

SKRIPSI

**PEMILIHAN MODEL REGRESI LOGISTIK LASSO
MENGUNAKAN AIC: STUDI KASUS PREDIKSI DEFAULT KREDIT**



LUCKY RAMADHANI

22106010087

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA**

2026

**PEMILIHAN MODEL REGRESI LOGISTIK LASSO
MENGUNAKAN AIC: STUDI KASUS PREDIKSI DEFAULT KREDIT**

Skripsi

Untuk memenuhi sebagian persyaratan

mencapai derajat Sarjana S-1

Program Studi Matematika



diajukan oleh:

LUCKY RAMADHANI

22106010087

Kepada

PROGRAM STUDI MATEMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA

YOGYAKARTA

2026



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI

Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-2215/Un.02/DST/PP.00.9/08/2026

Tugas Akhir dengan judul : Pemilihan Model Regresi Logistik LASSO Menggunakan AIC: Studi Kasus Prediksi Default Kredit

yang dipersiapkan dan disusun oleh:

Nama : LUCKY RAMADHANI
Nomor Induk Mahasiswa : 22106010087
Telah diujikan pada : Rabu, 12 Agustus 2026
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang

Sri Utami Zuliana, S.Si., M.Sc., Ph.D.

SIGNED

Valid ID: 6a87f28891d05



Penguji I

Dr. Mohammad Farhan Qudratullah, S.Si.,
M.Si

SIGNED

Valid ID: 6a8ff33507693



Penguji II

Muhamad Zaki Riyanto, S.Si., M.Sc.
SIGNED

Valid ID: 6a9107334895d



Yogyakarta, 12 Agustus 2026

UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi

Prof. Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 6a91488a1706c



SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi / Tugas Akhir

Lamp :

Kepada

Yth. Dekan Fakultas Sains dan Teknologi

UIN Sunan Kalijaga Yogyakarta

di Yogyakarta

Assalamu 'alaikum wr. wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : Lucky Ramadhani

NIM : 22106010087

Judul Skripsi : PEMILIHAN MODEL REGRESI LOGISTIK LASSO

MENGGUNAKAN AKAIKE INFORMATION CRITERION (AIC)

STUDI KASUS : PREDIKSI DEFAULT KREDIT

sudah dapat diajukan kembali kepada Program Studi Matematika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Matematika.

Dengan ini kami berharap agar skripsi/tugas akhir Saudara tersebut di atas dapat segera dimunaqasyahkan. Atas perhatiannya kami ucapkan terima kasih.

Wassalamu 'alaikum wr. wb.

Yogyakarta, 5 Agustus 2026

Pembimbing

Sri Utami Zuliana, S.Si., M.Sc., Ph.D.

NIP. 19741003 200003 2 002

SURAT PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini:

Nama : Lucky Ramadhani
NIM : 22106010087
Program Studi : Matematika
Fakultas : Sains dan Teknologi

Dengan ini menyatakan bahwa isi skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar sarjana di suatu Perguruan Tinggi dan sesungguhnya skripsi ini merupakan hasil pekerjaan penulis sendiri sepanjang pengetahuan penulis, bukan duplikasi atau saduran dari karya orang lain kecuali bagian tertentu yang penulis ambil sebagai bahan acuan. Apabila terbukti pernyataan ini tidak benar, sepenuhnya menjadi tanggung jawab penulis.

Yogyakarta, 5 Agustus 2026


08E55A0X206264325
Lucky Ramadhani

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

HALAMAN PERSEMBAHAN

ALHAMDULILLAH, PUJI SYUKUR KE HADIRAT ALLAH SWT ATAS SEGALA RAHMAT DAN KEMUDAHAN-NYA SEHINGGA SKRIPSI INI DAPAT DISELESAIKAN. KARYA SEDERHANA INI PENULIS PERSEMBAHKAN KEPADA BAPAK DAN IBU, SUMBER KEKUATAN TERBