
Pendidikan terakhir (SMA/lebih rendah)	β_{92}	0,058	0,809
Status tempat tinggal (Milik sodara/teman)	β_{101}	4,042	0,044*
Status tempat tinggal (Kontrak)	β_{102}	4,627	0,031*
Status pernikahan	β_{11}	0,446	0,504
Jenis kelamin	β_{12}	3,892	0,049*

*: signifikan dengan taraf arti 5%

Pengujian hipotesis parameter model regresi logistik dilakukan dengan menggunakan hipotesis:

$H_0 : \beta_j = 0$; parameter β_j tidak signifikan

$H_1 : \beta_j \neq 0$; parameter β_j signifikan

Dengan taraf arti sebesar 5%, berdasarkan Tabel 5.2 diperoleh hasil variabel yang berpengaruh secara signifikan adalah variabel nilai nominal pinjaman, usia debitur, status tempat tinggal (Milik sodara/teman), status tempat tinggal (Kontrak) dan jenis kelamin. Nilai taksiran parameter untuk variabel nilai nominal pinjaman sebesar 0,003. Artinya jika nilai nominal pinjaman debitur bertambah sebesar 1 juta rupiah, maka peluang debitur akan mengalami macet membayar adalah 1,003 kali lipat lebih besar dibandingkan terhadap debitur yang meminjam 1 satuan lebih rendah. Dengan kata lain peningkatan nilai nominal pinjaman debitur akan meningkatkan peluang terjadinya macet. Nilai taksiran parameter untuk variabel usia debitur adalah 0,138. Artinya jika usia debitur bertambah sebesar 1 tahun, maka peluang debitur akan mengalami macet membayar adalah 1,148 kali lipat lebih besar dibandingkan

terhadap debitur yang usianya 1 tahun lebih rendah. Hal ini menunjukkan bahwa makin tua usia debitur maka makin besar peluang terjadinya macet. Nilai taksiran parameter untuk variabel status tempat tinggal debitur milik saudara/teman adalah 1,456. Artinya peluang kredit macet dari debitur yang status tempat tinggal milik saudara/teman adalah 4,291 kali lipat lebih besar dibandingkan dengan debitur yang status tempat tinggal kontrak atau lainnya. Nilai taksiran untuk parameter untuk variabel tempat tinggal debitur dengan status kontrak adalah 1,468. Artinya peluang kredit macet dari debitur yang tempat tinggal dengan status kontrak adalah 4,339 kali lipat lebih besar dibandingkan dengan debitur yang status tempat tinggal milik saudara/teman atau lainnya. Nilai taksiran parameter untuk variabel jenis kelamin adalah 1,297. Artinya peluang kredit macet jenis kelamin laki-laki 3,659 kali lipat lebih besar dibandingkan jenis kelamin perempuan.

5.3.4 Nilai Taksiran Peluang Kredit Macet ($\hat{\pi}$)

Setelah diperoleh taksiran model regresi logistik, maka selanjutnya dihitung taksiran peluang kredit macet ($\hat{\pi}$) untuk seluruh debitur yang termasuk dalam data *training* dan hasilnya ditampilkan dalam Tabel 5.3.

Tabel 5.3 Nilai Taksiran Peluang Kredit Macet ($\hat{\pi}$) untuk Seluruh Debitur yang Termasuk dalam Data *Training*

No	i	Peluang Kredit Macet ($\hat{\pi}$)
1	5	0.01103
2	7	0.03076
3	8	0.01886
⋮	⋮	⋮
252	504	0.02127

5.3.5 *Cut-off* optimal

Nilai taksiran peluang kredit macet ($\hat{\pi}$) untuk setiap debitur dalam data *training*, selanjutnya akan dikonversi menjadi taksiran variabel responnya (\hat{y}). Akan tetapi sebelumnya perlu dicari dulu berapakah nilai batas $\hat{\pi}$ atau *cut-off* yang optimal dengan menggunakan kurva

ROC. Prosedur pencarian nilai *cut-off* optimal melalui perhitungan sensitivitas dan spesifisitas terlebih dahulu. Hasilnya ditampilkan dalam Tabel 5.4.

Tabel 5.4 Proses Mencari *Cut-off* Optimal Model Regresi Logistik

No	Positive if Greater Than or Equal To	Sensitivity	1-Specificity	Sensitivity+(Spesificity-1)
1	0.0000000	1.000	1.000	0.0000000
2	0.0017306	1.000	0.996	0.0044248
3	0.0028850	1.000	0.991	0.0088496
⋮	⋮	⋮	⋮	⋮
253	1.0000000	0.000	0.000	0.0000000

Nilai *cut-off* optimal adalah nilai maksimum dari sensitivitas + (spesifisitas-1) yaitu sebesar 0,625. Hasil dari proses selengkapnya dilampirkan dalam lampiran 5.

5.3.6 Tabel Klasifikasi Pada Data *Training*

Setelah diperoleh nilai *cut-off* optimal melalui kurva ROC, maka ditaksir nilai \hat{y} yang hasilnya ditampilkan dalam Tabel 5.5 dengan cara sebagai berikut :

Jika peluang debitur mengalami macet $\geq 0,625$ maka $\hat{y} = 1$ dan

Jika peluang debitur mengalami macet $< 0,625$ maka $\hat{y} = 0$

Tabel 5.5 Taksiran Nilai \hat{y} Model Regresi Logistik untuk Seluruh Debitur dalam Data *Training*

No	x	y	\hat{y}
1	5	0	0
2	7	0	0
3	8	0	0
⋮	⋮	⋮	⋮
252	504	0	0

Adapun nilai taksiran \hat{y} untuk data *training* selengkapnya dilampirkan pada lampiran 9.

Kemudian dibuat tabel klasifikasi antara y dan \hat{y} untuk data *training* dan hasilnya

ditampilkan dalam Tabel 5.6.

Tabel 5.6 Tabel Klasifikasi Pada Data *Training*

Aktual		Prediksi	
		\hat{y}	
		0	1
y	0	224	2
	1	22	4

$$\begin{aligned} \text{Mean cost} &= \frac{f_{01} + 20f_{10}}{n} \\ &= \frac{2 + (20 \times 22)}{252} \\ &= 1,753968 \end{aligned}$$

Ongkos atau kerugian akibat salah memprediksi model regresi logistik pada data *training* adalah sebesar 1,753968.

5.3.7 Tabel Klasifikasi Pada Data *Testing*

Dengan menggunakan model dan *cut-off* optimal yang diperoleh dari data *training*, selanjutnya dihitung peluang kredit macet ($\hat{\pi}$) dan ditaksir nilai \hat{y} untuk seluruh debitur yang termasuk dalam data *testing* dan hasilnya ditampilkan dalam Tabel 5.7 dan Tabel 5.8.

Tabel 5.7 Nilai Taksiran Peluang Kredit Macet ($\hat{\pi}$) untuk Seluruh Debitur yang Termasuk dalam Data *Testing*

No.	i	Peluang Kredit Macet ($\hat{\pi}$)
1	1	0.15515
2	2	0.05474
3	3	0.07595
⋮	⋮	⋮
252	500	0.12202

Tabel 5.8 Taksiran Nilai \hat{y} Model Regresi Logistik untuk Seluruh Debitur dalam Data *Testing*

No	x_1	x_2	\hat{y}
1	1	0	1
2	2	0	1
3	3	0	1
⋮	⋮	⋮	⋮
252	500	0	1

Kemudian dibuat Tabel klasifikasi antara y dan \hat{y} untuk data *testing* dan hasilnya ditampilkan dalam Tabel 5.9.

Tabel 5.9 Tabel Klasifikasi Pada Data *Testing*

Aktual		Prediksi	
		\hat{y}	
		0	1
y	0	217	5
	1	29	1

$$\text{Mean cost} = \frac{f_{01} + 20f_{10}}{n}$$

$$= \frac{5 + (20 \times 29)}{252}$$

$$= 2,321429$$

Ongkos atau kerugian akibat salah memprediksi model regresi logistik pada data *testing* adalah sebesar 2,321429.

5.4 Model Cox *proportional hazard*

5.4.1 Taksiran Parameter Model Cox *proportional hazard*

Model Cox *proportional hazard* yang memodelkan data *survival* dengan variabel bebas yang mempengaruhi fungsi *hazard* dicocokkan terhadap data *training* menghasilkan taksiran model sebagaimana ditampilkan pada Tabel 5.10.

Tabel 5.10 Taksiran Parameter untuk Model *Cox proportional hazard*

Variabel	Parameter	Taksiran	Galat Baku
Niai nominal pinjaman	β_1	0,002	0,000
Debt Ratio	β_2	-0,018	0,077
Lama angsuran	β_3	-0,025	0,025
Loan to value	β_4	0,004	0,013
Lama masa kerja	β_5	-0,036	0,038
Usia debitur	β_6	0,157	0,042
Jumlah tanggungan	β_7	0,537	0,511
Pekerjaan debitur	β_8	-0,089	0,682
Pendidikan terakhir (Universitas)	β_{91}	-0,253	0,599
Pendidikan terakhir (SMA/lebih rendah)	β_{92}	-0,008	0,578
Status tempat tinggal (Milik saudara/teman)	β_{101}	0,839	0,650
Status tempat tinggal (Kontrak)	β_{102}	0,351	0,595
Status pernikahan	β_{11}	0,215	0,574
Jenis kelamin	β_{12}	0,758	0,594

Berdasarkan hasil perhitungan yang tersaji pada Tabel 5.10, diperoleh nilai-nilai taksiran parameter untuk model *Cox proportional hazard*. Model ini menunjukkan hubungan variabel bebas terhadap fungsi hazard kemacetan. Nilai taksiran parameter β_1 sampai β_{12} untuk variabel nilai nominal pinjaman, *debt ratio*, lama angsuran, *loan to value*, lama masa kerja, usia debitur, jumlah tanggungan, pekerjaan debitur, pendidikan terakhir (universitas), pendidikan terakhir (SMA/lebih rendah), status tempat tinggal (milik saudara/teman), status tempat tinggal (kontrak), status pernikahan dan kovariat jenis kelamin berturut-turut adalah 0,002 (dengan galat baku 1,002), -0,018 (dengan galat baku 0,982), -0,025 (dengan galat baku 0,975), 0,004 (dengan galat baku 1,004), -0,036 (dengan galat baku 0,964), 0,157 (dengan galat baku 1,170), 0,537 (dengan galat baku 1,711), -0,089 (dengan galat baku 0,915), -0,253 (dengan galat baku 0,776), -0,008 (dengan galat baku 0,992), 0,839 (dengan galat baku 2,315), 0,351 (dengan galat baku 1,420), 0,215 (dengan galat baku 1,239), 0,758 (dengan galat baku 2,133).

Dengan demikian *model Cox proportional hazard* yang dinyatakan dalam bentuk persamaan log adalah sebagai berikut:

$$\begin{aligned} \log \left\{ \frac{h_i(t)}{h_0(t)} \right\} &= 0,002x_{i1} - 0,018x_{i2} - 0,025x_{i3} + 0,004x_{i4} - 0,036x_{i5} + 0,157x_{i6} \\ &+ 0,537x_{i7} - 0,089x_{i8} - 0,253x_{i91} - 0,008x_{i92} + 0,839x_{i101} + 0,351x_{i102} \\ &+ 0,215x_{i11} + 0,758x_{i12} \end{aligned}$$

5.4.2 Pengujian Signifikansi Parameter Model Secara Simultan

Pada bagian ini dilakukan pengujian signifikansi parameter secara simultan yang bertujuan untuk memeriksa kemaknaan variabel secara serentak. Hipotesis yang digunakan adalah :

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_{12} = 0$$

$$H_1 : \text{paling tidak terdapat satu } \beta \neq 0$$

Statistik Uji:

$$G = -2 \ln L_0 - (-2 \ln L_p)$$

$$= 232,776 - 187,122$$

$$= 45,654$$

Dari hasil $G = 45,654$ diperoleh nilai *p-value* sebesar $3,19 \times 10^{-5}$. Dengan taraf arti $\alpha = 5\%$ maka nilai *p-value* $< \alpha$ sehingga H_0 ditolak. Artinya, minimal ada satu variabel bebas berpengaruh terhadap peluang terjadinya macet debitur bank.

5.4.3 Pengujian Signifikansi Parameter Model Secara Parsial

Pengujian signifikansi parameter model regresi logistik menggunakan uji *Wald* untuk mengetahui variabel mana yang mempengaruhi fungsi hazard kemacetan debitur secara signifikan.

Tabel 5.11 Nilai Statistik Uji *Wald* dan *P-value* Model *Cox proportional hazard*

Parameter	Parameter	Wald	P-value
Niai nominal pinjaman	β_1	17,535	0,000*
Debt Ratio	β_2	0,055	0,814
Lama angsuran	β_3	0,989	0,320
Loan to value	β_4	0,077	0,781
Lama masa kerja	β_5	0,903	0,342
Usia debitur	β_6	13,728	0,000*
Jumlah tanggungan	β_7	1,103	0,294
Pekerjaan debitur	β_8	0,017	0,896
Pendidikan terakhir (Universitas)	β_{91}	0,179	0,672
Pendidikan terakhir (SMA/lebih rendah)	β_{92}	0,000	0,989
Status tempat tinggal (Milik saudara/teman)	β_{101}	1,666	0,197
Status tempat tinggal (Kontrak)	β_{102}	0,347	0,556
Status pernikahan	β_{11}	0,140	0,709
Jenis kelamin	β_{12}	1,626	0,202

*: signifikan dengan taraf arti 5%

Pengujian hipotesis parameter model *Cox proportional hazard* dilakukan dengan menggunakan hipotesis:

$H_0 : \beta_j = 0$; parameter β_j tidak signifikan

$H_1 : \beta_j \neq 0$; parameter β_j signifikan

Dengan taraf arti sebesar 5%, berdasarkan Tabel 5.11 diperoleh hasil variabel yang berpengaruh secara signifikan adalah variabel nilai nominal pinjaman dan usia debitur. Nilai taksiran parameter untuk variabel nilai nominal pinjaman sebesar 0,002. Artinya jika nilai nominal pinjaman debitur bertambah sebesar 1 juta rupiah, maka peluang debitur akan mengalami macet membayar adalah 1,002 kali lipat lebih besar dibandingkan terhadap debitur yang meminjam 1 satuan lebih rendah. Hal ini menunjukkan bahwa peningkatan nilai nominal pinjaman debitur akan meningkatkan peluang terjadinya macet. Nilai taksiran parameter untuk variabel usia debitur adalah 0,157. Artinya jika usia debitur bertambah sebesar 1 tahun, maka peluang debitur akan mengalami macet membayar adalah 1,170 kali lipat lebih besar

dibandingkan terhadap debitur yang usianya 1 tahun lebih rendah. Hal ini menunjukkan bahwa makin tua usia debitur maka makin besar peluang terjadinya macet.

5.4.4 Nilai Taksiran Peluang Kredit Macet ($1 - \hat{S}(12)$)

Setelah diperoleh taksiran model *Cox proportional hazard*, maka selanjutnya dihitung taksiran peluang kredit macet pada bulan ke-12, $1 - \hat{S}(12)$, untuk seluruh debitur yang termasuk dalam data *training*. Karena SPSS 21 mengeluarkan output $\hat{S}(12|x = \bar{x}) = 0,992$, sementara kita memerlukan $\hat{S}(12|x = 0) = \hat{S}_0(12)$. Maka berdasarkan model *Cox proportional hazard* yang dinyatakan dalam fungsi survival:

$$\hat{S}(t; x_i) = [\hat{S}_0(t)]^{\exp(x_i^T \beta)}$$

$$\hat{S}_0(t) = \hat{S}(t; x_i)^{\exp(-x_i^T \beta)}$$

$$\hat{S}_0(12) = \hat{S}(12; x = \bar{x})^{\exp(-\bar{x}^T \beta)}$$

Diketahui :

$$\bar{x}_1 = 231,692$$

$$\bar{x}_2 = 9,062$$

$$\bar{x}_3 = 25,014$$

$$\bar{x}_4 = 68,286$$

$$\bar{x}_5 = 7,830$$

$$\bar{x}_6 = 37,877$$

$$\bar{x}_7 = 0,705$$

$$\bar{x}_8 = 0,881$$

$$\bar{x}_{91} = 0,463$$

$$\bar{x}_{92} = 0,352$$

$$\bar{x}_{101} = 0,300$$

$$\bar{x}_{102} = 0,282$$

$$\bar{x}_{11} = 0,678$$

$$\bar{x}_{12} = 0,674$$

$$\text{Sehingga, } \hat{S}_0(12) = (0,992)^{\exp(-\bar{x}_j\beta)} = 0.999991058$$

selanjutnya dicari $\hat{S}(12; x = x_i)$; $i = 1, 2, \dots, n$ anggota data *training* dan hasilnya ditampilkan dalam tabel 5.12.

Tabel 5.12 Nilai Taksiran Peluang Kredit Macet ($1 - \hat{S}(12)$) untuk Seluruh Debitur yang Termasuk dalam Data *Training*

No.			
1	5	0.99447	0.00553
2	7	0.99277	0.00723
3	8	0.99831	0.00169
⋮	⋮	⋮	⋮
252	504	0.9941	0.00590

5.4.5 *Cut-off optimal*

Nilai taksiran peluang kredit macet ($1 - \hat{S}(12)$) untuk setiap debitur dalam data *training*, selanjutnya akan dikonversi menjadi taksiran variabel responnya ($\hat{\delta}$). Akan tetapi sebelumnya perlu dicari dulu berapakah nilai batas peluang kredit macet atau *cut-off* yang optimal dengan menggunakan kurva ROC. Prosedur pencarian nilai *cut-off* optimal melalui perhitungan sensitivitas dan spesifisitas terlebih dahulu. Hasilnya ditampilkan dalam tabel 5.13.

Tabel 5.13 Proses Mencari *Cut-off* Optimal Model *Cox PH*

No	Positive if Greater Than or Equal To ^a	Sensitivity	1-Specificity	Sensitivity+(Spesificity-1)
1	0.00000000	1.000	1.000	0.00000000
2	0.00025081	1.000	0.996	0.00442478
3	0.00064515	1.000	0.991	0.00884956
⋮	⋮	⋮	⋮	⋮
253	1.00000000	0.000	0.000	0.00000000

Nilai *cut-off* optimal adalah nilai maksimum dari sensitivitas + (spesifisitas - 1) yaitu sebesar 0,55.

5.4.6 Tabel Klasifikasi Pada Data *Training*

Setelah diperoleh nilai *cut-off* optimal melalui kurva ROC, maka ditaksir nilai $\hat{\delta}$ yang hasilnya ditampilkan dalam tabel 4.14 dengan cara sebagai berikut :

Jika peluang debitur mengalami macet $\geq 0,55$ maka $\hat{\delta} = 1$ dan

Jika peluang debitur mengalami macet $< 0,55$ maka $\hat{\delta} = 0$

Tabel 5.14 Taksiran Nilai $\hat{\delta}$ Model Cox PH untuk Seluruh Debitur dalam Data *Training*

No	δ	$\hat{\delta}$
1	5	1
2	7	1
3	8	1
⋮	⋮	⋮
252	504	1

Kemudian dibuat tabel klasifikasi antara δ dan $\hat{\delta}$ untuk data *training* dan hasilnya ditampilkan dalam Tabel 5.15.

Tabel 5.15 Tabel Klasifikasi Pada Data *Training*

Aktual		Prediksi	
		$\hat{\delta}$	
δ	0	226	0
	1	25	1

$$\begin{aligned} \text{Mean cost} &= \frac{f_{01} + 20f_{10}}{n} \\ &= \frac{(20 \times 25)}{252} \\ &= 1,984127 \end{aligned}$$

Ongkos atau kerugian akibat salah memprediksi model *Cox proportional hazard* pada data *training* adalah sebesar 1,984127.

5.4.7 Tabel Klasifikasi Pada Data *Testing*

Dengan menggunakan model dan *cut-off* optimal yang diperoleh dari data *training*, selanjutnya dihitung nilai peluang kredit macet ($1 - \hat{S}(12)$) dan ditaksir nilai $\hat{\delta}$ untuk seluruh debitur yang termasuk dalam data *testing* dan hasilnya ditampilkan dalam Tabel 5.16 dan Tabel 5.17.

Tabel 5.16 Nilai Taksiran Peluang Kredit Macet ($1 - \hat{S}(12)$) untuk Seluruh Debitur yang Termasuk dalam Data *Testing*

No	i	Peluang Kredit Macet ($1 - \hat{S}(12)$)	Taksiran Peluang Kredit Macet
1	1	0.99008	0.00992
2	2	0.98603	0.01397
3	3	0.97976	0.02024
⋮	⋮	⋮	⋮
252	500	0.99392	0.00608

Tabel 5.17 Taksiran Nilai $\hat{\delta}$ Model Cox PH untuk Seluruh Debitur dalam Data *Testing*

No	i	δ	$\hat{\delta}$
1	1	0	1
2	2	0	1
3	3	0	1
⋮	⋮	⋮	⋮
252	500	0	1

Kemudian dibuat tabel klasifikasi antara δ dan $\hat{\delta}$ untuk data *testing* dan hasilnya ditampilkan dalam Tabel 5.18.

Tabel 5.18 Tabel Klasifikasi Pada Data *Testing*

Aktual		Prediksi	
		$\hat{\delta}$	
		0	1
δ	0	222	0
	1	30	0

$$\text{Mean cost} = \frac{f_{01} + 20f_{10}}{n}$$

$$= \frac{(20 \times 30)}{252}$$

$$= 2,380952$$

Ongkos atau kerugian akibat salah memprediksi model *Cox proportional hazard* pada data *testing* adalah sebesar 2,380952.

5.5 Perbandingan Nilai *Mean Cost* Model Regresi Logistik dan *Cox Proportional Hazard*

Tabel 5.19 Nilai *Mean Cost* Model Regresi Logistik dan *Cox Proportional Hazard*

	Regresi Logistik	<i>Cox Proportional Hazard</i>
<i>Training</i>	1,753968	1,984127
<i>Testing</i>	2,321429	2,380952

Nilai *mean cost* atau kerugian akibat salah memprediksi model regresi logistik pada data *training* adalah sebesar 1,753968 sedangkan pada data *testing* adalah sebesar 2,321429. Dan nilai *mean cost* atau kerugian akibat salah memprediksi model *Cox proportional hazard* pada data *training* adalah sebesar 1,984127 sedangkan pada data *testing* adalah sebesar 2,380952.

5.6 Model Cure Log-logistik

Model Cure Log-logistik memodelkan proporsi individu yang tidak mengalami peristiwa yang diperhatikan (tidak macet). Dengan menggunakan *software* SAS, diperoleh model taksiran Cure Log-logistik sebagai berikut :

$$\hat{S}(t) = \left(\frac{1}{1 + e^{0,4312}} \right) \left(\frac{1}{1 + e^{\frac{\log t - 3,8227}{0,3641}}} \right) + \left(1 - \frac{1}{1 + e^{0,4312}} \right)$$

$$\hat{S}(t) = \left(\frac{1}{1 + e^{0,4312}} \right) \left(\frac{1}{1 + e^{2,7465 \log t - 10,4990}} \right) + \left(1 - \frac{1}{1 + e^{0,4312}} \right) \quad \dots(5.1)$$

Dimana
$$\hat{\pi}(z) = \frac{1}{1 + e^{\beta z}}$$

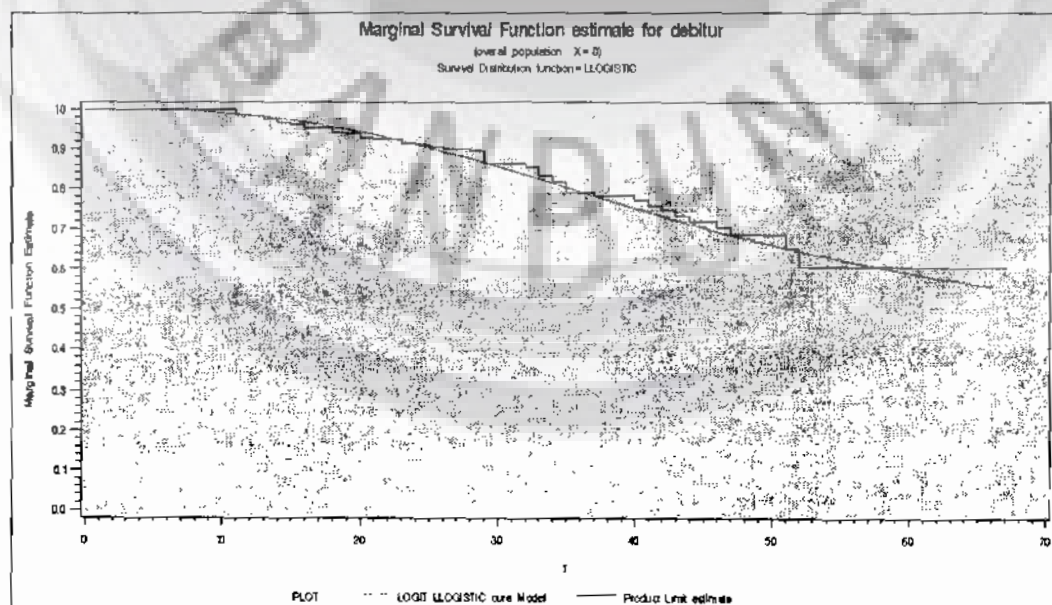
Maka, tanpa dipengaruhi oleh berbagai kovariat, proporsi debitur KPR Bank ABC, Tbk yang dinyatakan macet atau tidak berhasil melunasi cicilannya dapat diperoleh :

$$\hat{\pi}(z) = \frac{1}{1 + e^{0,4312}} = 0,39$$

Dengan demikian, dapat disimpulkan taksiran fraksi *uncure* atau proporsi debitur KPR Bank ABC, Tbk yang mengalami macet tanpa dipengaruhi oleh berbagai kovariat adalah sebesar 39%, dan debitur yang tidak mengalami macet (*cure*) adalah sebesar $1 - \hat{\pi}(z)$ atau 61%.

5.7 Analisis Ketahanan Mencicil Debitur Tanpa Kovariat

Pola ketahanan mencicil debitur KPR bank ABC, Tbk diperoleh dari hasil penaksiran fungsi ketahanan dalam bentuk kurva fungsi ketahanan marginal untuk keseluruhan sampel data. Menggunakan *software* SAS, kurva taksiran fungsi ketahanan marginal adalah sebagai berikut :



Gambar 5.1 Kurva Fungsi Ketahanan Marginal

Pada Gambar 5.1 terlihat dua buah pola yang terdiri dari pola fungsi survival dan pola Kaplan Meier (*product limit*) yang termasuk metode nonparametrik. Yang menjadi fokus dalam kurva ketahanan mencicil debitur KPR adalah pola fungsi survival, dimana sumbu tegak pada pola fungsi survival merupakan taksiran dari fungsi survival Cure Log-logistik dan sumbu datar merupakan waktu mencicil debitur KPR.

Dari Gambar 5.1 juga dapat dilihat bahwa mulai pada bulan ke-11 mulai terjadi penurunan fungsi ketahanan. Sehingga dapat dikatakan bahwa bulan ke-11 ketahanan pembayaran pada debitur peminjam KPR mulai mengalami macet meskipun peluang ketahanan dalam mencicil masih cukup besar. Penurunan fungsi ketahanan yang paling besar terjadi sekitar bulan ke-29, bulan ke-35, hingga bulan ke-52. Sehingga diindikasikan debitur KPR mengalami macet pembayaran paling besar sekitar bulan ke -29, bulan ke-35, hingga bulan ke-52.

5.8 Model Regresi Cure Log-logistik

Model regresi Cure Log-logistik memodelkan hubungan fungsional antar kovariat dengan waktu bertahan hidup dimana terdapat sekelompok individu yang tidak mengalami peristiwa yang diperhatikan berdasarkan model distribusi peluang Log-logistik, Dengan menggunakan *software* SAS, diperoleh output taksiran untuk parameter model regresi Cure Log-logistik sebagai berikut :

Pada Gambar 5.1 terlihat dua buah pola yang terdiri dari pola fungsi survival dan pola Kaplan Meier (*product limit*) yang termasuk metode nonparametrik. Yang menjadi fokus dalam kurva ketahanan mencicil debitur KPR adalah pola fungsi survival, dimana sumbu tegak pada pola fungsi survival merupakan taksiran dari fungsi survival Cure Log-logistik dan sumbu datar merupakan waktu mencicil debitur KPR.

Dari Gambar 4.1 juga dapat dilihat bahwa mulai pada bulan ke-11 mulai terjadi penurunan fungsi ketahanan. Sehingga dapat dikatakan bahwa bulan ke-11 ketahanan pembayaran pada debitur peminjam KPR mulai mengalami macet meskipun peluang ketahanan dalam mencicil masih cukup besar. Penurunan fungsi ketahanan yang paling besar terjadi sekitar bulan ke-29, bulan ke-35, hingga bulan ke-52. Sehingga diindikasikan debitur KPR mengalami macet pembayaran paling besar sekitar bulan ke -29, bulan ke-35, hingga bulan ke-52.

5.8 Model Regresi *Cure Log-logistik*

Model regresi *Cure Log-logistik* memodelkan hubungan fungsional antar kovariat dengan waktu bertahan hidup dimana terdapat sekelompok individu yang tidak mengalami peristiwa yang diperhatikan berdasarkan model distribusi peluang *Log-logistik*. Dengan menggunakan *software SAS*, diperoleh output taksiran untuk parameter model regresi *Cure Log-logistik* sebagai berikut :

Tabel 5.20 Taksiran Parameter untuk Model Regresi Cure Log-logistik

Parameter	Taksiran	Galat Baku
Model Logistik		
Intersep	-5,7231	5,8656
Nilai Nominal	0,0159	0,004347
<i>Debt Ratio</i>	-0,1149	0,1305
Lama Angsuran	0,02319	0,06406
<i>Loan to Value</i>	-0,02806	0,03064
Lama Masa Kerja	0,07528	0,09116
Usia	0,05577	0,09992
Jumlah Tanggungan	0,716	0,9115
Pekerjaan	-1,0823	1,093
Pendidikan Terakhir (Universitas)	2,4248	1,296
Pendidikan Terakhir (SMA/lebih rendah)	1,1993	1,1397
Status Tempat Tinggal (Milik saudara/teman)	-3,2529	1,6064
Status Tempat Tinggal (Kontrak)	-1,3957	1,3063
Status Pernikahan	1,8374	1,1714
Jenis Kelamin	1,2507	1,3172
Model Survival		
Intersep	5,3078	0,7953
Nilai Nominal	0,000941	0,000587
<i>Debt Ratio</i>	0,1101	0,06854
Lama Angsuran	-0,02168	0,03971
<i>Loan to Value</i>	0,03436	0,01473
Lama Masa Kerja	-0,00444	0,04691
Usia	0,08868	0,06143
Jumlah Tanggungan	-0,1217	0,4195
Pekerjaan	-0,387	0,8256
Pendidikan Terakhir (Universitas)	-0,2969	0,8812
Pendidikan Terakhir (SMA/lebih rendah)	-0,02989	0,6194
Status Tempat Tinggal (Milik saudara/teman)	1,3794	0,7131
Status Tempat Tinggal (Kontrak)	-0,03426	0,8425
Status Pernikahan	-0,158	0,6703
Jenis Kelamin	0,9349	0,6812
Shape	0,2535	0,02788

Berdasarkan hasil perhitungan yang tersaji pada tabel 5.20, diperoleh nilai-nilai taksiran parameter untuk model regresi Cure Log-logistik. Model logistik menunjukkan parameter-parameter untuk regresi logistik yaitu pengaruh dari z_q dalam peluang $\pi(z)$ atau peluang terjadinya peristiwa macet (*uncure*). Nilai taksiran parameter β_0 sampai β_{12} untuk intersep, kovariat nilai nominal, *debt ratio*, lama angsuran, *loan to value*, lama masa kerja,

usia, jumlah tanggungan, pekerjaan, pendidikan terakhir (SMA/lebih rendah), pendidikan terakhir (universitas), status tempat tinggal (kontrak), status tempat tinggal (milik saudara/teman), status pernikahan dan kovariat jenis kelamin berturut-turut adalah, -5,7231 (dengan galat baku 5,8656), 0,0159 (dengan galat baku 0,004347), -0,1149 (dengan galat baku 0,1305), 0,02319 (dengan galat baku 0,06406), -0,02806 (dengan galat baku 0,03064), 0,07528 (dengan galat baku 0,09116), 0,05577 (dengan galat baku 0,09992), 0,716 (dengan galat baku 0,9115), -1,0823 (dengan galat baku 1,093), 2,4248 (dengan galat baku 1,296), 1,1993 (dengan galat baku 1,1397), -3,2529 (dengan galat baku 1,6064), -1,3957 (dengan galat baku 1,3063), 1,8374 (dengan galat baku 1,1714), 1,2507 (dengan galat baku 1,3172) dan untuk parameter σ nilai taksirannya adalah 0,2535 (dengan galat baku 0,02788).

Dengan demikian model regresi ketika pengaruh dari z_i dalam peluang $\pi(z)$ atau peluang terjadinya peristiwa macet (*uncure*) adalah sebagai berikut :

$$\begin{aligned} \log it(\hat{\pi}(z)) = & -5,7231 + 0,0159z_1 - 0,1149z_2 + 0,02319z_3 - 0,02806z_4 + 0,07528z_5 + \\ & 0,05577z_6 + 0,716z_7 - 1,0823z_8 + 2,4248z_{9I} + 1,1993z_{9II} - 3,2529z_{10I} - \\ & 1,3957z_{10II} + 1,8374z_{11} + 1,2507z_{12} \end{aligned} \quad \dots(5.2)$$

Sedangkan model survival menunjukkan pengaruh kovariat x_p terhadap waktu mencuil debitur. Nilai-nilai taksiran parameter model regresi Cure Log-logistik untuk parameter γ_0 sampai γ_{12} berturut-turut adalah 5,3078 (dengan galat baku 0,7953), 0,000941 (dengan galat baku 0,000587), 0,1101 (dengan galat baku 0,06854), -0,02168 (dengan galat baku 0,03971), 0,03436 (dengan galat baku 0,01473), -0,00444 (dengan galat baku 0,04691), 0,08868 (dengan galat baku 0,06143), -0,1217 (dengan galat baku 0,4195), -0,387 (dengan galat baku 0,8256), -0,2969 (dengan galat baku 0,8812), -0,02989 (dengan galat baku 0,6194), 1,3794 (dengan galat baku 0,7131), -0,03426 (dengan galat baku 0,8425), -0,158 (dengan galat baku 0,6703), dan 0,9349 (dengan galat baku 0,6812).

Model regresi Cure Log-logistik, dimana kovariat x_i mempengaruhi parameter waktu mencuil μ adalah sebagai berikut :

$$\begin{aligned} \hat{\mu} = & 5,3078 + 0,000941x_1 + 0,1101x_2 - 0,2168x_3 + 0,03436x_4 - 0,00444x_5 + \\ & 0,08868x_6 - 0,1217x_8 - 0,387x_8 - 0,2969x_{9,l} - 0,02989x_{9,II} + 1,3794x_{10,l} - \\ & 0,03426_{10,II} - 0,158x_{11} + 0,9349z_{12} \end{aligned} \quad \dots(5.3)$$

Maka, model regresi Cure Log-logistik dimana kovariat x dapat mempengaruhi μ dan z dapat mempengaruhi $\pi(z)$ adalah sebagai berikut :

$$\hat{S}(t | x, z) = \hat{\pi} \left(\frac{1}{1 + e^{\frac{\log t - \hat{\mu}}{0,2535}}} \right) + (1 - \hat{\pi})$$

Dimana $\hat{\pi}$ sebagaimana pada persamaan 4.2 dan $\hat{\mu}$ sebagaimana pada persamaan 4.3.

5.9 Pengujian Signifikansi Parameter Model Secara Simultan

Pada bagian ini dilakukan pengujian signifikansi parameter secara simultan yang bertujuan untuk memeriksa kemaknaan koefisien secara serentak. Hipotesis yang digunakan adalah :

$$H_0 : \gamma_1 = \gamma_2 = \dots = \gamma_{12} = \beta_1 = \beta_2 = \dots = \beta_{12} = 0$$

H_1 : Paling sedikit terdapat satu γ atau satu $\beta \neq 0$

Dengan menggunakan *software* SAS diperoleh statistik uji rasio kemungkinan adalah 135,3. Dengan taraf arti sebesar 5%, nilai *p-value* adalah sebesar 0,000. Karena nilai $0,000 < 0,05$ maka dapat disimpulkan untuk menolak H_0 . Artinya, kovariat nilai nominal, *debt ratio*, lama angsuran, *loan to value*, lama masa kerja, usia, jumlah tanggungan, pekerjaan, pendidikan terakhir, status tempat tinggal, status pernikahan, dan jenis kelamin debitur secara simultan berpengaruh terhadap peluang terjadinya macet dan waktu mencicil debitur KPR Bank ABC, Tbk.

5.10 Pengujian Signifikansi Parameter Model Secara Parsial

Pengujian signifikansi parameter model regresi Cure Log-logistik menggunakan uji t untuk mengetahui kovariat mana yang mempengaruhi waktu mencicil debitur secara signifikan.

Tabel 5.21 Nilai Statistik Uji dan P -Value Model Regresi Cure Log-logistik

Parameter	t Value	Pr > t
Model Logistik		
Intersep	-0.98	0.3297
Nilai Nominal	3.66	0.0003*
<i>Debt Ratio</i>	-0.88	0.3792
Lama Angsuran	0.36	0.7176
<i>Loan to Value</i>	-0.92	0.3603
Lama Masa Kerja	0.83	0.4093
Usia	0.56	0.577
Jumlah Tanggungan	0.79	0.4325
Pekerjaan	-0.99	0.3225
Pendidikan Terakhir (Universitas)	1.87	0.0619
Pendidikan Terakhir (SMA/lebih rendah)	1.05	0.2932
Status Tempat Tinggal (Milik saudara/teman)	-2.02	0.0434*
Status Tempat Tinggal (Kontrak)	-1.07	0.2858
Status Pernikahan	1.57	0.1174
Jenis Kelamin	0.95	0.3428
Model Survival		
Intersep	6.67	<.0001
Nilai Nominal	1.6	0.1096
<i>Debt Ratio</i>	1.61	0.1088
Lama Angsuran	-0.55	0.5853
<i>Loan to Value</i>	2.33	0.0201*
Lama Masa Kerja	-0.09	0.9246
Usia	1.44	0.1494
Jumlah Tanggungan	-0.29	0.7719
Pekerjaan	-0.47	0.6394
Pendidikan Terakhir (SMA/lebih rendah)	-0.34	0.7363
Pendidikan Terakhir (Universitas)	-0.05	0.9615
Status Tempat Tinggal (Milik saudara/teman)	1.93	0.0536
Status Tempat Tinggal (Kontrak)	-0.04	0.9676
Status Pernikahan	-0.24	0.8138
Jenis Kelamin	1.37	0.1706
Shape	9.09	<.0001

* : Signifikan dengan taraf arti 5%

Pengujian hipotesis parameter model regresi Cure Log-logistik ketika z mempengaruhi peluang $\pi(z)$ dilakukan dengan menggunakan hipotesis:

$H_0 : \beta_i = 0$; parameter β_i tidak signifikan

$H_1 : \beta_i \neq 0$; parameter β_i signifikan

Dengan taraf arti sebesar 5%, berdasarkan tabel 5.21 diperoleh hasil kovariat yang berpengaruh secara signifikan adalah kovariat nilai nominal dan status tempat tinggal (milik saudara/teman). Artinya, jika nilai nominal pinjaman debitur bertambah sebesar 1 juta rupiah, maka peluang debitur akan mengalami macet membayar adalah 1,0160 kali lipat ($\exp(0,0159) = 1,0160$) dari peluang tidak terjadinya macet. Dengan kata lain peningkatan nilai nominal pinjaman debitur akan meningkatkan peluang terjadinya macet. Nilai taksiran untuk status tempat tinggal debitur dengan status milik saudara/teman adalah -3,2529. Hal ini menunjukkan bahwa dengan meningkatnya status tempat tinggal debitur menjadi 1 tingkat lebih baik dari status dengan tempat tinggal milik saudara/teman, maka peluang terjadinya macet adalah sebesar 0,0387 kali lipat ($\exp(-3,2529) = 0,0387$) lebih rendah dibandingkan debitur dengan status tempat tinggal milik saudara/teman. Dengan kata lain peningkatan status tempat tinggal debitur akan menurunkan peluang terjadinya macet.

17 6189

Sedangkan pada model survival, untuk melihat pengaruh kovariat yang mempengaruhi waktu mencicil secara signifikan menggunakan hipotesis sebagai berikut :

$H_0 : \gamma_i = 0$; parameter γ_i tidak signifikan

$H_1 : \gamma_i \neq 0$; parameter γ_i signifikan

Dengan taraf arti sebesar 5%, berdasarkan tabel 5.21 kovariat yang berpengaruh secara signifikan terhadap waktu mencicil hanya kovariat *loan to value*. Nilai taksiran untuk koefisien *loan to value* adalah 0,03436, hal ini menunjukkan bahwa dengan meningkatnya *loan to value* debitur sebesar 1 persen, maka akan menaikkan waktu mencicil debitur sebesar 0,03436 bulan.

5.11 Analisis Proporsi Debitur Macet dengan Pengaruh Kovariat

Debitur yang menjadi sampel dalam penelitian ini memiliki karakteristik-karakteristik yang berbeda, sehingga dapat diperoleh nilai proporsi debitur macet yang dipengaruhi oleh

berbagai kovariat yaitu kovariat nilai nominal, *debt ratio*, lama angsuran, *loan to value*, lama masa kerja, usia, jumlah tanggungan, pekerjaan, pendidikan terakhir (SMA/lebih rendah), pendidikan terakhir (universitas), status tempat tinggal (kontrak), status tempat tinggal (milik saudara/teman), status pernikahan dan kovariat jenis kelamin.

Taksiran fraksi *uncure* dan fraksi *cure* untuk data debitur KPR bank ABC, Tbk yang dipengaruhi oleh kovariat berjenis numerik dan kovariat berjenis kategorik dengan kategori 0 adalah :

$$\hat{\pi}(z) = \frac{1}{1 + e^{-1,5515}} = 0,83$$

Dengan demikian, taksiran fraksi *uncure* atau proporsi debitur KPR Bank ABC, Tbk yang mengalami macet dengan dipengaruhi oleh kovariat nilai nominal, *debt rasio*, lama angsuran, *loan to value*, lama masa kerja, usia, jumlah tanggungan > 2, pekerjaan debitur dengan status bukan karyawan, pendidikan terakhir debitur selain SMA/lebih rendah, pendidikan terakhir debitur selain universitas, status tempat tinggal debitur dengan status selain milik saudara/teman, status tempat tinggal debitur dengan status selain kontrak, debitur dengan status pernikahan belum menikah, dan debitur dengan jenis kelamin perempuan adalah sebesar 83% dan debitur yang tidak mengalami macet (*cure*) adalah sebesar $1 - \hat{\pi}(z)$ atau 17%.

Sedangkan, taksiran fraksi *uncure* dan fraksi *cure* untuk data debitur KPR bank ABC, Tbk yang dipengaruhi oleh kovariat berjenis numerik dan kovariat berjenis kategorik dengan kategori 1 adalah :

$$\hat{\pi}(z) = \frac{1}{1 + e^{0,1458}} = 0,46$$

Dengan demikian, taksiran fraksi *uncure* atau proporsi debitur KPR Bank ABC, Tbk yang mengalami macet dengan dipengaruhi oleh kovariat nilai nominal, *debt rasio*, lama angsuran, *loan to value*, lama masa kerja, usia, jumlah tanggungan ≤ 2 , pekerjaan debitur dengan status karyawan, pendidikan terakhir debitur SMA/lebih rendah, pendidikan terakhir debitur

universitas, status tempat tinggal debitur dengan status milik saudara/teman, status tempat tinggal debitur dengan status kontrak, debitur dengan status pernikahan menikah, dan debitur dengan jenis kelamin laki-laki adalah sebesar 46% dan debitur yang tidak mengalami macet (*cure*) adalah sebesar $1 - \hat{\pi}(z)$ atau 54%.

5.12 Luaran Penelitian

- a. Kudus, A., Suliadi dan Nurwahidah, A. I., 2015, *Model Credit Scoring Menggunakan Regresi Logistik Beserta Validasinya*, Prosiding KNMSA 2015 (ISBN: 978-979-99168-1-5), halaman 241-251
- b. Suliadi, Kudus, A., dan Wistara, R.R.A., 2015, *Regresi Logistik pada Data Rare Event*, Prosiding Statistika (ISSN: 2460-6456), halaman 1-7.
- c. Suliadi, Kudus, A., 2016, *Reducing Bias for Logistic Regression with Rare Event Data*, Proceedings The 2nd International Conference on Applied Statistics (ICAS II 2016) - AIP Proceedings Scopus Index
- d. Kudus, A., and Suliadi, 2016, *Comparison of Credit Scoring Model between Logistic Regression and Cox Proportional Hazard Model Based on Mean Cost Criteria*, submitted to Open Journal of Statistics

BAB 6. KESIMPULAN DAN SARAN

6.1 Kesimpulan

Berdasarkan hasil dan pembahasan yang telah dilakukan yaitu dengan memodelkan dan menghitung nilai *mean cost* data debitur KPR bank ABC, Tbk menggunakan model regresi logistik dan model *Cox proporsional hazard* dapat ditarik beberapa kesimpulan, diantaranya :

1. Model Regresi Logistik

- Model taksiran logit peluang kredit macet adalah sebagai berikut:

$$\begin{aligned}\hat{g}(x_i) = & -7,888 + 0,003x_{i1} - 0,077x_{i2} - 0,021x_{i3} - 0,014x_{i4} - 0,014x_{i5} \\ & + 0,138x_{i6} + 0,663x_{i7} - 1,039x_{i8} - 0,467x_{i9} - 0,168x_{i92} \\ & + 1,456x_{i101} + 1,468x_{i102} + 0,389x_{i11} + 1,297x_{i12}\end{aligned}$$

- Nilai *mean cost* atau kerugian akibat salah memprediksi model regresi logistik pada data *training* adalah sebesar 1,753968 sedangkan pada data *testing* adalah sebesar 2,321429.

2. Model *Cox proporsional hazard*

- Model *Cox proporsional hazard* yang dinyatakan dalam bentuk persamaan log adalah sebagai berikut:

$$\begin{aligned}\log \left\{ \frac{h_i(t)}{h_0(t)} \right\} = & 0,002x_{i1} - 0,018x_{i2} - 0,025x_{i3} + 0,004x_{i4} - 0,036x_{i5} + \\ & 0,157x_{i6} + 0,537x_{i7} - 0,089x_{i8} - 0,253x_{i9} - 0,008x_{i92} + \\ & 0,839x_{i101} + 0,351x_{i102} + 0,215x_{i11} + 0,758x_{i12}\end{aligned}$$

- Nilai *mean cost* atau kerugian akibat salah memprediksi model *Cox proporsional hazard* pada data *training* adalah sebesar 1,984127 sedangkan pada data *testing* adalah sebesar 2,380952.

3. Karena nilai *mean cost* model regresi logistik lebih kecil dari pada model *Cox proporsional hazard*, maka model regresi logistik sedikit lebih baik dari pada model *Cox proporsional hazard*.

Pemodelkan data debitur KPR bank ABC, Tbk dengan menggunakan model regresi Cure Log-logistik dapat ditarik beberapa kesimpulan, diantaranya :

1. Ketahanan mencicil debitur KPR bank ABC, Tbk melalui kurva fungsi ketahanan marginal, debitur KPR mengalami macet pembayaran paling besar sekitar bulan ke-11, bulan ke-29, bulan ke-35, hingga bulan ke-52.
2. Taksiran proporsi debitur KPR Bank ABC, Tbk yang mengalami macet tanpa dipengaruhi oleh berbagai kovariat adalah sebesar 39%. Sedangkan untuk debitur yang tidak mengalami macet adalah $1 - \hat{\pi}(z)$ atau 61%.
3. Kovariat nilai nominal, *debt rasio*, lama angsuran, *loan to value*, lama masa kerja, usia, jumlah tanggungan, pekerjaan, pendidikan terakhir, status tempat tinggal, status pernikahan, dan jenis kelamin debitur secara simultan berpengaruh terhadap peluang terjadinya macet dan waktu mencicil debitur KPR Bank ABC, Tbk.
4. Kovariat nilai nominal dan status tempat tinggal berpengaruh secara signifikan terhadap peluang terjadinya peristiwa macet, dan kovariat *loan to value* berpengaruh secara signifikan terhadap waktu mencicil debitur KPR bank ABC, Tbk.
5. Taksiran fraksi *uncure* atau proporsi debitur KPR Bank ABC, Tbk yang mengalami macet dengan dipengaruhi oleh kovariat nilai nominal, *debt rasio*, lama angsuran, *loan to value*, lama masa kerja, usia, jumlah tanggungan > 2 , pekerjaan debitur dengan status bukan karyawan, pendidikan terakhir debitur selain SMA/lebih rendah, pendidikan terakhir debitur selain universitas, status tempat tinggal debitur dengan status selain milik saudara/teman, status tempat tinggal debitur dengan status selain kontrak, debitur dengan status pernikahan belum menikah, dan debitur dengan jenis kelamin perempuan adalah

sebesar 83% dan debitur yang tidak mengalami macet (*cure*) adalah sebesar $1 - \hat{\pi}(z)$ atau 17%. Sedangkan taksiran fraksi *uncure* atau proporsi debitur KPR Bank ABC, Tbk yang mengalami macet dengan dipengaruhi oleh kovariat nilai nominal, *debt ratio*, lama angsuran, *loan to value*, lama masa kerja, usia, jumlah tanggungan ≤ 2 , pekerjaan debitur dengan status karyawan, pendidikan terakhir debitur SMA/lebih rendah, pendidikan terakhir debitur universitas, status tempat tinggal debitur dengan status milik saudara/teman, status tempat tinggal debitur dengan status kontrak, debitur dengan status pernikahan menikah, dan debitur dengan jenis kelamin laki-laki adalah sebesar 46% dan debitur yang tidak mengalami macet (*cure*) adalah sebesar $1 - \hat{\pi}(z)$ atau 54%.

6.2 Saran

1. Karena nilai \hat{S} untuk berbagai nilai t mulai dari t terkecil sampai dengan t terbesar dalam data ternyata \hat{S} masih disekitar 0,99, maka disarankan untuk menggunakan data dengan durasi penelitian yang lebih panjang.
2. Model *Cox proportional hazard* bisa digunakan untuk memodelkan data dengan variabel bebas yang bertipe *time dependent covariate*, dengan demikian jika model *Cox proportional hazard* dan model regresi logistik diterapkan untuk data tersebut maka harapannya model *Cox proportional hazard* akan lebih baik.
3. Disarankan untuk membahas model Cure untuk menaksir proporsi individu yang tidak mengalami peristiwa yang diperhatikan dengan distribusi peluang lain seperti weibull, eksponensial, dan log-normal.

DAFTAR PUSTAKA

- Abreu, A.M., Rocha, C.S. 2006. A Parametric Cure Model with Covariates. *Journal of the Computer Methods and Programs in Biomedicine*, 85, 151-162.
- Allen & Rose. 2006. Financial Survival Analysis of Defaulted Debtors. *Journal of the Operational Research Society*, 57, 630-636.
- Andreeva, G., 2006, European generic scoring models using survival analysis, *J. Oper. Res. Soc.*, 57, 1180-1187.
- Argesti, A. 1990. *An Introduction to Categorical Data Analysis*. Second ed. New Jersey: A Wiley Interscience Publication.
- Athoillah & Sudarno. 2012. Model Regresi Data Tahan Hidup Tersensor Tipe III Berdistribusi Log-logistik. *Jurnal Gaussian I*, no.1. 83-92
- Bravo, C., Maldonado, S. & Weber, R. 2013. Granting and managing loans for micro-entrepreneurs: new developments and practical experiences. *European Journal of Operational Research*, xxx, xxx-xxx.
- Budisantoro, Totok & Triandaru, Sigit. 2006, *Bank dan Lembaga Keuangan Lain*. Edisi 2, Jakarta : Salemba Empat.
- Cantor, Alan B. 2003. *SAS Survival Analysis Techniques for Medical Research, Second Edition*, North Carolina : SAS Institute Inc.
- Coffman, J. Y. 1986, The Proper Role of Tree Analysis in Forecasting the Risk Behavior of Borrowers, *MDS Reports, Management Decision Systems*, Atlanta, GA, 3, 4, 7, 9.
- Collett, D. 2003. *Modelling Survival Data in Medical Research*. US: Chapman & Hall.
- Corbiere & Joly. 2007. A SAS Macro For Parametric and Semiparametric Mixture Cure Models. *Journal of the Computer Methods and Programs in Biomedicine*, 85,173-180.
- Edelman, D. B., 2008, Credit this: how the banks decide your credit score, *Significance*, June, pp: 59 - 61
- Francesca, G. 2012. A Discrete-time models for loans: some evidence from italian banking system. *American Journal of Applied Sciences*, 9(9), 1337-1346.
- Grablowsky, B. J. and Talley, W. K., 1981, Probit and discriminant functions for classifying credit applicants: A comparison, *J. Econom. Business*, 33, 254-261.
- Hajarisman, Nusar. 2009. *Buku Ajar Analisis Data Kategorik*. Bandung: Program Studi Statistika Universitas Islam Bandung.
- Herli, Ali S. 2013. *Pengelolaan BPR dan Lembaga Keuangan Mikro*. Andi. Yogyakarta.
- Hosmer, D. W., & Lemeshow. (2000). *Applied Logistic Regression*. New York: John Wiley and Sons.
- Ibrahim, N. A., Kudus, A., Daud, I. and Abu Bakar, M. R. Decision Tree for Competing Risks Survival Probability in Breast Cancer Study. *Int. J. of Biomed. Sci.* 3(1):25-29 (2008).
- Kleinbaum, D. G., & Klein, M. (2005). *Survival Analysis: A Self-Learning Text Second Edition*. USA: Springer Science+Business Media, Inc.

- Klugman, S. A., Panjer, H. H., dan Willmot, G. E. (2004). *Loss Models: From Data to Decisions*. Edisi kedua, Wiley, New York.
- Kudus, A. Model competing risks dengan parameter proporsi yang sembuh. *Prosiding Konferensi Nasional Matematika, Sains dan Aplikasinya (KnMSA2013)*. Universitas Islam Bandung 20 Juni 2013. pp: 331-339. ISBN: 978-602-19356-1-3.
- Lawless, J. F. 2007. *The Statistical Analysis of Recurrent Event*. USA: Springer Science+Business Media, Inc.
- Lee, E.T. 2003. *Statistical Methods for Survival Data Analysis*. Second ed. New York: A Wiley Interscience Publication.
- Liu, Xian. 2012, *Survival Analysis Models and Applications*. USA: Jhon Wiley & Sons Inc.
- Makowski, P., 1985, Credit scoring branches out, *Credit World*, 75, 30-37.
- McCullagh, P., & J.A. Nelder. 1983. *Generalized Linear Models*. 2nd Ed. New York: Chapman and Hall
- Othus, Megan., Barlogie, Bart., Leblanc, Michael L., Crowley, John J. (2012). Cure Models as a Useful Statistical Tool for Analyzing Survival. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*. 18(14): 3731-3736.
- Prasetya, Laniati. 2006. *Penerapan Metode Survival Analysis dalam KPR (Studi Kasus pada PT. Bank ABC, Tbk)*, Program Pasca Sarjana, Magister Manajemen, Universitas Indonesia.
- Rezac, M. 2011. *How to Measure the Quality of Credit Scoring Models*. *Journal of Economics and Finance*, 5.
- Sarlija, N., Bencic, M. and Zekic-Susac, M., 2009, Comparison procedure of predicting the time to default in behavioral scoring, *Expert Systems with Applications*, 36, pp: 8778-8788.
- Siddiqi, N. 2006. *Credit Risk Scorecard Developing and Implementing Intelligent Credit Scoring*. New Jersey (US) : John Willey & Sons.
- Sjafitri, Henny. 2011. *Faktor-Faktor yang Mempengaruhi Kualitas Kredit dalam Dunia Perbankan*. *Jurnal Manajemen dan Kewirausahaan*.
- Suwondo & Santosa, Stefanus. 2014. *Credit Scoring Menggunakan Metode Support Vector Machine dengan Teknik Seleksi Atribut Berbasis Chi Squared Statistic dan Particle Swarm Optimization*.
- Suyatno, Thomas., dkk. 2010. *Dasar-Dasar Perkreditan Edisi Keempat*. Raja Grafindo Persada.
- Tong, E. N. C., Mues, C. & Thomas, L. 2012. Mixture cure models in credit scoring: if and when borrowers default. *European Journal of Operational Research*, 218, 132-139.
- Tonny, B., & Jonathan, C. 2007. *Credit Scoring With Macroeconomic Variables Using Survival Analysis*, Management School and Economics, University of Edinburgh.
- Turvey, C. G., 1991, Credit scoring for agricultural loans: a review with applications, *Agricultural Finance Review*, 51, pp:43-54

Wiginton, J. C., 1980, A note on the comparison of logit and discriminant models of consumer credit behaviour, *J. Financial Quantitative Anal.*, 15, 757-770.



BIODATA DOSEN PENELITI

I. Identitas Diri

Nama	:	Abdul Kudus
NIK	:	D.95.0.223
NIDN	:	0421036901
Tempat dan Tanggal Lahir	:	Subang, 21 Maret 1969
Golongan / Pangkat	:	IIID
Jabatan Fungsional	:	Lektor
Fakultas/Program Studi	:	MIPA, Statistika
Alamat Rumah	:	Komplek Unisba No. 38 Jalan Unisba III Desa Mandalamekar Kecamatan Cimencyan Kab Bandung 40197
Telp./Fax	:	-
Nomor HP	:	085624607194
Alamat e-mail	:	akudus69@yahoo.com

II. Riwayat Pendidikan

Program	S1	S2	S3
Nama Perguruan Tinggi	IPB	IPB	Universiti Putra Malaysia
Bidang Ilmu	Statistika	Statistika	Statistika
Tahun Masuk	1989	1996	2003
Tahun Lulus	1994	1999	2009
Judul Skripsi/Thesis /Disertasi	Analisis produksi tanaman padi dengan model CESS	Penggunaan metode regresi pohon untuk analisis data ketahanan hidup	Tree-structured and direct parametric regression models for the subdistribution of competing risks

III. Pengalaman Penelitian (5 tahun terakhir)

No	Tahun	Judul Penelitian	Ketua/Anggota	Sumber Dana, Jumlah (Rp Juta)
1.	2014	Peningkatan Tata Kelola Untuk Percepatan Akreditasi Institusi	Anggota	Unisba Rp 14
2.	2014	Uji Kesesuaian dan Performanya pada Model-Model Semiparametrik Untuk Data Biner Berkorelasi (tahun ke-2)	Anggota	DIKTI Rp 45
3.	2013	Pemodelan Peluang Kesembuhan Pasien untuk Kasus Penyakit Komplikasi (tahun ke-2)	Ketua	DIKTI Rp 36
4.	2012	Pemodelan Peluang Kesembuhan Pasien untuk Kasus Penyakit Komplikasi (tahun ke-1)	Ketua	DIKTI Rp 37
5.	2011	Parameter Estimation in the Presence of Nondetect (Left-censored) Data	Ketua	TWAS - The Academy of Sciences for the Developing World - UNESCO US\$7000
6.	2011	Pemodelan Peluang Kesembuhan Pasien	Ketua	Unisba Rp 12
7.	2010	Estimation of Default Probability Using Survival Analysis	Anggota	Ministry of Higher Education Malaysia RM20000

IV. Pengalaman Pengabdian kepada Masyarakat (5 tahun terakhir)

No	Tahun	Judul Penelitian	Ketua/Anggota	Sumber Dana, Jumlah (Rp)
-	-	-	-	-

V. Pengalaman Penulisan Artikel Ilmiah dalam jurnal (5 tahun terakhir)

No.	Tahun	Judul Artikel	Penulis Utama/Anggota	Nama Jurnal, Vol., No.	Terakreditasi/Belum Terakreditasi
1	2010	Identification of Multiple High Leverage Points in Weibull Censored Regression Model	Anggota	<i>Journal of Applied Probability & Statistics</i> , Vol. 5, No. 2, th. 2010	Terakreditasi
2	2010	Optimal Cutoff Point for Dichotomization of Continuous Independent Variable in Competing Risks Data Analysis	Penulis Utama	<i>Math Digest</i> , Vol.3, No. 2, th. 2010	Terakreditasi
3	2010	Simulation on Group Deleted Potentials for Diagnostics of Survival Regression	Anggota	<i>Journal of Materials Science and Engineering</i> , Vol. 4, No.7, th. 2010	Terakreditasi

VI. Pengalaman Penulisan Buku (5 tahun terakhir)

No.	Tahun	Judul Buku	Penulis Utama/ Anggota	Penerbit	Jumlah Halaman
-	-	-	-	-	-

Biodata sebagai salah satu syarat dalam pengajuan Proposal Penelitian LPPM UNISBA, dan apabila di kemudian hari ternyata dijumpai ketidaksesuaian, saya sanggup menerima sanksinya.

Bandung, 30 November 2016

Peneliti,

Abdul Kudus, Ph.D.
D.95.0.223

BIODATA ANGGOTA PENELITI

A. Identitas Diri

1	Nama lengkap	Dr. Suliadi, S.Si., M.Si.
2	Jenis Kelamin	L
3	Jabatan Fungsional	Lektor
4	NIK	D.97.0.267
5	NIDN	0416117202
6	Tempat dan Tanggal Lahir	Malang, 16 Nopember 1972
7	Alamat E-mail	suliadi@gmail.com
8	Nomor Telepon/ HP	022-7217343/085846252822
9	Alamat Kantor	Jl. Ranggamalela No. 01 Bandung 40116
10	Nomor Telepon/Faks	022-4203368 ext 136/022-4263895
12	Lulusan yang Telah Dihasilkan	S-1=10 Orang; S-2= - ; S-3 = -
14	Mata Kuliah yang Diampu	1. Analisis Regresi - Semester Genap (sekarang) 2. Matematika Keuangan - Semester Genap (sekarang) 3. Prak. Analisis Regresi - Semester Genap (sekarang) 4. Pengantar Peluang- Statistika (Semester Ganjil) 5. Pengantar Peluang- Matematika (Semester Ganjil) 6. Prak. Rancangan Percobaan (Semester Ganjil)

B. Riwayat Pendidikan

	S1	S2	S3
Nama Perguruan Tinggi	Institut Pertanian Bogor	Institut Pertanian Bogor	Universiti Putra Malaysia - Malaysia
Bidang Ilmu	Statistika	Statistika	Statistika
Tahun Masuk-Lulus	1990-1996	2000-2003	2006-2011
Judul Skripsi/Thesis /Disertasi	Analisis Pengaruh Lingkungan Terhadap Genotip Galur-galur Kedelai	Model Fuzzy untuk Menentukan Indeks Pembangunan Berkelanjutan di Indonesia	GEE-Smoothing Spline for Semiparametric Estimation of Longitudinal Categorical Data
Nam Pembimbing /Promotor	Dr. Totong Martono	Dr. Bambang Djuanda	Prof. Dr. Noor Akma Ibrahim

C. Pengalaman Penelitian dalam Lima Tahun Terakhir

No	Tahun	Judul Penelitian	Pendanaan	
			Sumber	Jumlah (Juta Rp)
1	2012	Pemodelan Peluang Ketersembuhan Pasien Untuk Kasus Penyakit Kompilasi (Tahun ke-1)	Fundamental	37
2	2013	Pemodelan Peluang Ketersembuhan Pasien Untuk Kasus Penyakit Kompilasi (Tahun ke-2)	Fundamental	40
3	2013	Uji Kesesuaian dan Performanya pada Model-Model Semiparametric Untuk Data Biner Berkorelasi (Tahun ke-1)	Fundamental	30

D. Pengalaman Pengabdian Kepada Masyarakat dalam Lima Tahun Terakhir

No	Tahun	Judul Pengabdian Kepada Masyarakat	Pendanaan	
			Sumber	Jumlah (Juta Rp)
1	2012	-	-	-

E. Pengalaman Penulisan Artikel Ilmiah Dalam Jurnal Dalam Lima Tahun Terakhir

No	Judul Artikel Ilmiah	Volume/Nomor /Tahun	Nama Jurnal
1	Semiparametric Estimation with Profile Algorithm for Longitudinal Binary Data	42/1/2013	Communication in Statistics-Simulations and Computations
2	Generating correlated discrete ordinal data using R and SAS IML	104/3 /2011	Computer Methods and Programs in Biomedicine
3	Nonparametric Regression for Longitudinal Binary Data Based on GEE-Smoothing Spline	5/1/2010	Journal of Applied Statistics and Probability
4	GEE-Smoothing Spline for Semiparametric Estimation of Longitudinal Binary Data.	18/S10/2010	International Journal of Applied Mathematics and Statistics
5	Nonparametric Regression for Correlated Data	8/7/2009	WSEAS Transaction on Mathematics

6	Comparison of Some Smoothing Parameter Selection Methods in GEE-Smoothing Spline	7/11/2009	World Applied Sciences Journal
---	--	-----------	--------------------------------

F. Pengalaman Penyampaian Makalah Secara Oral Pada Pertemuan/Seminar Ilmiah Dalam Lima Tahun Terakhir

No	Nama Pertemuan	Judul Artikel Ilmiah	Waktu dan Tempat
1	The 3rd International Conference on Mathematics and Statistics (ICoMS-3), Institut Pertanian Bogor	GEE-Smoothing Spline For Longitudinal Data	Bogor, August 5-6, 2008
2	Seminar Kebangsaan Pengoptimuman Berangka dan Penyelidikan Operasi Ke-2	Semiparametric Estimation for Longitudinal Data: GEE-Smoothing Spline Approach	Kuala Terengganu, Terengganu, Malaysia, December 13-14, 2008
3	The 1st Regional Conference on Applied and Engineering Mathematics (RCAEM-1),	Semiparametric Models for Correlated Nominal Data	Penang, Malaysia, June 2-3, 2010
4	Seminar Kebangsaan Aplikasi Sains dan Matematik 2010 (SKASM2010) sempena Simposium Kebangsaan Sains Matematik 2010 (SKSM2010)	The Implementation of Lee's Algorithm in Generating Correlated Nominal Data with SAS IML	Johor Bahru, Johor, Malaysia, December 8-11, 2010
5	Konferensi Nasional Matematika, Sains dan Aplikasinya 2013	Hat Matrik pada GLM	Unisba, Bandung, 22 Juni 2013
6	The 9th International Conference on Mathematics, Statistics and Applications (ICMSA) 2013	Goodness of Fit Test of Semiparametric Model for Correlated Binary Data Based on Standardized Residual	Medan, 12-13 Desember 2013

G. Pengalaman Penulisan Buku Dalam Lima Tahun Terakhir

No	Judul Buku	Tahun	Jumlah Halaman	Penerbit
1	-			

H. Pengalaman Perolehan HKI Dalam 5-10Tahun Terakhir

No	Judul Tema HKI	Tahun	Jenis	Nomor P/ID
1	-			

I. Pengalaman Merumuskan Kebijakan Publik/Rekayasa Sosial Lainnya Tahun Terakhir

No	Judul/Tema Rekayasa Sosial Lainnya yang Telah Diterapkan	Tahun	Tempat Penerapan	Respons Masyarakat
1				

J. Penghargaan yang Pernah Diraih Dalam 10 Tahun Terakhir (dari Pemerintah, Asosiasi atau Institusi Lainnya)

No	Jenis Penghargaan	Institusi Pemberi Penghargaan	Tahun
1	The Invention, Research and Innovation Exhibition – Medali Perak	Universiti Putra Malaysia	2010
2	Pre World Statistic Day Inter-Varsity Poster Competition on Post-Graduate Research – Runner up	University of Malaya – Institut Statistik Malaysia	2010
3	The Invention, Research and Innovation Exhibition – Medali Perunggu	Universiti Putra Malaysia	2011

Semua data yang saya isikan dan tercantum dalam biodata ini adalah benar dan dapat dipertanggungjawabkan secara hukum. Apabila dikemudian hari ternyata dijumpai ketidaksesuaian dengan kenyataan, saya sanggup menerima resikonya.

Demikian biodata ini saya buat dengan sebenarnya untuk memenuhi salah satu persyaratan dalam pengajuan Hibah Penelitian Fundamental.

Bandung, 30 November 2014

Peneliti,

Dr. Suliadi, S.Si., M.Si.

CURRICULUM VITAE

A. Identitas Diri

1	Nama Lengkap (denganelar)	Dr. Pupung Purnamasari, SE, M.Si, Ak.,CA
2	Jenis Kelamin	Perempuan
3	Jabatan Fungsional	Lektor
4	NIP/NIK/Identitas lainnya	D.93.0.186
5	NIDN	0416046901
6	Tempat dan Tanggal Lahir	Subang, 16 April 1969
7	E-mail	p_purnamasari@yahoo.co.id
9	Nomor Telepon/HP	08122381608/ 082138093344
10	Alamat Kantor	Jl. Tamansari no 1 Bandung
11	Nomor Telepon/Faks	(022) 4264064 / 4264064
12	Lulusan yang Telah Dihasilkan	S-1 = 165 orang; S-2 = 0 orang; S-3 = 0 orang
13.	Mata Kuliah yang Diampu	1. Pengauditan
		2. Pengauditan Internal
		3. Seminar Pengauditan
		4. Akuntansi Forensik

B. Riwayat Pendidikan

	S1	S2	S3
Nama Perguruan Tinggi	UNISBA	UNPAD	UNDIP
Bidang Ilmu	Akuntansi	Akuntansi	Akuntansi
Tahun Masuk-Lulus	1988 – 1993	1995 – 1999	2010-2014
Judul Skripsi/Tesis/Disertasi	Peranan Anggaran Sebagai Alat untuk Mengendalikan Pembelian Alat Medis	Pengaruh Prosedur Akuntansi, Prosedur Kepatuhan Terhadap Kepatuhan Pengusaha EPTE (Enterport Produksi Tujuan Ekspor).	“Perilaku Penghentian Prematur atas Prosedur Audit dalam Perspektif Keputusan berbasis Etika (Studi Empiris pada Kantor Akuntan Publik)”,
Nama Pembimbing/Promotor	Wachyudin Zarkasy, SE, Ak., M.Si	Dr. Sumarno Zain, Akt.	Prof. Dr. Abdul Rohman, M.Si., Akt

C. Pengalaman Penelitian (4 tahun terakhir)

No	Tahun	Judul Penelitian
1	2010	Ketua Peneliti dalam Penelitian LPPM UNISBA Pengaruh Kecerdasan Emosional dan Kecerdasan Spiritual Terhadap Sikap Etis Mahasiswa (Studi Kasus pada Perguruan Tinggi Swasta di Bandung)
2	2012	Ketua Peneliti pada Penelitian LPPM UNISBA Pengaruh Faktor Eksternal dan Internal Auditor terhadap Perilaku Penghentian Prematur Atas Prosedur Audit (Studi Kasus pada Kantor Akuntan Publik di Bandung)
3	2012	Diagnosa Model Sosial sebagai Strategi Pengembangan Usaha Kecil Menengah
4	2013	How to optimize the potential of zakat
5	2014	Islamic Corporate Social Responsibility Reporting Disclosure: Its Determinants
6	2014	Determinant factors of cognitive moral development in audit activities: Ethical decision perspective (empirical study on public accounting firms)
7	2015	Influence of employee fraud on Asset misappropriation analyzed by fraud diamond dimension,
8	2015	Fraud Prevention : Relevance to religiosity and spirituality in the workplace
9	2015	Pluriform motivation as antecedent and its relationships to budgeting participation and managerial performance
10	2015	The impact of the values of Islamic religiosity on Islamic employee's work ethics and job satisfaction in Tasikmalaya West Java, Indonesia, Industrial center

D. Pengalaman Pengabdian Kepada Masyarakat Dalam 5 Tahun Terakhir

No	Tahun	Judul PKM	Sumber Dana	Jumlah Dana
4	2009	Pelatihan Software MYOB Terhadap Guru SMU Sekota Madya Bandung, yang dilaksanakan pada tanggal 13 November 2009. Kerjasama Program studi Akuntansi dan Dinas Pendidikan	Program Studi Akuntansi Unisba dengan Dinas Pendidikan Pemerintah Kota Bandung	Rp.15.000.000
5	2013	Pelatihan kompetensi akuntansi bagi UMKM di	IPF	Rp 15.000.000

No	Tahun	Judul PKM	Sumber Dana	Jumlah Dana
		kecamatanKiangrokePangalengan		
6	2013	Pelatihankompetensiakuntansibagi Guru SD dan SMP untukpenyusunan dan BOS di kecamatanKiangrokePangalengan	IPF	Rp 15.000.000
7	2013	Pelatihankompetensiakuntansibagipengurus masjid di kecamatanKiangrokePangalengan	IPF	Rp 15.000.000
8	2013	Pelatihankompetensiakuntansibagisiswa SMA dan SMK di kecamatanKiangrokePangalengan	IPF	Rp 15.000.000
9	2015	Pelatihankompetensiakuntansibagi UMKM di kecamatanCibadakdanCitatayanKabupatenSukabumi	IPF	Rp. 15.000.000
10	2015	Pengobatan Gratis di kecamatanCibadakdanCitatayanKabupatenSukabumi	IPF	Rp15.000.000
11	2015	Pelatihankompetensiakuntansibagipengurus masjid di kecamatanCibadakdanCitatayanKabupatenSukabumi	IPF	Rp15.000.000
12	2015	Pelatihankompetensiakuntansibagisiswa SMA dan SMK di kecamatanCibadakdanCitatayanKabupatenSukabumi	IPF	Rp15.000.000

E. Publikasi Artikel Ilmiah (4 tahun terakhir)

No	Judul Artikel Ilmiah	Nama Jurnal	Volume/ Nomor/ Tahun
1	Pengaruh Kecerdasan Emosional dan Kecerdasan Spiritual Terhadap Sikap Etis Mahasiswa (Studi Kasus pada Perguruan Tinggi Swasta di Bandung)	Performa Jurnal Manajemen, UNISBA	Volume VIII No. 2 September 2011
2	Diagnosa Model Sosial sebagai Strategi Pengembangan Usaha Kecil Menengah	Prosiding Seminar Nasional dan Call for Papers, FEB Unisba, Bandung.	2012
3	Determinant factors of cognitive moral development in audit activities: Ethical decision perspective (empirical study on public accounting firms)	International Journal of Research in Business and Technology, CANADA	Volume 5 NO. 2 October 2014

4	Islamic Corporate Social Responsibility Reporting Disclosure: Its Determinants	Conference Proceeding. FEB Universitas Indonesia, Depok	2014
5	Fraud Prevention : Relevance to religiosity and spirituality in the workplace,	Elsevier Procedia Journal of Sosial and Behavioral Sciences (scienceDirect) Multidisciplinary perspective on Management and Society	Volume 211, 2015
6	Pluriform motivation as antecedent and its relationships to budgeting participation and managerial performance	Elsevier Procedia Journal of Sosial and Behavioral Sciences (scienceDirect) Multidisciplinary perspective on Management and Society	Volume 211, 2015
7	The impact of the values of Islamic religiosity on Islamic employee's work ethics and job satisfaction in Tasikmalaya West Java, Indonesia, Industrial center	Elsevier Procedia Journal of Sosial and Behavioral Sciences (scienceDirect) Multidisciplinary perspective on Management and Society	Volume 211, 2015

F. Pemakalah Seminar Ilmiah (*Oral Presentation*) (4 tahun terakhir)

No	Nama Pertemuan Ilmiah / Seminar	Judul Artikel Ilmiah	Waktu dan Tempat
1	Seminar Nasional & Call For Paper	Pemberdayaan Industri Kecil dan Menengah dalam Upaya Membangun Ekonomi Kreatif	9 – 10 Feb 2012, FEB Unisba, Bandung. (Presenter)
2	International Conference and Call for Papers	Analysis Of Effect Of Implementation Of Internal Control Organizational Culture And Total Quality Management To Application Of Good Governance	27 th – 28 th June 2013, Hotel Jayakarta– FEB Unisba, Bandung. (Presenter)

3	International Conference and Call for Papers	How to optimize the potential of zakat	27 th – 28 th June 2013, Hotel Jayakarta-FEB Unisba, Bandung. (Presenter)
4	International Accounting Conference & Student Forum (IACSF)	Islamic Corporate Social Responsibility Reporting Disclosure: Its Determinants	24-25 November 2014, FEB Universitas Indonesia, Depok (Presenter)
5	International conference on Accounting Studies (ICAS)	Influence of employee fraud on Asset misappropriation analyzed by fraud diamond dimension,	17 th -20 th August 2015 Thistle Hotel, Johor Bahru Malaysia (Presenter)
6	2 nd Global Conference on Business and Social Sciences	Fraud Prevention : Relevance to religiosity and spirituality in the workplace,	17 th – 18 th September, Bali Indonesia. 2015(Presenter)

G. Pengalaman Penulisan Buku dalam 5 Tahun Terakhir

No	Judul Buku	Tahun	Jumlah Halaman	Penerbit
1	Perkembangan Moral Kognitif dan Perilaku Penghentian Prematur Dalam Audit	2014	194	P2U LPPM Unisba

Semua data yang saya isikan dan tercantum dalam biodata ini adalah benar dan dapat dipertanggungjawabkan secara hukum. Apabila di kemudian hari ternyata dijumpai ketidak-Sesuaian dengan kenyataan, saya bersedia mempertanggungjawabkannya.

Bandung, 30 November 2016
Peneliti,

Dr. Pupung Purnamasari, SE., M.Si., Ak., CA

LAMPIRAN



Model Credit Scoring Menggunakan Regresi Logistik Beserta Validasinya

Ade Irma Nurwahidah, Abdul Kudus, Suliadi

Program Studi Statistika Universitas Islam Bandung
e-mail: adeirmanurwahidah@gmail.com; akudusmillis@gmail.com; suliadi@gmail.com

Abstrak

Model credit scoring merupakan suatu alat dan teknik untuk meminimalkan risiko pada lembaga keuangan khususnya bank. Dari hasil model credit scoring akan didapat kartu skor hasil diskretisasi recursive partitioning dengan CART yaitu nilai untuk setiap kategori dari karakteristik calon debitur (variabel bebas) dengan menggunakan regresi logistik. Untuk mengukur seberapa baik model credit scoring dalam mengklasifikasikan nasabah yang baik dan buruk dan memperkuat ketetapan prediksi model maka dilakukan validasi terhadap model. Dalam skripsi ini, terdapat beberapa ukuran-ukuran untuk melakukan validasi model yaitu kurva Receiver Operating Characteristic (ROC), Statistik Kolmogorov-Smirnov (KS), Indeks Gini dan C-Statistik. Dalam mengaplikasikan penelitian ini, penulis menggunakan data kredit Jerman 1994. Setelah dilakukan analisis, bahwa model baik atau ideal sehingga model credit scoring dapat digunakan untuk penyeleksian nasabah yang akan menerima kredit.

Kata Kunci: Model Credit Scoring, Scorecard, Diskretisasi Recursive Partitioning, Regresi Logistik, Receiver Operating Characteristic (ROC), Statistik Kolmogorov-Smirnov (KS), Indeks Gini, C-Statistik.

1. Pendahuluan

Dalam menjalani kehidupan, setiap orang atau suatu lembaga menginginkan keuntungan bukan kerugian yang mengakibatkan risiko. Secara umum risiko dapat diartikan sebagai suatu keadaan yang dihadapi seseorang atau perusahaan dimana terdapat kemungkinan yang merugikan. Di zaman sekarang ini, banyak risiko yang terjadi di berbagai lingkup lembaga, misalnya saja pada lembaga keuangan yaitu bank. Istilah risiko yang terjadi di bidang tersebut yaitu risiko kredit. Risiko kredit terjadi karena ketidakmampuan nasabah atas kewajiban pembayaran utangnya baik utang pokok maupun bunganya atau kedunya (kredit macet). Faktor penyebab timbulnya kredit macet salah satunya yaitu penyempurnaan credit scoring yang kurang mampu dilakukan oleh pihak bank (Sjafitri, 2011).

Untuk kebanyakan bank, risiko kredit merupakan risiko terbesar yang dihadapinya karena dapat menguras modal bank dengan cepat. Selain itu, peranan bank sebagai lembaga intermediasi tidak dapat berfungsi sehingga akan memperkecil kesempatan peluang bisnis, proyek baru, lapangan kerja baru, dan sebagainya. Terdapat cara untuk meminimalkan risiko ini yaitu dengan melakukan analisis risiko melalui model credit scoring.

Model credit scoring merupakan suatu alat dan teknik prediksi yang membantu lembaga keuangan dalam pemberian kredit (Rezac, 2011). Suwondo dan Santosa (2014) menyebutkan bahwa tujuan pembuatan model credit scoring untuk menganalisa dan membuat keputusan yang lebih cepat, tepat dan efisien terhadap penyeleksian nasabah yang akan menerima kredit. Model credit scoring akan menghasilkan scorecard yaitu nilai untuk setiap kategori dari karakteristik calon debitur (variabel bebas). Regresi logistik merupakan teknik yang umum digunakan untuk mengembangkan scorecard di sebagian lembaga keuangan, dimana variabel yang diprediksi adalah variabel kategori (Siddiqi, 2006). Dalam skripsi ini regresi logistik yang digunakan menggunakan regresi logistik biner.

Di samping itu, untuk mengukur seberapa baik model credit scoring dalam mengklasifikasikan nasabah yang baik dan buruk dan memperkuat ketetapan prediksi model maka dilakukan validasi terhadap model. Model yang baik akan berdampak terhadap penyeleksian calon nasabah yang akan

menerima pinjaman secara akurat. Terdapat beberapa ukuran yang dapat digunakan untuk melakukan validasi model dan pihak lembaga dapat memilih ukuran yang untuk melakukan validasi model. Tujuan dari penelitian ini yaitu mempraktikkan beberapa ukuran validasi model credit scoring diantaranya kurva Receiver Operating Characteristic (ROC), Area Under the Curve (AUC), Statistik Kolmogorov-Smirnov (KS), Indeks Gini, dan C-Statistik.

2. Kajian Pustaka

2.1. Kredit

Menurut undang-undang Nomor 14 tahun 1967 tentang Pokok-Pokok Perbankan, yang dimaksud dengan kredit adalah: "Penyediaan uang atau tagihan-tagihan yang dapat disamakan dengan itu berdasarkan persetujuan pinjam-meminjam antara bank dengan pihak lain dalam hal mana pihak peminjam berkewajiban melunasi utangnya setelah jangka waktu tertentu dengan jumlah bunga yang telah ditentukan".

2.2 Diskritisasi

Diskritisasi merupakan proses transformasi data kuantitatif menjadi pengkategorian yang berguna untuk proses scorecard. Menurut Kotsiantis dan Kanellopoulos (2006) terdapat empat tahapan diskritisasi, yaitu:

1. Mengurutkan nilai kontinu yang akan didiskritisasi.
2. Mengevaluasi titik potong sebagai pemisah selang atau penggabung selang yang berdekatan.
3. Berdasarkan kriteria tertentu dilakukan pemisahan atau penyatuan selang nilai.
4. Menghentikan proses pada titik tertentu.

Salah satu metode diskritisasi yaitu metode tersupervisi dan tidak tersupervisi. Dalam penelitian ini menggunakan metode diskritisasi tersupervisi yaitu recursive partitioning dengan CART (Classification and Regression Trees). CART adalah salah satu metode atau algoritma dari salah satu teknik eksplorasi data yaitu teknik keputusan. Algoritma pembentukan pohon klasifikasi terdiri dari empat (Kardiana dkk, 2006), yaitu pemilihan pemilih, penentuan simpul terminal, penandaan label kelas, dan penentuan pohon dengan ukuran tepat.

1. Pemilihan pemilih

Pada tahap ini dicari pemilih dari setiap simpul yang menghasilkan penurunan tingkat keheterogenan paling tinggi. Heterogenan yaitu simpul diukur berdasarkan nilai impurity-nya. Fungsi impuritas yang dapat digunakan adalah indeks Gini. Bila impuritas suatu simpul semakin besar maka semakin heterogen simpul tersebut (Brieman dkk, 1993).

Nilai impuritas menggunakan indeks Gini pada simpul t , $i(t)$, dapat ditulis sebagai berikut:

$$i(t) = 1 - \sum_j p_j^2(t) \quad \dots(2.1)$$

dimana $p_j(t)$ adalah peluang unit pengamatan dalam kelas ke- j dari simpul t yang dinyatakan sebagai berikut:

$$p_j(t) = \frac{\pi_j N_j(t) / N_t}{\sum_j \pi_j N_j(t) / N_t} \quad \dots(2.2)$$

Dengan π_j adalah peluang awal kelas ke- j , N_t adalah banyaknya unit pengamatan dalam ke- t , dan $N_j(t)$ adalah banyaknya unit pengamatan yang termasuk ke dalam kelas ke- j pada simpul t .

2. Penentuan simpul terminal

Suatu simpul t akan menjadi simpul terminal atau tidak akan dipilih kembali, jika jumlah pengamatannya kurang dari jumlah minimum. Umumnya jumlah pengamatan minimum pada

simpul besar 5 dan terkadang berjumlah 1 (Brieman dkk, 1993). Maka selanjutnya t tidak dipilih lagi tetapi dijadikan simpul terminal dan hentikan pembuatan pohon.

3. Penandaan label kelas

Label kelas dri simpul terminal ditentukan berdasarkan aturan jumlah terbanyak, yaitu jika $P(j_0 | t) = \max_j P(j | t)$, maka label kelas untuk terminal t adalah j_0 (Brieman dkk, 1993).

4. Penentuan pohon optimum

Menurut Brieman dkk. (1993), salah satu cara mendapatkan pohon optimum yaitu dengan pemangkasan (pruning). Pemangkas berturut-turut memangkas pohon bagian yang kurang penting. Tingkat kepentingan sebuah pohon bagian diukur berdasarkan ukuran biaya kompleksitas (cost-complexity). Persamaanya adalah:

$$R_{\alpha}(T_k) = R(T_k) + \alpha |T_k| \quad \dots(2.3)$$

Hasil proses pemangkasan berupa sederet pohon klasifikasi T_k dan dengan validasi saling (cross-validation sample) dapat ditentukan pohon optimum T_{k^0} sebagai berikut:

$$R^{cv}(T_{k^0}) = \min(R^{cv}(T_k))$$

2.3. Weight of Evidence (WoE)

Weight of Evidence (WoE) adalah suatu nilai yang digunakan untuk mengukur kekuatan setiap kategori dari variabel bebas setelah didiskritisasi (Siddiqi, 2006). Perhitungan nilai WoE dilakukan untuk setiap kategori di variabel bebas hasil diskritisasi. WoE dapat dirumuskan sebagai berikut:

$$WoE_i = \left[\ln \left(\frac{DistrGood_i}{DistrBad_i} \right) \right] \times 100 \quad \dots(2.4)$$

Dimana:

Distr Good_i : Presentase jumlah nasabah kategori ke-i pada kelompok nasabah yang baik terhadap jumlah nasabah kelompok baik.

Distr Bad_i : Presentase jumlah nasabah kategori ke-i pada kelompok nasabah yang buruk terhadap jumlah nasabah kelompok buruk.

i: 1, 2, 3, ..., k

k : Banyaknya kategori untuk variabel bebas tertentu.

2.4. Information Value (InV)

Information Value (InV) merupakan suatu nilai yang digunakan untuk mengukur kekuatan dari variabel bebas setelah didiskritisasi (Siddiqi, 2006). Nilai InV digunakan untuk menyeleksi variabel bebas mana saja yang akan masuk kedalam model regresi logistik berdasarkan nilai batas tingkat prediksi tertentu. Berdasarkan SAS Institute (2012) jika nilai InV kurang dari 0,02 maka variabel bebas tidak dimasukkan kedalam model. Information Value dapat dirumuskan sebagai berikut:

$$InV = \sum_{i=1}^k (DistrGood_i - DistrBad_i) * \ln \left(\frac{DistrGood_i}{DistrBad_i} \right) \quad \dots(2.5)$$

Menurut Siddiqi (2006) tingkat prediksi InV dibagi kedalam beberapa kategori, yaitu:

1. Nilai $\ln V < 0,02$: variabel bebas tidak prediktif
2. Nilai $0,02 < \ln V < 0,1$: variabel bebas memiliki tingkat prediktif lemah
3. Nilai $0,1 < \ln V < 0,3$: variabel bebas memiliki tingkat prediktif medium
4. Nilai $\ln V \geq 0,3$: variabel bebas memiliki tingkat prediktif kuat

Berdasarkan SAS Institute Inc (2012) jika nilai $\ln V$ kurang dari 0,02 maka variabel bebas dikatakan tidak prediktif sehingga variabel bebas tersebut tidak dimasukkan kedalam model.

2.5. Regresi Logistik

Regresi logistik merupakan metode analisis statistika yang digunakan untuk menganalisis hubungan antara variabel tak bebas yang bersifat biner atau dikotomis dengan satu atau lebih variabel bebas (Hosmer dan Lemeshow, 2000). Pada regresi logistik, variabel tak bebas berskala kategorik. Variabel tak bebas yang dinotasikan dengan y bersifat biner atau dikotomis yang mempunyai dua nilai yaitu 0 dan 1. Dalam keadaan demikian, variabel y mengikuti distribusi Bernoulli untuk setiap observasi tunggal. Fungsi probabilitas untuk setiap observasi diberikan sebagai berikut:

$$f(y) = \pi^y(1 - \pi)^{1-y} \quad y = 0,1 \quad \dots(2.6)$$

sehingga diperoleh:

$$\text{jika } y = 0 \text{ maka } f(y) = \pi^0(1 - \pi)^{1-0} = 1 - \pi$$

$$\text{jika } y = 1 \text{ maka } f(y) = \pi^1(1 - \pi)^{1-1} = \pi$$

Fungsi regresi logistiknya dapat dituliskan sebagai berikut:

$$f(y) = \frac{1}{1+e^{-y}} \quad \text{atau} \quad f(y) = \frac{e^y}{1+e^y} \quad \dots(2.7)$$

dimana $y = \beta_0 - \beta_1 x_1 - \dots + \beta_p x_p$ dengan p = banyak variabel prediktor. Nilai y antara $-\infty$ dan $+\infty$ sehingga nilai $f(y)$ terletak antara 0 dan 1 untuk setiap nilai y yang diberikan. Hal tersebut menunjukkan bahwa model logistik sebenarnya menggambarkan probabilitas atau risiko dari suatu objek. Model regresi logistik adalah sebagai berikut.

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}, \text{ dimana } p = \text{banyaknya variabel prediktor.} \quad \dots(2.8)$$

Fungsi $\pi(x)$ di atas berbentuk non linear sehingga untuk membuatnya menjadi fungsi linier harus dilakukan transformasi logit sebagai berikut:

$$g(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x_1 - \dots - \beta_p x_p \quad \dots(2.9)$$

2.5.1. Pendugaan Parameter Regresi Logistik

Pendugaan parameter yang digunakan regresi logistik adalah metode kemungkinan maksimum (maximum likelihood, ML). Prinsip dasar dari metode kemungkinan maksimum adalah memilih suatu penaksir parameter sedemikian rupa sehingga dapat memaksimumkan fungsi peluang yang diamati. Jika x_i dan y_i adalah pasangan variabel prediktor dan terikat pada pengamatan ke - i dan diasumsikan bahwa setiap pasangan pengamatan saling bebas dengan pasangan pengamatan lainnya, $i = 1, 2, \dots, n$, maka fungsi probabilitas untuk setiap pasangan adalah sebagai berikut:

$$f(x_i) = \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \quad ; y_i = 0, 1 \quad \dots (2.10)$$

dengan,

$$\pi_i = \frac{e^{\sum_{j=0}^p \beta_j x_{ij}}}{1 + e^{\sum_{j=0}^p \beta_j x_{ij}}} \quad \dots (2.11)$$

dimana ketika $j = 0$ maka nilai $x_{0j} = x_{j0} = 1$. Fungsi likelihood untuk distribusi Bernoulli adalah :

$$l(\beta) = \prod_{i=1}^n \pi_i^{y_i} [1 - \pi_i]^{1-y_i}$$

$$= \prod_{i=1}^n \left(\frac{\pi_i}{1 - \pi_i} \right)^{y_i} (1 - \pi_i)$$

dengan persamaan logistik $\pi_i = \frac{e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}}} = \frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}}$

$$1 - \pi_i = 1 - \frac{e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}}} = 1 - \frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}} = (1 + e^{x_i^T \beta})^{-1}$$

Dari persamaan diatas maka diperoleh $\frac{\pi_i}{1 - \pi_i} = e^{x_i^T \beta}$. Sehingga fungsi likelihood dapat diperoleh menjadi :

$$l(\beta) = \prod_{i=1}^n e^{x_i^T \beta y_i} (1 + e^{x_i^T \beta})^{-1}$$

...(2.12)

Setelah fungsi likelihood didapat, langkah selanjutnya yaitu memperoleh nilai log-likelihood yang dapat dinyatakan sebagai berikut :

$$L(\beta) = \ln l(\beta)$$

$$= \sum_{i=1}^n \left\{ y_i \beta^T x_i - \ln(1 + e^{x_i^T \beta}) \right\}$$

...(2.13)

Untuk mendapatkan nilai penaksiran koefisien regresi logistik ($\hat{\beta}$) dilakukan dengan penurunan $L(\beta)$ terhadap β dan disamakan dengan 0. Turunan pertama dari $x_i^T \beta$ terhadap β_j adalah x_{ij} , sehingga penaksir β di hitung menggunakan rumus:

$$\beta^{(t+1)} = \beta^{(t)} + (X^T W^{(t)} X)^{-1} X^T (y - \mu^{(t)})$$

...(2.14)

Kita ulangi proses tersebut sampai dengan konvergen, artinya nilai $\beta^{(t+1)}$ sangat mendekati nilai $\beta^{(t)}$. Demikian juga pada saat konvergen, maka $(X^T W^{(t)} X)^{-1}$ merupakan penaksir matriks kovarians bagi $\hat{\beta}$.

2.6. Model Credit Scoring

Model credit scoring merupakan suatu metode untuk mengevaluasi kelayakan kredit seseorang berdasarkan rumus tertentu atau suatu aturan tertentu. Dalam hal ini, model credit scoring menghasilkan kartu skor yaitu nilai skor setiap kategori di variabel bebas. Teknik ini mengacu pada jangkauan dan format skor dalam kartu skor (Siddiqi, 2006).

Perhitungan skor untuk setiap kategori pada satu variabel bebas, disajikan sebagai berikut :

$$score_i = \sum_{j=1}^p \left(\left(Woe_j * \hat{\beta}_j + \frac{\hat{a}}{p} \right) * factor - \frac{offset}{p} \right)$$

...(2.15)

Dimana :

Woe_j : Nilai *weight of evidence* untuk setiap kategori

$\hat{\beta}_j$: Nilai dugaan koefisien parameter setiap variabel bebas

\hat{a} : Nilai *intercept* dari hasil regresi logistik

p : banyaknya variabel bebas

j : indeks untuk variabel bebas ; $j = 1, 2, \dots, p$

2.7. Ukuran-Ukuran Validasi Model

Beberapa ukuran yang dapat dilakukan untuk validasi yaitu Receiver Operating Characteristic (ROC), Area Under the Curve (AUC), Statistik Kolmogorov-Smirnov (KS), Indeks Gini, dan C-Statistik.

2.7.1 Receiver Operating Characteristic (ROC)

Kurva ROC adalah plot kombinasi nilai sensitivitas dengan nilai 1-spesivitas dengan berbagai *cut off* yang mungkin. Suatu model yang dikatakan baik jika mendekati 100% sebaliknya model yang tidak baik mendekati 50%. Kurva ROC merupakan hasil dari tabel klasifikasi. Untuk memperoleh tabel klasifikasi, kita harus menetapkan *cutpoint c* misalnya berdasarkan desil. Setelah nilai desil diperoleh, kemudian membandingkan nilai desil dengan setiap skor sehingga diperoleh \hat{y}_i .

$$\hat{y}_i = \begin{cases} 1, & \text{iklas}_i \geq c \\ 0, & \text{iklas}_i < c \end{cases}$$

Kemudian nilai-nilai tak bebas y_i yang sebenarnya dirangkum ke dalam tabel kontingensi bersama nilai-nilai prediksinya \hat{y}_i . Hasil tabulasi silang tersebut disebut *confusion matrix*. Bentuk dari *confusion matrix* diperlihatkan pada Tabel 2.1.

Tabel 2.1 Bentuk dari *confusion matrix* untuk *cutpoint c* tertentu (Fawcett, 2003)

		Kelas Sebenarnya	
		Benar	Salah
Kelas Prediksi	Benar	Benar Positif (True Positives (TP))	Salah Positif (False Positives (FP))
	Salah	Salah Negatif (False Negatives (FN))	Benar Negatif (True Negatives (TN))

Jumlah Total Kolom: P N

Beberapa parameter pengukur kinerja ditunjukkan dengan Persamaan (2.16) sampai dengan Persamaan (2.20).

$$FP\ rate = \frac{FP}{N} \quad \dots(2.16)$$

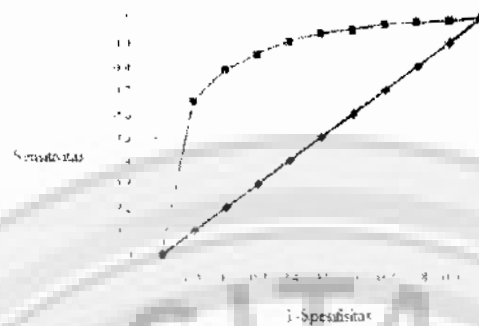
$$TP\ rate = \frac{TP}{p} = Recall \quad \dots(2.17)$$

$$Sensitivitas = Recall \quad \dots(2.18)$$

$$Spesifisitas = \frac{TN}{FP+TN} = 1 - FP\ rate \quad \dots(2.19)$$

$$1 - Spesifisitas = 1 - (1 - FP\ rate) = FP\ rate \quad \dots(2.20)$$

Kurva Receiver Operating Characteristic (ROC) dibentuk berdasar pasangan koordinat 1-spesifisitas dan sensitivitas untuk berbagai nilai *c*, seperti pada gambar 2.3.



Gambar 2.1 Kurva ROC

2.7.2. Area Under the Curve (AUC)

Dari prosedur ROC akan sekaligus mendapatkan nilai Area Under the Curve (AUC) yang letaknya dibawah kurva ROC. Nilai AUC akan mudah didapat dengan menghitung luas daerah dibawah kurva ROC. Untuk suatu kurva ROC yang memadai, maka letaknya selalu berada di daerah sebelah atas dari garis diagonal (0,0) dan (1,1). Nilai AUC 1 merupakan nilai AUC terbaik sehingga model akan semakin baik ketika nilai AUC mendekati 1.

2.7.3. Statistik Kolmogorov-Smirnov

Statistik Kolmogorov-Smirnov (KS) digunakan untuk melihat seberapa besar model untuk memisahkan nasabah baik dan buruk. Nilai KS memuat nilai antara 0 sampai 1. Jika nilai statistik KS mendekati nilai 0 maka model semakin tidak dapat membedakan nasabah baik dan buruk begitupun sebaliknya jika nilai statistik KS mendekati nilai 1 maka model semakin ideal dalam membedakan nasabah baik dan buruk (Rezac, 2011). Asumsikan bahwa untuk setiap nasabah terdapat informasi mengenai nilai skor s dan keterangan sebagai berikut:

$$D_k = \begin{cases} 1, & \text{nasabah yang baik} \\ 0, & \text{sebaliknya} \end{cases}$$

Sehingga KS dapat didefinisikan sebagai:

$$KS = \max_{c \in [L, H]} |F_{m, BAD}(c) - F_{n, GOOD}(c)| \quad \dots(2.21)$$

2.7.4. Indeks Gini

Indeks ini menggambarkan kualitas global dari fungsi skor yang nilainya berkisar antara -1 dan 1 (Rezac, 2011). Model yang ideal memiliki indeks Gini mendekati dengan nilai 1, yaitu fungsi skor yang sempurna memisahkan nasabah yang baik dan buruk. Di sisi lain, model yang memberikan fungsi scoring acak kepada nasabah akan memiliki indeks Gini sama dengan 0. Nilai negatif mempunyai makna fungsi skor yang terbalik yaitu jika skor semakin kecil maka peluang prediksi nasabah yang baiknya semakin tinggi. Indeks Gini dapat dirumuskan sebagai berikut:

$$Gini-1 = \frac{n-m}{\sum_{k=2}^{n+m} (F_{m, BAD}_k - F_{m, BAD}_{k-1}) (F_{n, GOOD}_k - F_{n, GOOD}_{k-1})} \quad \dots(2.22)$$

dimana :

$F_{m, BAD}$, ($F_{n, GOOD}$): nilai vektor ke- k dari fungsi distribusi empirik nasabah buruk (baik).

2.7.5. C-Statistik

Ukuran lain dari kualitas model adalah c-statistik. Nilai c-statistik diantara 0,5 sampai 1, dimana nilai c-statistik sebesar 0,5 menunjukkan nilai kualitas model acak dan 1 menunjukkan nilai kualitas model yang ideal (Rezac, 2011).

Sehubungan dengan indeks Gini, c-statistik (Siddiqi, 2006) dirumuskan sebagai:

$$c\text{-statistik} = \frac{1 - \text{Gini}}{2} \quad \dots(2.23)$$

3. Data dan Hasil

Sumber data yang digunakan merupakan data sekunder nasabah yang terdapat pada data kredit di Jerman yang terdiri 1000 nasabah yaitu 700 nasabah baik dan 300 nasabah buruk. Data kredit yang akan digunakan terdapat 1 variabel tak bebas dan 20 variabel bebas yang terdiri dari 7 variabel bebas dengan skala pengukuran numerik dan 13 lainnya dengan skala pengukuran kategori. Variabel tak bebas dalam penelitian ini adalah status kolektibilitas nasabah berupa status baik ($Y=1$) dan buruk ($Y=0$). Pada tabel 3.1 disajikan contoh data yang memuat variabel respon dan variabel prediktor.

Tabel 3.1 Data Kredit Jerman Tahun 1994

X_1	X_2	X_3	...	X_{18}	X_{19}	X_{20}	Y
A11	6	A34	...	1	A192	A201	0
A12	48	A32	...	1	A191	A201	1
A14	12	A34	...	2	A191	A201	0
A12	45	A34	...	1	A191	A201	0

Sumber: [https://archive.ics.uci.edu/ml/datasets/Statlog+\(German+Credit+Data\)](https://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data))

3.1. Hasil Diskritisasi, Weight of Evidence (WoE) dan Information Value (InV)

Dalam pembuatan model credit scoring, hal yang harus dilakukan terlebih dahulu yaitu melakukan diskritisasi terhadap variabel yang bertipe numerik, melakukan perhitungan WoE dan InV. Dibawah ini merupakan hasil dari diskritisasi, WoE, dan InV yang disajikan dalam tabel 3.2.

Tabel 3.2 Hasil Diskritisasi, WoE, dan InV

Variabel	Kategori	WoE	InV	Tingkat Prediksi	Kesimpulan
X_2	A11	-0,7409	0,5581	Kuat	Masuk ke dalam model
	A12	-0,4080			
	A13	0,4987			
	A14	1,0242			
X_3	$0 < \text{lamanya kredit} \leq 12$	0,3905	0,2226	Medium	Masuk ke dalam model
	$12 < \text{lamanya kredit} \leq 15$	0,8782			
	$15 < \text{lamanya kredit} \leq 24$	-0,0203			
	$24 < \text{lamanya kredit} \leq 30$	-0,0403			
	$\text{lamanya kredit} > 30$	-0,8135			
X_{20}	A201	-0,0367	0,0463	Lemah	Tidak masuk ke dalam model
	A202	1,2659			

Dengan melihat tabel 3.1 dapat disimpulkan bahwa yang masuk kedalam model regresi logistik terdapat 15 variabel bebas dan sisanya yaitu 5 variabel bebas tidak masuk kedalam model regresi logistik yang selanjutnya akan digunakan untuk pembuatan skor beserta validasinya.

3.2. Membentuk Model Credit Scoring

Sebelum melakukan validasi terlebih dahulu membentuk model credit scoring yang berisikan skor untuk setiap kategori dalam suatu variabel bebas. Pembuatan skor memuat nilai taksiran parameter dari model regresi logistik. Di bawah ini hasil taksiran model regresi logistik berdasarkan persamaan 2.14 dengan bantuan software R 3.2.1.

Tabel 3.3 Taksiran Parameter untuk Model Regresi Logistik

Y	Parameter	Taksiran
Kolektibilitas Nasabah	Intercept	0,8104
	X_1	0,7629
	X_2	0,6859

	X_{22}	0,3236
	X_{23}	0,8725

Setelah nilai taksiran parameter diperoleh langkah selanjutnya pembuatan kartu skor berdasarkan persamaan 2.15. Di bawah ini akan dipaparkan hasil kartu skor untuk setiap kategori dalam suatu variabel bebas terpilih yang disajikan dalam tabel 3.4.

Tabel 3.4 Kartu Skor untuk Setiap Kategori

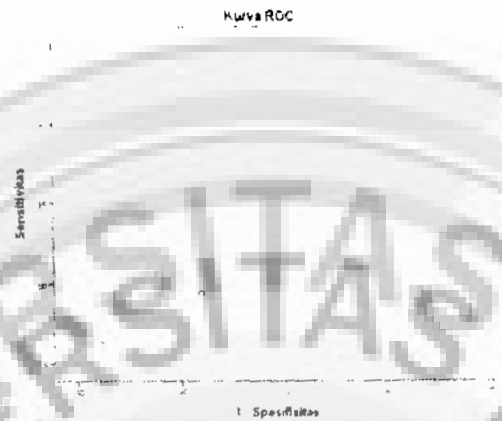
Variabel	Kategori	Skor
X_1	A11	18
	A12	25
	A13	45
	A14	57
X_2	$0 < \text{lamanya kredit} \leq 12$	42
	$12 < \text{lamanya kredit} \leq 15$	51
	$15 < \text{lamanya kredit} \leq 24$	34
	$24 < \text{lamanya kredit} \leq 30$	33
	lamanya kredit > 30	18
...
X_{22}	A201	33

3.1. Validasi Model

Setelah dilakukan pembuatan model credit scoring, perlu dilakukan validasi model terhadap data testing untuk memperkuat ketetapan prediksi dan melihat seberapa baik model credit scoring dalam mengklasifikasikan nasabah yang baik dan buruk. Beberapa ukuran yang dapat dilakukan untuk validasi yaitu Receiver Operating Characteristic (ROC), Area Under the Curve (AUC), Statistik Kolmogorov-Smirnov (KS), Indeks Gini, dan C-Statistik.

- **Receiver Operating Characteristic (ROC)**

Dibawah ini merupakan kurva ROC hasil tabulasi silang atau confusion matrix yang memuat nilai sensitivitas dan 1-spesifisitas.



Gambar 3.1 Kurva ROC Data Kredit Jerman

Pada kurva ROC diatas menunjukkan bahwa kurva mendekati kurava 100% yang dapat diinterpretasikan bahwa model baik atau akurat dalam penyeleksian yang akan menerima kredit.

Tabel 3.5 Hasil Validasi AUC, Statistik K-S, Indeks Gini, dan C-Statistik

Area Under the Curve (AUC)	Statistik Kolmogorov-Smirnov (K-S)	Indeks Gini	C-Statistik
0,5	0,5207	0,6443	0,8321

Pada tabel 3.5 nilai AUC, statistic K-S, indeks Gini, dan c-statistik mendekati nilai 1 yang dapat diinterpretasikan bahwa model baik atau kualitas model ideal dalam memisahkan nasabah baik dan buruk.

4. Kesimpulan

Berdasarkan analisis yang dilakukan terhadap data kredit Jerman tahun 1994 dapat disimpulkan bahwa model credit scoring yang dibuat menggunakan regresi logistik baik atau ideal dan dapat mengklasifikasikan nasabah baik dan buruk, itu terlihat dari hasil validasi model. Dari hasil validasi model yang pertama yaitu Receiver Operating Characteristic (ROC) dapat disimpulkan bahwa model baik atau akurat dalam penyeleksian yang akan menerima kredit dengan berbagai cutt-point berdasarkan desil dan berdasarkan Area Under the Curve (AUC) model mendekati nilai 1 sehingga dapat disimpulkan bahwa model baik dalam penyeleksian yang akan menerima kredit. Selanjutnya ukuran validasi model dengan statistik K-S model hanya dapat memisahkan nasabah baik dan buruk sebesar 0,5171 sehingga model efektif dalam penyeleksian yang akan menerima kredit. Begitupun dengan hasil ukuran validasi model indeks Gini dan c-statistik yang menunjukkan bahwa fungsi scoring yang sempurna dan kualitas model yang ideal sehingga model credit scoring dapat digunakan untuk penyeleksian nasabah yang akan menerima kredit.

Daftar Pustaka

- Ardiati, I. D. 2014. Perbandingan Metode Diskretisasi dalam Model Regresi Logistik (Studi Kasus: Pembentukan Model Penskoran Kredit Bank X) Skripsi. Bogor : Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor.
- Brieman, dkk. 1993. Classification and Regression Trees New York: Champan and Hall.

- Fawcett, T. 2003. ROC Graphs: Notes and Practical Considerations for Data Mining Researchers. HP Laboratories Working Paper, 861-874.
- Hajarisman, Nusar. 2009. Buku Ajar Analisis Data Kategorik. Bandung: Program Studi Statistika Universitas Islam Bandung.
- Hosmer, D. W., dan Lemeshow. 2000. Applied Logistic Regression. New York: John Wiley and Sons.
- Kardiana. dkk. 2006. Metode Klasifikasi Berstruktur Pohon Biner: Kasus Perkiraan Sifat Hujan Bulanan di Bogor. Yogyakarta, 17 Juni 2006. Seminar Nasional Aplikasi Teknologi Informasi (SNATI): G21-G25.
- Kotsiantis S, Kannellpoulos D. 2006. Discretization Technique: A recent survey International Transaction On Computer Science and Engineering, 32 : 47-58.
- Rezac, M. 2011. How to Measure the Quality of Credit Scoring Models. Journal of Economics and Finance, 5 : 486-507.
- Siddiqi, N. 2006. Credit Risk Scorecard Developing and Implementing Intelligent Credit Scoring. New Jersey (US) : John Willey & Sons.
- Sjafitri, Henny. 2011. Faktor-Faktor yang Mempengaruhi Kualitas Kredit dalam Dunia Perbankan. Jurnal Manajemen dan Kewirausahaan, 2: 160-120.
- Suwondono dan Santosa, Stefanus. 2014. Credit Scoring Menggunakan Metode Support Vector Machine dengan Teknik Seleksi Atribut Berbasis Chi Squared Statistic dan Particle Swarm Optimization, 10 : 1-18.
- Suyatno, Thomas., dkk 2010. Dasar-Dasar Perkreditasi Edisi Keempat. Raja Grafindo Persada.

Regresi Logistik pada Data *Rare Event*

¹Rundy Rumi Ari Wistara, ²Suliadi, ³Abdul Kudus

^{1,2,3} Statistika, Fakultas MIPA, Universitas Islam Bandung,

Jl. Ranga Malela No. 1 Bandung 40116

e-mail: ¹rundy_ra@ymail.com, ²suliadi@gmail.com, ³akudus69@ahoo.com

Abstrak. Regresi logistik merupakan salah satu metode statistika yang digunakan untuk menganalisis hubungan beberapa faktor dengan sebuah variabel respon. Pada regresi logistik variabel respon terdiri dari dua kategori yaitu “sukses” dan “gagal” yang dinotasikan dengan $y = 1$ (sukses) dan $y = 0$ (gagal). Regresi logistik baik digunakan jika persentase $y = 0$ dan $y = 1$ tidak jauh berbeda. Dalam kasus kredit macet dapat dijumpai kondisi di mana persentase kredit macet jauh lebih kecil dibandingkan dengan persentase kredit lancar pada variabel respon. Kondisi seperti itu disebut dengan *rare event*. Pada data *rare event* akan menyebabkan $\Pr(Y = 1)$ *underestimates* sedangkan untuk $\Pr(Y = 0)$ *overestimates*. Masalah sampel terbatas (*finite sample*) dapat menyebabkan (i) model yang terbentuk akan menghasilkan penaksir parameter yang bias; (ii) kesalahan baku bagi penaksir yang lebih kecil (*underestimates*) dan (iii) dapat menyebabkan $\Pr(Y = 1)$ *underestimates*. Skripsi ini membahas bagaimana mengoreksi penduga parameter yang bias dan koreksi peluang pada regresi logistik jika data respon jarang terjadi (*rare event*). Model regresi logistik *rare event* akan diterapkan pada data kasus kredit bank di Amerika dengan $Y = 1$ jika nasabah mengalami kredit macet lebih dari 90 hari. Hasil koreksi bias terhadap koefisien regresi adalah bahwa bias pada $\hat{\beta}_0$ lebih besar di bandingkan dengan yang lainnya. Namun untuk hasil bias pada keseluruhan taksiran parameter kecil, hal ini karena sampel yang cukup besar yaitu sebanyak 12013. Hasil kesalahan baku penaksir terlihat bahwa kesalahan baku penaksir parameter terkoreksi lebih kecil dibandingkan dengan kesalahan baku pada penaksir parameter regresi logistik. Oleh karena itu regresi logistik pada data *rare event* lebih baik digunakan daripada regresi logistik. Sedangkan untuk koreksi taksiran peluang ($\tilde{\pi}$) lebih kecil dibandingkan dengan taksiran peluang jika tidak menggunakan koreksi pada regresi logistik.

Kata kunci: Regresi Logistik, *Rare Event*, Bias, Peluang.

A. Pendahuluan

Dalam kehidupan sehari-hari semua orang pasti memiliki kebutuhan. Kebutuhan ada yang bersifat mendesak dan ada yang tidak. Kebutuhan yang mendesak menuntut untuk segera dipenuhi. Namun pemenuhan tersebut tidak terlepas dari masalah biaya atau dana. Dana yang diperlukan biasanya tidak sedikit jumlahnya, sementara dana yang tersedia acapkali tidak mencukupi.

Kebanyakan orang dalam menghadapi kekurangan dana salah satu jalan keluar yang dapat dilakukan adalah dengan berutang kepada pihak bank. Para nasabah yang telah memperoleh fasilitas kredit dari bank tidak seluruhnya dapat mengembalikan utangnya dengan lancar sesuai dengan waktu yang telah diperjanjikan. Akibat nasabah tidak dapat membayar lunas utangnya, maka akan tergambar perjalanan kredit menjadi macet atau terhenti.

Salah satu metode yang dapat dipergunakan untuk memetakan nasabah ke dalam kategori kredit macet dan lancar yaitu metode regresi logistik. Regresi logistik merupakan salah satu metode statistika yang digunakan untuk menganalisis hubungan antara satu variabel respon (Y) dengan satu atau lebih variabel bebas (X_j). Dimana variabel respon terdiri dari dua kategori yaitu “sukses” dan “gagal” yang dinotasikan dengan $Y = 1$ (sukses) dan $Y = 0$ (gagal). Sebagai contoh pada kasus kartu kredit, $Y = 0$

jika variabel responnya menyatakan kredit lancar dan $Y = 1$ jika variabel responnya menyatakan kredit macet.

Regresi logistik baik digunakan jika persentase $Y = 0$ dan $Y = 1$ tidak jauh berbeda. Dalam kasus kredit macet dapat dijumpai kondisi di mana persentase kredit macet jauh lebih kecil dibandingkan dengan persentase kredit lancar pada variabel respon. Kondisi seperti itu disebut dengan *rare event*. Pada data *rare event* akan menyebabkan $\Pr(Y = 1)$ *underestimates* sedangkan untuk $\Pr(Y = 0)$ *overestimates*. Masalah sampel terbatas (*finite sample*) dapat menyebabkan (i) model yang terbentuk akan menghasilkan penaksir parameter yang bias; (ii) kesalahan baku bagi penaksir yang lebih kecil (*underestimates*) dan (iii) dapat menyebabkan $P(Y = 1)$ *underestimates*. Skripsi ini membahas bagaimana mengoreksi penduga parameter yang bias dan koreksi peluang pada regresi logistik jika data respon *rare event* dan diaplikasikan pada kasus kredit macet.

1. Rumusan Masalah

Berdasarkan uraian dari latar belakang yang telah diungkapkan, maka masalah yang dapat diidentifikasi adalah:

1. Bagaimana perbandingan penaksir parameter regresi logistik dengan regresi logistik pada data *rare event* ?
2. Bagaimana perbandingan kesalahan baku penaksir parameter regresi logistik dengan regresi logistik pada data *rare event* ?
3. Bagaimana perbandingan taksiran peluang regresi logistik dengan regresi logistik pada data *rare event* ?

2. Tujuan Penelitian

Berdasarkan identifikasi masalah maka tujuan dalam penulisan skripsi ini adalah:

1. Membandingkan penaksir parameter regresi logistik dengan regresi logistik pada data *rare event*.
2. Membandingkan kesalahan baku penaksir parameter regresi logistik dengan regresi logistik pada data *rare event*.
3. Membandingkan taksiran peluang regresi logistik dengan regresi logistik pada data *rare event*.

B. Tinjauan Pustaka

1. Regresi Logistik

Menurut Hosmer dan Lemeshow (1989) model regresi logistik yang dipengaruhi oleh k variabel bebas dapat dinyatakan sebagai nilai harapan dari Y dengan diberikan nilai x .

$$E(Y | x) = \pi_i = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)} \quad (2.1)$$

Untuk mempermudah dalam menaksir parameter regresi, maka π_i pada persamaan (2.1) ditransformasikan dengan menggunakan transformasi logit. Sehingga dapat ditulis sebagai berikut:

$$\text{logit}(\pi_i) = g(x) = \ln \left[\frac{\pi_i}{1 - \pi_i} \right] = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (2.2)$$

2. Penaksiran Parameter Model

Metode penaksiran parameter yang biasa digunakan dalam regresi logistik adalah metode MLE (*Maximum Likelihood Estimation*). Variabel respon Y memiliki sebaran Bernoulli dengan parameter π_i dan fungsi sebaran peluangnya adalah:

$$P(y_i | x_i) = \begin{cases} \pi_i^{y_i} [1 - \pi_i]^{1-y_i}, & \text{untuk } y_i = 0 \text{ atau } 1 \\ 0 & \text{, untuk } y_i, \text{ yang lain} \end{cases}$$

Menurut Hosmer dan Lemeshow (1989), fungsi *likelihood* distribusi Bernoulli untuk n sampel bebas adalah

$$l(\beta) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \quad (2.3)$$

Untuk memudahkan mencari nilai $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ yang memaksimumkan fungsi *likelihood* digunakan bentuk logaritma natural dari fungsi *likelihood*, yang disebut sebagai fungsi *log-likelihood*. Logaritma natural fungsi peluang bersamanya dapat ditulis sebagai berikut:

$$L(\beta) = \ln l(\beta) = \sum_{i=1}^n \{y_i (\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}) + \ln[1 - \pi_i]\} \quad (2.4)$$

Selanjutnya dihitung turunan pertama dari $L(\beta)$ masing-masing terhadap $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ kemudian disyaratkan sama dengan nol.

$$\begin{aligned} \frac{dL(\beta)}{d\beta_0} &= \frac{d}{d\beta_0} \sum_{i=1}^n \{y_i (\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}) + \ln[1 - \pi_i(\beta)]\} \\ &= \sum_{i=1}^n \{y_i - \pi_i\} \end{aligned} \quad (2.5)$$

$$\begin{aligned} \frac{dL(\beta)}{d\beta_1} &= \frac{d}{d\beta_1} \sum_{i=1}^n \{y_i (\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}) + \ln[1 - \pi_i(\beta)]\} \\ &= \sum_{i=1}^n x_{i1} [y_i - \pi_i] = 0 \end{aligned} \quad (2.6)$$

$$\begin{aligned} \frac{dL(\beta)}{d\beta_p} &= \frac{d}{d\beta_p} \sum_{i=1}^n \{y_i (\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}) + \ln[1 - \pi_i(\beta)]\} \\ &= \sum_{i=1}^n x_{ik} [y_i - \pi_i] = 0 \end{aligned} \quad (2.7)$$

Dari persamaan (2.7), (2.8), dan (2.9) masih terkandung y_i , dari turunan pertama di atas sulit untuk dihitung secara manual oleh sebab itu digunakan bantuan *software*.

Selanjutnya akan dihitung turunan kedua, turunan kedua ini akan dilihat apakah ada solusi atau tidak.

Bentuk umum dari turunan parsial kedua fungsi *log-likelihood* adalah:

$$\frac{\delta^2 L(\beta)}{\delta \beta_k^2} = -\sum_{i=1}^n x_{ki}^2 \pi_i (1 - \pi_i) < 0$$

$$\frac{\delta^2 L(\beta)}{\delta \beta_i \delta \beta_k} = -\sum_{i=1}^n x_{ki} x_{ki} \pi_i (1 - \pi_i) < 0$$

dimana $i, j = 0, 1, 2, \dots, k$. Dan penaksir matriks variansnya adalah

$$V(\hat{\beta}) = \left[\sum_{i=1}^n \pi_i (1 - \pi_i) x_i x_i' \right]^{-1} \quad (2.8)$$

dimana π_i adalah peluang sukses, $1 - \pi_i$ adalah peluang gagal dan x_i adalah variabel bebas dengan $i = 1, 2, \dots, k$.

3. Regresi Logistik pada Data *Rare Event*

Misalkan variabel respon $Y_1, Y_2, \dots, Y_i, \dots, Y_n$ merupakan sampel acak yang berdistribusi Bernoulli dengan $\pi_i = P(Y = 1)$ dan $1 - \pi_i = P(Y = 0)$ untuk $i = 1, 2, \dots, n$. Dalam model regresi logistik peluang π_i adalah fungsi distribusi kumulatif logistik padapersamaan 2.1. Transformasi logit sebagaimana dijelaskan pada bagian regresi logistik yaitu persamaan 2.2.

1) Koreksi Bias terhadap Koefisien

Untuk mengoreksi bias $\hat{\beta}$ dapat ditaksir oleh *weighted least-squared*:

$$bias(\hat{\beta}) = (X'WX)^{-1} X'W\xi \quad (2.9)$$

dimana

$$\xi_i = 0.5 Q_{ii} ((1 + \bar{w}_i) \hat{\pi}_i - \bar{w}_i) \quad (2.10)$$

$$Q_{ii} = \text{elemen diagonal utama} \{X(X'WX)^{-1} X'\} \quad (2.11)$$

$$W = \text{diag}\{\hat{\pi}_i (1 - \hat{\pi}_i) w_i\} \quad (2.12)$$

Dengan vektor pembobot w_i sebagai berikut,

$$w_i = \bar{w}_1 Y_i + \bar{w}_0 (1 - Y_i) \quad (2.13)$$

dimana $\bar{w}_1 = \frac{\tau}{\bar{y}}$ sebagai pembobot untuk nilai satu dan $\bar{w}_0 = \frac{1 - \tau}{1 - \bar{y}}$ sebagai

pembobot nilai nol. Sedangkan τ adalah proporsi kejadian sukses dalam populasi dan \bar{y} adalah proporsi kejadian sukses dalam sampel. Metode WLS pada regresi logistik *rare event* mudah untuk diterapkan karena komponennya sama dengan metode WLS pada regresi logistik. Dengan ξ sebagai variabel respon, X sebagai variabel bebas dan W sebagai pembobot. Sedangkan untuk penaksir koreksi biasnya yaitu,

$$\tilde{\beta} = \hat{\beta} - bias(\hat{\beta}) \quad (2.14)$$

Untuk mendapatkan prediksi peluang maka bisa dilakukan dengan memasukkan koreksi penaksir bias ($\tilde{\beta}$) ke dalam persamaan logit sebagai berikut:

$$\tilde{\pi} = \Pr(\hat{y}_i = 1 | \tilde{\beta}) = \frac{\exp(x_i \tilde{\beta})}{1 + \exp(x_i \tilde{\beta})} \quad (2.15)$$

Namun, hal ini tidak optimal karena mengabaikan ketidakpastian pada $\tilde{\beta}$. Oleh karena itu perlu dilakukan koreksi ulang terhadap $\tilde{\pi}_i$. Bentuk koreksi peluangnya sebagai berikut:

$$\Pr(y_i = 1) \approx \tilde{\pi}_i + C_i \quad (2.16)$$

dengan faktor koreksinya adalah

$$C_i = (0.5 - \tilde{\pi}_i) \tilde{\pi}_i (1 - \tilde{\pi}_i)^{-1} V(\tilde{\beta}) x_i \quad (2.17)$$

dimana

$$V(\tilde{\beta}) = \left(\frac{n}{n+k} \right)^2 V(\hat{\beta})$$

dan matriks varians $V(\hat{\beta})$ sebagaimana dijelaskan pada bagian penaksiran parameter model regresi logistik.

C. Bahan dan Metode Penelitian

1. Bahan

Data yang digunakan untuk mengaplikasikan analisis regresi logistik *rare event* ini berasal dari "Give me some credit" yang diluncurkan dalam situs Kaggle. Data tersebut berisi tentang nasabah yang memiliki fasilitas kredit (debitur). Variabel respon yang digunakan adalah status kredit yaitu:

$$y_i = \begin{cases} 0, & \text{tidak mengalami kredit macet lebih dari 90 hari} \\ 1, & \text{mengalami kredit macet lebih dari 90 hari} \end{cases}$$

Sedangkan variabel bebas ada sebanyak 10 variabel yang terdiri dari Jumlah Saldo Kartu Kredit, Usia Debitur, Frekuensi Mengalami Kredit Macet 30-59 Hari, Pembayaran Utang Bulanan, Pendapatan Bulanan, Jumlah Pinjaman Terbuka dan Kredit, Frekuensi Mengalami Kredit Macet >90 Hari, Jumlah Kredit KPR dan Properti, Frekuensi Mengalami Kredit Macet 60-89 Hari, Jumlah Tanggungan Keluarga.

2. Metode

Metode dan tahap-tahap penelitian yang dilakukan untuk mencapai tujuan penulisan adalah sebagai berikut:

1. Melakukan penaksiran koefisien parameter regresi logistik.
2. Melakukan penaksiran model regresi logistik pada data *rare event* dengan langkah-langkah sebagai berikut:
 - a) Menghitung nilai proporsi kejadian sukses dalam populasi (τ).
 - b) Menghitung nilai proporsi kejadian sukses dalam sampel (\bar{y}).
 - c) Masukkan langkah 1 dan 2 ke dalam persamaan (2.15) lalu hitung vektor pembobot (w_i).
 - d) Menentukan vektor pembobot W pada persamaan (2.14).
 - e) Kemudian tentukan nilai Q_{ii} pada persamaan (2.13).
 - f) Lalu tentukan vektor dari ξ_i pada persamaan (2.12).
 - g) Menghitung $bias(\hat{\beta})$ dengan menggunakan persamaan (2.11).
 - h) Menghitung penaksir terkoreksi ($\tilde{\beta}$) pada persamaan (2.16).

Langkah-langkah di atas dilakukan dengan menggunakan perintah 'relogit' pada *package* 'Zelig' software R.

3. Mengoreksi kesalahan baku bagi penaksir pada regresi logistik pada persamaan (2.20).
4. Koreksi terhadap $P(Y = 1)$ model regresi logistik pada persamaan (2.18) dengan langkah-langkah sebagai berikut:
 - a) Menghitung prediksi peluang dengan memasukkan penaksir terkoreksi ke dalam persamaan (2.17).
 - b) Menghitung faktor koreksi (C_i) pada persamaan (2.19)

Langkah 4.a dan 4.b dilakukan dengan menggunakan *software* SAS IML.

D. Pembahasan

Penaksiran parameter untuk model regresi logistik dilakukan dengan menggunakan metode *maximum likelihood*. Variabel respon yang digunakan adalah status kredit yaitu:

$$y_i = \begin{cases} 0, & \text{tidak mengalami kredit macet lebih dari 90 hari} \\ 1, & \text{mengalami kredit macet lebih dari 90 hari} \end{cases}$$

dimana banyaknya nilai $Y = 1$ yaitu sebesar 827 dan banyaknya nilai $Y = 0$ sebesar 11186. Namun demikian data yang mengandung fenomena *rare event* dimana persentase $Y = 1$ hanya sebesar $\frac{827}{12013} = 6,88\%$. Hal tersebut akan mengakibatkan

underestimated pada $P(Y = 1)$ yang artinya terdapat bias pada penaksir parameter. Dengan demikian harus dilakukan koreksi terhadap koefisien parameter regresi logistik. Bias pada penaksir parameter kecil, hal ini dikarenakan sampel yang cukup besar yaitu sebanyak 12013. Sedangkan bias pada $\hat{\beta}_0$ lebih besar di bandingkan dengan yang lainnya.

Kesalahan baku pada penaksir parameter sangat penting digunakan dalam suatu analisis salah satunya pengujian hipotesis. Kesalahan baku bagi penaksir parameter terkoreksi lebih kecil dibandingkan dengan kesalahan baku bagi penaksir parameter regresi logistik. Ketika data *rare event* koreksi terhadap kesalahan baku lebih baik digunakan daripada regresi logistik.

Koreksi peluang pada data *rare event* dilakukan dengan menggunakan penaksir terkoreksi untuk mendapatkan prediksi peluang $\tilde{\pi}_i$. Koreksi peluang ($\tilde{\pi}$) lebih kecil dibandingkan dengan taksiran peluang jika tidak menggunakan koreksi pada regresi logistik.

E. Kesimpulan

Kesimpulan dari skripsi ini adalah:

1. Hasil koreksi bias terhadap koefisien regresi adalah bahwa bias pada $\hat{\beta}_0$ lebih besar di bandingkan dengan yang lainnya. Namun untuk hasil bias pada keseluruhan taksiran parameter kecil, hal ini karena sampel yang cukup besar yaitu sebanyak 12013.
2. Hasil kesalahan baku penaksir terlihat bahwa kesalahan baku penaksir parameter terkoreksi lebih kecil dibandingkan dengan kesalahan baku pada penaksir parameter regresi logistik. Oleh karena itu regresi logistik pada data *rare event* lebih baik digunakan daripada regresi logistik.

3. Sedangkan untuk koreksi taksiran peluang ($\tilde{\pi}$) lebih kecil dibandingkan dengan taksiran peluang jika tidak menggunakan koreksi pada regresi logistik.

DAFTAR PUSTAKA

- Agresti, Alan. 2002. *Categorical Data Analysis*. New York: Inc. John Wiley and Sons.
- Bahsan, M. 2012. *Hukum Jaminan dan Jaminan Kredit Perbankan Indonesia*. Jakarta: PT. Raja Grafindo Persada.
- D.W. Hosmer, dan S. Lemeshow. 1989. *Applied Logistic Regression*. New York: Inc. John Wiley and Sons.
- Hajarisman. N. 2009. *Analisis Data Kategorik*. Bandung: Tidak diterbitkan.
- King Gary, dan Zeng Langche. 2001. *Logistic Regression in Rare Events Data*. <http://gking.harvard.edu/files/0s.pdf>.
- Nasution, Aatiqah. 2012. *Regresi Logistik untuk Menentukan Peluang yang Mempunyai Kartu Yogya dan yang Tidak Mempunyai Kartu Yogya Berdasarkan Kepuasan Konsumen*. Bandung: tidak diterbitkan.
- Supramono, Gatot. 2009. *Perbankan dan Masalah Kredit – Suatu Tinjauan di Bidang Yuridis*. Jakarta: PT. Rineka Cipta.

Reducing Bias for Logistic Regression with Rare Event Data

Suliadi Suliadi^{1,a)}, Abdul Kudus^{1,b)} and Teti Sofia Yanti^{1,c)}

¹Dept. of Statistics - Bandung Islamic University
Jl. Ranggamalela No. 1 Bandung (40116)
West Java - Indonesia

^{a)}Corresponding author: suliadi@gmail.com, suliadi@unisba.ac.id

^{b)}akudus69@yahoo.com

^{c)}buitet@yahoo.com

Abstract. It is well known that maximum likelihood estimation (MLE) for logistics regression results good estimate of the regression coefficients as long as the number of success and fail are not largely different. In the case of rare event data, where the proportion of success and fail are large different, then MLE results biased estimate on the regression coefficients and the probability of success as well. To reduce this bias, we may combine under sampling approach and bootstrap procedure. We implement this method to real data and obtain the results as follows. When the ratio of number of success and number of fail event tends to one, then the estimates tend to unbiased but the variances increase and vice verse. This result is promising for the next research in obtaining method that can reduce bias in the case of rare event data.

Keywords: Resampling, Bootstrap like, Bias Estimator, Rare Event, Logistic Regression

INTRODUCTION

Logistic regression is commonly used to model binary data. This method uses maximum likelihood (MLE) to estimate the regression parameter. The property of the estimate of regression parameters for logistics regression by using MLE is unbiased and has minimum variance. However, this unbiased property holds if the proportion of success ($Y=1$) and proportion fail ($y=0$) has no large different (McCullagh & Nelder, 1989, Chap 4 & 15). If the proportion of those two categories has large different then the estimate of regression coefficients will be biased, including the estimate of the probability of success and fail (King & Zeng, 2001; Qiu, et al, 2013). This case is known as rare event problem, where the proportion of success (fail) less than 10% from all data. King & Zeng (2001) and Qiu, et al (2013) gave several cases in rare event: war, landslides, fraudulent of credit cards, international conflict, oil spill etc.

Several procedures have been proposed to correct the bias. McCullagh & Nelder (1989) proposed to correct the estimate of regression coefficients, whilst King & Zeng (2001) proposed to correct the intercept coefficient. Qiu, et al (2013) used different approach, by using weighted maximum likelihood that employed to the reconstructed data and combine with correction of McCullagh & Nelders and King & Zengs methods.

Discriminant analysis also suffers from rare event data. Several methods have been proposed to handle this problem. It commonly uses resampling method to balance the proportion of ($Y=1$) and ($Y=0$), such as under sampling and over sampling method. The principle of these methods is balancing the number of fail ($Y=0$) and success ($Y=1$).

Several methods/algorithms have been proposed to reduce bias by balancing the number of success and fail events that known as under sampling, over sampling and others type sampling (Kosmidis & Firth, 2010; Abd Elrahman & Abraham, 2013; Satyasree & Muthy, 2013; Weiss, 2004; Yap, et al., 2014). Basic principle of these method is sampling design, i.e balancing the number of fail ($Y=0$) and success ($Y=1$). Under sampling method balances the data by keeping data of the minority class and taking some observations from the majority class randomly, such that, both minority and majority classes have the desired proportion. Meanwhile, over sampling method balances the data by keeping data of the majority class and replicating data of the minority class. Both methods have severe drawback, under sampling approach will lose some information, due to removing some observations from majority class, whilst over sampling approach has redundant data and increases the size of data without increase the information.

In this paper we use different approach to overcome the MLE's drawback when applied to the case of rare event data. We use bootstrap method that applied to sampling design approach (Kosmidis & Firth, 2010; Abd Elrahman & Abraham, 2013; Satyasree & Muthy, 2013; Weiss, 2004; Yap, et al., 2014) and estimate the model using ordinary MLE. Hence we may use any statistical software that commonly provide logistic regression analysis.

The outline of this paper is as follows. Section 2 gives a brief steps in the estimation of logistic regression parameters, and followed by Section 3 that discusses the method of bootstrap procedure. Section 4 discusses the characteristics of estimates based on bootstrap and the conclusion and suggestion is given in Section 5.

LOGISTIC REGRESSION

Suppose there are n subject and y_i be a binary response for the i -th subject where the $y_i \in \{0, 1\}$, for $i = 1, 2, \dots, n$. Response $y_i = 1$ if the i -th subject has a specific characteristics (success event) and $y_i = 0$ if the i -th subject has no that characteristics (fail event). For each subject is also measured k -dimensional vector of covariates x_i that affects the response. The response y_i has follow Bernoulli distribution with parameter π_i and $E(y_i) = P(y_i = 1) = \pi_i$. The probability mass function (pmf) of Y_i random variable is $f(y_i|\pi_i) = \pi_i^{y_i}(1 - \pi_i)^{1-y_i}$. This pmf can be written in the form of exponential family distribution as

$$f(y, \pi) = \exp \left\{ y \log \left(\frac{\pi}{1 - \pi} \right) - \log \left(1 + \frac{\pi}{1 - \pi} \right) + \log \left(\frac{1}{y} \right) \right\}$$

where $\theta = \log \left\{ \frac{\pi}{1 - \pi} \right\}$; $a(\phi) = 1$; $b(\theta) = \log \left(1 + \frac{\pi}{1 - \pi} \right) = \log(1 + \exp(\theta))$; and $c(y, \phi) = \log \left(\frac{1}{y} \right)$. Parameter θ is known as canonical parameter.

Suppose the response y_i is related the covariates through a link function

$$g(\pi) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k.$$

Logistic regression takes canonical parameter as the link function. It will take

$$\pi_i = P(Y_i = 1) = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} = \frac{\exp(\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki})}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki})}$$

Let $y = (y_1, y_2, \dots, y_n)^T$; $\pi = (\pi_1, \pi_2, \dots, \pi_n)^T$; $\eta = (\eta_1, \eta_2, \dots, \eta_n)^T$; and $X = (x_1, x_2, \dots, x_n)^T$. The ML estimator for $\beta = (\beta_0, \beta_1, \dots, \beta_k)^T$ is the solution of estimating equation

$$U(\beta) = X^T W_1 (y - \pi) = 0$$

where $W_1 = \text{Diag} \left[\frac{1}{\text{Var}(y_i)} \left[\frac{\partial \pi_i}{\partial \eta_i} \right] \right]$. For canonical link function we have $\frac{\partial \pi_i}{\partial \eta_i} = \text{Var}(y_i)$. The iterative procedure by using Fisher Scoring algorithm for β is given by

$$\hat{\beta}^{(m+1)} = \hat{\beta}^m + I^{-1} U$$

where

$$\begin{aligned} I &= E \left[\frac{\partial l}{\partial \beta} \frac{\partial l}{\partial \beta} \right] = X^T E_1 E(y - \pi)(y - \pi)^T W_1 X \\ &= X^T W X \end{aligned}$$

with

$$W = W_1 \text{Diag}(\text{Var}(y_i)) W_1 = \text{Diag} \left\{ \frac{1}{\text{Var}(y_i)} \left(\frac{\partial \pi_i}{\partial \eta_i} \right)^2 \right\}$$

BOOTSTRAPING IN LOGISTICS REGRESSION WITH RARE EVENT DATA

In order to overcome bias problem in rare event data, we suggest bootstrap method combined with resampling approach (King & Zeng, 2001; Qiu, et al, 2013), specifically under sampling approach. In this frame work, all observations which are $Y=1$ (rare event or minority class) will be included to bootstrap data and combined with observations obtained from resampling some observations of $Y=0$ (majority class). Number observation of resampling data from majority class depends on p , the proportion the drawn data from of $Y=0$. This bootstrap data are applied to ordinary logistic regression. Without loss of generality, we assume that event $Y=1$ is the rare event or minority class, whilst event $Y=0$ is the majority class.

The idea of this procedure is motivated from the fact that bias of the estimate is caused by the unbalanced proportion of $Y=0$ and $Y=1$. Thus, by taking only several part of data with $Y = 0$ in estimation, we may balance the proportion of both classes such that the ideal condition of ordinary logistic regression achieved. All observations of the minority class will be included to the bootstrap data, since each observation of this class is very valuable, in which it has very large amount of information compared observation to the other group (King & Zeng, 2001). Hence to balance the proportion of both groups, we may re-sample only from non-rare event data. In order to overcome the lost information due to removing several observations form majority class, this process will be repeated several times, e.g 1000 times. Thus this approach is not really bootstrap method, since it does not re-sample observation from both classes.

The steps in estimating the regression parameters based on bootstrap approach is as follows.

1. Take all observations where it has $Y=1$ (rare event)
2. Draw m observations randomly from group of observations where $Y=0$, such that the proportion the drawn data from $Y=0$ is p .
3. Combine data from step a) and b) and do the ordinary logistic regression for this data
4. Repeat steps 1 - 3 above k times.

PERFORMANCE OF BOOTSTRAP METHOD

Bootstrap's Scenario

To evaluate the characteristics of estimate of the bootstrap method, we applied steps in Section 3 above using data of bad loan (credit), obtained from www.kaggle.com/c/GiveMeSomeCredit. After data cleaning, it consists of 117,971 observations with 109,794 (93.07%) of observations with $Y=0$ and the rest 8,177 (6.93%) observations with $Y=1$. This shows that the rare event problem exists.

We run five scenarios as follows.

1. Scenario I (G01) where $p = 8,177/109,794$. This means we draw randomly from majority class ($Y=0$) about $p \times 109,794 \approx 8,177$ observations. In this scenario, the ratio of observation $Y = 0$ and $Y = 1$ for bootstrap is 1 : 1 (approximate) and proportion of ($Y = 1$) is $p^* \approx 50\%$.
2. Scenario II (G02) where $p = 16,000/109,794$. In this scenario, the ratio of observation $Y = 0$ and $Y = 1$ for bootstrap is 2 : 1 (approximate) and the proportion of ($Y = 1$) is $p^* \approx 33.33\%$.
3. Scenario III (G03) where $p = 25,000/109,794$. The ratio of observation $Y = 0$ and $Y = 1$ for bootstrap is 3 : 1 (approximate) and the proportion of ($Y = 1$) is $p^* \approx 25\%$.
4. Scenario IV (G04) where $p = 33,000/109,794$. The ratio of observation $Y = 0$ and $Y = 1$ for bootstrap is 4 : 1 (approximate) and the proportion of ($Y = 1$) is $p^* \approx 20\%$.
5. Scenario V (G05) where $p = 41,000/109,794$. The ratio of observation $Y = 0$ and $Y = 1$ for bootstrap is 5:1 (approximate) and the proportion ($Y = 1$) is $p^* \approx 0.1667$.

For each scenario was run 250 times. As a note that all scenarios above are run using algorithm as given in section above.

TABLE 1. The Average, Median, and Standard Deviation of Bootstrap Estimate for Each Coefficient Estimate

Scenario	b0	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10
Average											
G01	-0.909	1.881	-0.018	0.563	0.0023	-0.000040	0.038	0.998	0.133	0.931	0.0397
G02	-1.482	1.797	-0.018	0.528	-0.0004	-0.000043	0.036	0.920	0.131	0.836	0.0463
G03	-1.852	1.741	-0.017	0.506	-0.0023	-0.000044	0.034	0.859	0.129	0.783	0.0497
G04	-2.058	1.680	-0.017	0.493	-0.0032	-0.000046	0.032	0.826	0.129	0.749	0.0520
G05	-2.224	1.640	-0.017	0.485	-0.0038	-0.000047	0.031	0.799	0.130	0.721	0.0541
OLR	-2.867	1.331	-0.018	0.461	-0.0060	-0.000051	0.022	0.681	0.127	0.626	0.0632
Median											
G01	-0.930	1.923	-0.018	0.564	0.0019	-0.000041	0.039	0.996	0.133	0.929	0.0403
G02	-1.499	1.812	-0.018	0.528	-0.0002	-0.000043	0.036	0.920	0.132	0.834	0.0463
G03	-1.849	1.748	-0.017	0.507	-0.0024	-0.000045	0.034	0.860	0.130	0.785	0.0492
G04	-2.060	1.692	-0.017	0.493	-0.0033	-0.000046	0.032	0.827	0.129	0.749	0.0523
G05	-2.221	1.644	-0.017	0.485	-0.0038	-0.000047	0.031	0.801	0.130	0.722	0.0542
Standard Deviation											
G01	10.166	1.233	10.131	2.482	5.391	3.0	3.011	7.315	12.421	6.477	10.369
G02	9.905	1.197	8.105	2.031	3.489	2.4	2.429	5.198	9.591	4.972	8.806
G03	9.57	1.15	7.323	1.667	3.182	2.0	2.283	3.69	7.689	3.81	6.473
G04	10.02	1.168	7.213	1.382	2.437	1.7	2.301	3.446	6.411	3.36	5.835
G05	9.733	1.159	5.848	1.16	2.536	1.6	2.266	3.109	5.918	2.756	4.844
	(x100)	(x10)	(x10000)	(x100)	(x1000)	(x10-6)	(x1000)	(x100)	(x1000)	(x100)	(x1000)

Bootstrap's Results

Results for this bootstrap method are given in Table 1 and depicted in Figure 1 that also give the result of ordinary logistics regression. This table shows that for each scenario, there is no large different between average and median for each coefficient estimate. This means that average of b_i and its median has the same location. But when we compare average (median) among scenarios, the average and median tends to increase or decrease consistently (see also Figure 1). If the proportion of rare event is getting smaller (from scenario G01 to G05), then the average of b_i is getting further from the average of b_i from scenario G01. As a note that Scenario G01 is an ideal condition since the proportion of $Y = 1$ is approximately 0.5 and G05 is the worst condition. The value of median also has the same pattern as average. Since condition of Scenario G01 is the ideal condition, the distance widening from the average of b_i of scenario G01 also indicates the bias pattern or behavior when the proportion or $n(Y = 0)$ and $n(Y = 1)$ becomes more unbalance: decreasing proportion of rare event ($Y = 1$) gives more bias.

The estimates of ordinary logistic regression (OLR) coefficients based on ordinary method using all data are also given in Table 1. Except for b_2 , all coefficients have the same pattern where the estimate of OLR's estimates are closest to bootstrap estimate of Scenario G05. This result is as expected, since the most unbalance scenario among G01-G05 is the G05. As we know that this OLR gives biased estimate (McCullagh & Nelder (1989); King & Zeng, 2001; Qiu, et al, 2013; and others). Since the OLR's results are very close to bootstrap approach scenario G05, thus it can be deduced that if rare event problem exists then bootstrap approach may reduce the bias estimate by taking scenario G01.

Figure 2. gives the pattern of variances (standard deviation) for each bootstrap estimate based on 250 replication for each regression parameters estimate. Except b_0 , all estimates have the same pattern of variance. They tend to decrease if the proportion of rare event tends to decrease too. Only b_0 has not the same pattern as others, but generally, the variance of b_0 decrease as the proportion of rare event does. This pattern can be understood, since as discussion above, the bias tends to increase. These results exhibit that the mean square error (MSE) remains steady whatever the proportion of rare event. If the proportion of rare event exists, it results bias of the estimate. But it is compensated by decreasing of their variances. Another explanation is that from G01 to G05 the number of observation (data) used increase as well as the number of observation of majority class, and it implies reducing variance of the estimates.

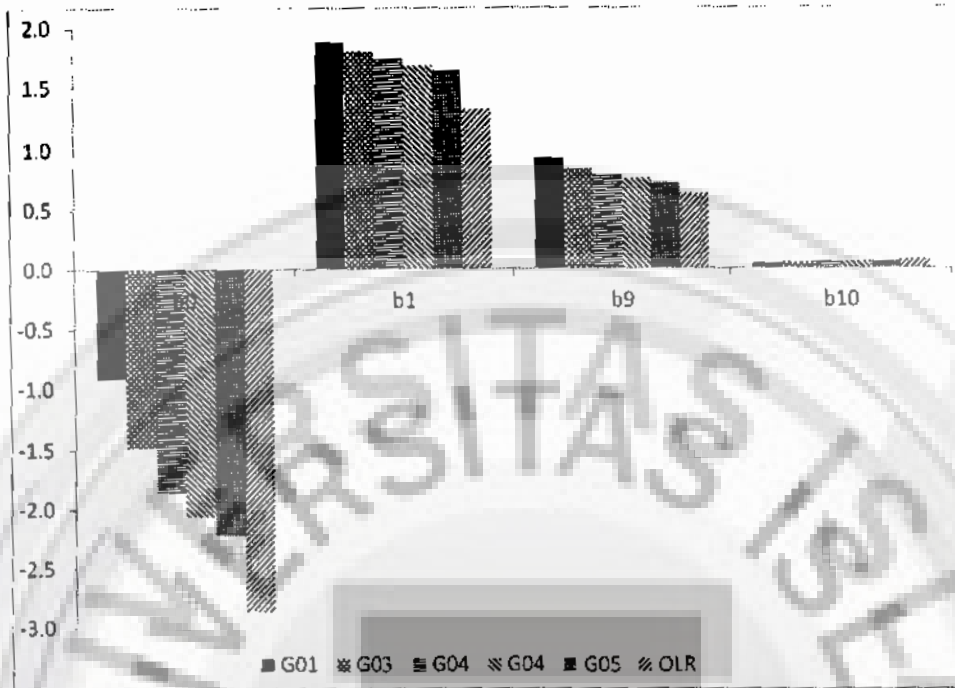


FIGURE 1. The Average of b_0 , b_1 , b_9 , and b_{10} .

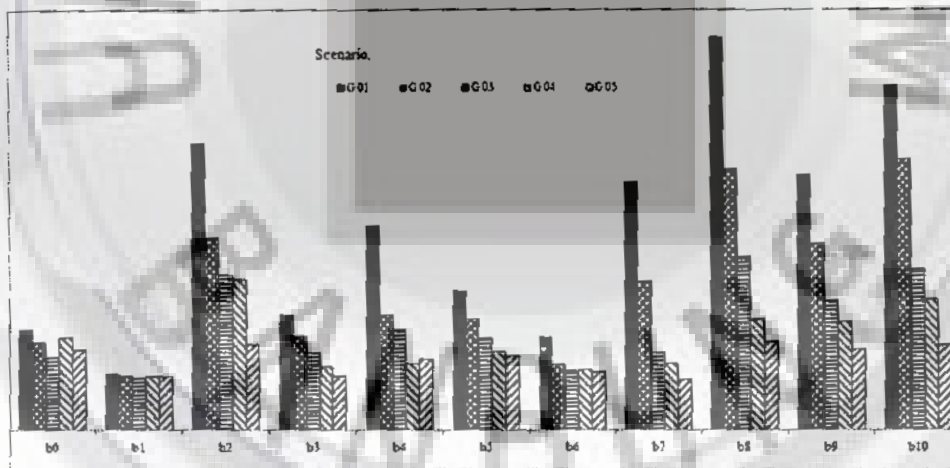


FIGURE 2. Variance Pattern of the bootstrap Estimate of The Regression Parameters.

DISCUSSION

It is well known that if there is rare event problem then the ordinary logistic gives bias estimates and the bias decreases if the proportion of rare event tends to 0.5. We used bootstrap approach to overcome the bias problem and obtained that combination of bootstrap method and under sampling approach to the data gives reducing bias. It is very promising us, that this approach can be used to reduce bias in rare event problem of logistic regression. Another advantage of this approach is its uses OLR as the basic computation, that usually provided by any statistical software. However, we do not conclude that this method results an unbiased estimate since we only compare with the results of ordinary logistic regression instead the true value of model parameter. Hence we need to do simulation under known true value of parameter and applied this method. Under known value of parameter we may result stronger conclusion than resulted

on this paper. We may consider to combine this algorithm and non-semi parametric approach (Suliadi, et al, 2010; Suliadi, et al, 2013) in reducing bias, since it is well known that nonparametric and semiparametric approaches are very effective in reducing bias.

ACKNOWLEDGMENTS

This research was supported by LPPM Unisba (Bandung Islamic University) by Contract No: 014/LPPM-SKP/I/2016. We would like to thank to the organizing committee and the reviewer that have accepted this paper to be presented in this conference.

REFERENCES

- [1] P. McCullagh and J. A. Nelder, *Generalized Linear Models* (Chapman and Hall, London, UK, 1989) p. Chapt. 4 and 15.
- [2] G. M. Weiss and H. Hirsh, *Learning to Predict Extremely Rare Events. In AAAI Workshop on Learning from Imbalance Data Sets* (-, -, 2000), pp. 64–68.
- [3] S. M. A. Elrahman^{*} and A. Abraham, *Journal of Network and Innovative Computing* **1**, 332–340 (2013).
- [4] W. Fithian and T. Hastie, *The Annals of Statistics* **42**, 1693–1724 (2014).
- [5] M. Guns and V. Vanacker, *Nat. Hazards Earth Syst. Sci* **12**, 1937–1947 (2012).
- [6] G. King and L. Zheng, *Political Analysis* **9**, 137–163 (2001).
- [7] G. King and L. Zheng, *World Politics* **54**, 623–658 (2001b).
- [8] I. Kosmidis and D. Firth, *Electronic Journal of Statistics* **4**, 1097–1112 (2010).
- [9] J. Quigley, T. Bedford, and L. Walls, *Reliability Engineering and System Safety* **92**, 619–627 (2007).
- [10] K. P. N. V. Satyasree and J. V. R. Murthy, *International Journal of Emerging Trends and Technology in Computer Science* **2**, 109–118 (2013).
- [11] T. Sei, *Journal of Statistical Planning and Inference* **149**, 116–124 (2014).
- [12] S. Suliadi, N. A. Ibrahim, I. Daud, and I. S. Krishnarajah, *Journal of Applied Probability and Statistics* **5**, 77–93 (2010).
- [13] S. Suliadi, N. A. Ibrahim, and I. Daud, *Communications in Statistics-Simulation and Computation* **42**, 138–152 (2013).
- [14] Z. Qiu, H. Li, H. Su, G. Ou, and T. Wang, “Logistic regression bias correction for large scale data with rare events,” in *Advanced Data Mining and Applications*, edited by L. C. O. Z. M. Y. H. Motoda, Z. Wu and W. Wang (Springer-Verlag, Berlin, Germany, 2013), pp. 133–143.
- [15] B. W. Yap, K. A. Rani, H. A. A. Rahman, S. Fong, Z. Khairudin, and N. N. Abdullah, “An application of oversampling, undersampling, bagging and boosting in handling imbalanced datasets,” in *Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013)*, Lecture Notes in Electrical Engineering 285 (Springer Science+Business Media, Singapore, 2014), pp. 13–22.
- [16] Anonymous, Give me some credit, <https://www.kaggle.com/c/GiveMeSomeCredit>. Cited: 2015-09-25.

Comparison of Credit Scoring Model between Logistic Regression and Cox Proportional Hazard Model Based on Mean Cost Criteria*

Abdul Kudus, Suliadi, Pupung Purnamasari, Devisa Negara

Department of Statistics, Bandung Islamic University, Bandung, Indonesia

Email: abdul.kudus@unisba.ac.id

How to cite this paper: Author 1, Author 2 and Author 3 (2016) Paper Title *****, *, *-*

http://dx.doi.org/10.4236/ojs.2016.****

Received: ****, **
Accepted: ****, **
Published: ****, **

Copyright © 2016 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Credit scoring is a technique that help bank in decided a loan. The purpose of making credit scoring is to select debtors efficiently. Logistic regression is a common technique used in credit scoring. Logistic regression formulates the default probability of debtors. On the other hand, it is interested on formulating the time of default occurred. The Cox proportional hazard model can be used for the second purpose. The comparison of both models will be presented in this paper. Mean cost criteria can be used to select the better one. Based on the application of both model to real data, the results shows that logistic regression model was better than Cox proportional hazard model.

Keywords

Credit scoring, logistic regression, Cox proportional hazard, optimal cut-off point, mean cost criteria.

1. Introduction

Loan activity has its long history. One of the financial institutions that conduct business loan is bank. Bank as a financial institution has the task to raise funds from the public and then distribute the funds back to the community in the form of loan.

Bank has some risks in doing business loan. One of the important risks is default risk. The default risk is due to the inability of customers to settle the bank loan. Those customers are called bad customer. The greater the percentage of bad customer, the more difficult for the bank to conduct business.

In addition, one of the important factors causing the presence of bad customer the weakness in conducting credit scoring [1]. Credit scoring is a predictive tools and techniques that help financial institutions in deciding a loan [2]. The objective of making credit

*Credit scoring comparison based on logistic regression and Cox proportional hazard model.

scoring is to analyze and make decisions more quickly, accurately and efficiently in the selection of customers to get a loan [3]. Credit scoring is also a method for evaluating the credit worthiness of a person based on a specific formula or a particular rule. At this moment the banking system still used manual method to decide a loan. This method is conducted based on survey-based assets. It is very ineffective for resolving the rising in number of credit demand which needs quick decision. Drawbacks of manual systems were slow, inaccurate, and it needs more time and operators to become an expert in deciding loan receiver. There are ways to minimize this risk is by means of credit scoring. credit scoring has become an important requirement, especially for banks and other financial institutions in minimizing the risk of loss.

Logistic regression is a technique that is very common in building credit scoring. The dependent variable in logistic regression is categorical variables [4]. For credit scoring purposes, logistic regression models the probability of default. On the other hand, the main interest is not on what is the probability of default, but also on when the default occurred. For the second purposes, it is powerful to use Cox proportional hazard model which is one kind of method in survival analysis.

The analysis of survival can build credit scoring to consider when default occurs. Survival analysis focus on modeling opportunities on the occurrence of an event that is monitored from the beginning until the emergence of phenomena taken, namely the "default". In this case, the incident in the form of development a debtor remained perform on its loan repayments since the beginning of the installments up to experience default events or managed to pay off the mortgage. Cox proportional hazard models is one of the special models in survival analysis which aims to model survival data with explanatory variables that affect the function of hazard. Through the Cox proportional hazard models can be seen the relationship between the independent variable on the dependent variable is the time of survival through hazardnya function.

Good models will have an impact on the selection of prospective customers who will receive loans accurately. In order for us to know how well the model, then the credit scoring produced by logistic regression models and Cox proportional hazard will be evaluated by means of mean cost corresponding to certain of the cut-off value. Cut-off is a value to determine how the predictions of each individual studied, whereas the mean cost is the cost or loss due to incorrect predict. Where there are two types of wrong prediction is a prediction of type I and type II incorrect prediction. One prediction of type I is stated predictions as "good" when in fact "bad" and a prediction predictions type II is expressed as "bad" when "good". Losses due to each one of these predictions are of course different, where one prediction I have the kind of cost that is greater than one prediction type II. Based on the description of the background that has been disclosed, then the problem can be identified in this thesis is How do you build a credit scoring is formed based on logistic regression and Cox proportional hazard? And How do you compare credit scoring produced by logistic regression and Cox proportional hazard with using the criteria of mean cost?.

2. Logistics Regression

Logistic regression is a statistical analysis methods were used to analyze the relationship between the dependent variable is a binary or dichotomous with one or more independent variables [5]. In regesi logistics, scale categorical dependent variable. The dependent variable is denoted by y is binary or dichotomous which has two values: 0 and 1. Under such circumstances, the variable y to follow the Bernoulli distribution for each single observation. Probability function for each observation is given as follows:

$$f(y) = \pi^y(1 - \pi)^{1-y} \quad y = 0,1 \quad \dots(2.1)$$

Logistic regression function can be written as follows:

$$\pi(x_i) = \frac{e^{(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})}}{1 + e^{(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})}} \quad \dots(2.2)$$

where the function $\pi(x_i)$ is non-linear shaped so as to make it a linear function logit transformation should be performed as follows:

$$g(x_i) = \ln\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \quad \dots(2.3)$$

3. Cox Proportional Hazard Model

To model the survival data with explanatory variables that affect the function of hazard is proportional hazard model proposed by Cox and is also known as Cox regression. Excess Cox is no need to have a function of the parametric distribution. Modeling assumptions only validates the assumption that the function must be proportional hazard every time. Assumptions proportional to the model can be seen through the plot $-\log[-\log S(t)]$ against survival time (t) for each category in the p explanatory variables that form a pattern that is parallel at different levels.

For example hazard depends on the values of the p -explanatory variables x_1, x_2, \dots, x_p . The values of the explanatory variables in the proportional hazard model is expressed in vector form x , so $x = x_1, x_2, \dots, x_p$. Suppose that is a function of hazard for people who all of the explanatory variables vector has a value of zero, then the function called baseline hazard function.

General proportional hazard model is as follows

$$h_i(t) = \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) h_0(t) \quad \dots(2.4)$$

can also be expressed in equation form log,

$$\log\left\{\frac{h_i(t)}{h_0(t)}\right\} = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \quad \dots(2.5)$$

4. Mean Cost Criteria

Mean cost is a risk of the banking system. Definition of mean cost is the cost or loss due to incorrect predict. The smaller the value of the mean cost model will then be better. Before knowing the mean cost, will set the value of the optimal cut-off. Cut off is a value to determine how the predictions of each individual studied. Cut-off value is usually determined by each bank policy, namely by 0.5. But that prediction is right, then it should be set optimal cut-off value.

If the probability of default \geq the optimal cut-off, $\hat{y} = 1$ and

If the probability of default $<$ the optimal cut-off $\hat{y} = 0$

With contingency table between y and \hat{y} :

		\hat{y}	
		0	1
y	0	f_{00}	f_{01}
	1	f_{10}	f_{11}

5. ROC curve

ROC curve is a plot of the combination of the sensitivity value with the 1-specificity with various cut-offs are possible. A model is said to be perfect if predict all either the debtor is represented by horizontal line or positive rate true. A point on the ROC curve is determined by first forming the classification table with a certain cut-off. The determination of the cut-offnya done by the probability of non-performing loans (). After the probability of non-performing loans () is obtained, then calculate the sensitivity and specificity in advance. Results of cross tabulation called the confusion matrix. The shape of the confusion matrix shown in Figure 2.1

		Prediction (\hat{y})	
		+ (Positive)	- (Negative)
Actual (y)	+ (Positive)	TP	FN
	- (Negative)	FP	TN

Figure 2.1 Forms of confusion matrix for certain c cutpoint

6. Results and Discussion

The first stage is to divide the data into two groups: training data by 50% and data testing by 50%.

The estimated value of bad debts opportunities for each debtor in the training data, then be converted into estimates the response variable. But previously necessary to find out what is the limit or cut-off optimized using ROC curves. Search procedures

optimal cut-off value through the calculation of sensitivity and specificity in advance. As explained in (2.14) and (2.15). Optimal cut-off value is the maximum value of the sensitivity + (1-specificity) that is equal to 0.625.

Table 3.1 Classification Table In Training Data

Actual		Prediction	
		\hat{y}	
		0	1
y	0	224	2
	1	22	4

Costs or losses as a result of logistic regression models predicting the training data is equal to 1.753968.

Table 3.2 Classification Table On Testing Data

Actual		Prediction	
		\hat{y}	
		0	1
y	0		5
	1	29	

Costs or losses as a result of logistic regression models predicting the testing of data is equal to 2.321429.

The estimated value of bad debts opportunities for each debtor in the training data, then be converted into estimates the response variable (δ). But previously necessary to find out what is the limit of bad credit opportunities or the optimal cut-off using ROC curves. Search procedures optimal cut-off value through the calculation of sensitivity and specificity in advance. As explained in the press (2.14) and (2.15). Optimal cut-off value is the maximum value of the sensitivity + (1-specificity) that is equal to 0.55.

Table 3.3 Classification Table In Training Data

Actual		Prediction	
		$\hat{\delta}$	
		0	1
δ	0		0
	1	25	1

Costs or losses as a result of Cox proportional hazard models to predict the training data is equal to 1.984127.

Table 3.4 Classification Table On Testing Data

Aktual		Prediksi	
		$\hat{\delta}$	
		0	1
δ	0		0
	1	30	

Costs or losses as a result of Cox proportional hazard models to predict the testing of data is equal to 2.380952.

Cost Comparison of Mean Values Logistic Regression Models and Cox Proportional Hazard

Table 3.5 Value Mean Cost Model Logistic regression and Cox Proportional Hazard

	Regresi Logistik	Cox Proporsional Hazard
Training	1,753968	1,984127
Testing	2,321429	2,380952

The mean value of cost or loss as a result of logistic regression models predicting the training data is equal to 1.753968, while the testing of data is equal to 2.321429. And the mean cost or loss due to incorrect Cox proportional hazard models to predict the training data is equal to 1.984127, while the testing of data is equal to 2.380952.

7. Conclusion

Based on the results and discussion that has been done is to model and calculate the mean cost of a data using logistic regression models and Cox proportional hazard models can be drawn several conclusions, namely: 1) The mean value of cost or loss as a result of logistic regression models predicting the training data is equal to 1.753968, while the testing of data is equal to 2.321429, 2) The mean value of cost or loss due to incorrect Cox proportional hazard models to predict the training data is equal to 1.984127, while the testing of data is equal to 2.380952.

Because the mean cost logistic regression model is smaller than the Cox proportional hazard models, the logistic regression model was slightly better than the Cox proportional hazard models. This is in contrast with the results in the journal because in this thesis does not involve time-dependent covariate.

References

- [1] Budisantoro, Totok & Triandaru, Sigit. (2006), Banks and Other Financial Institutions. Issue 2, Jakarta: Four Salemba.
- [2] Collett, D. (2003). Modelling Survival Data in Medical Research. US: Chapman & Hall.
- [3] Hajarisman. Nusar. 2009. Textbook analysis of categorical data. Bandung: Program Statistics

CERTIFICATE

The 1st International Conference Technology on Biosciences and Social Sciences


Title "Industry Based on Knowledges"


November 8-9th, November 2016, Convention Hall, University of Andalas
Padang - Indonesia

Awarded to

Abdul Kudus
As Oral Presenter




Prof. Dr. Herrywan Prayitno, M.Si., Psy.
Governor of West Sumatra

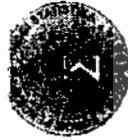

Prof. Dr. Infoli Husni, SE., MBA.
Rector of The University of Andalas


Prof. Dr. Hj. Endang Purwati, MS, Ph.D.
Chairman of The Committee





Department of Statistics
 Faculty of Mathematics and Natural Sciences
 Universitas Padjadjaran



CERTIFICATE

This is to certify that

Suljadi

has presented a paper with entitled

Statistical Inference Logistic Regression with Kere (Kerapok) Data

in The 2nd International Conference
 on Applied Statistics
 (ICAS II 2016)

Bandung, 27-28 September 2016

The Dean of Faculty of Mathematics and
 Natural Sciences, Universitas Padjadjaran

The Chairman of the Organizing Committee

Sudirat

Prof. Dr. Sudirat, MS.
 N.P. 19580519 198601 1 001

Yudie Andriyana

Dr. Yudie Andriyana, M.Sc.
 N.P. 19780722 200212 1 005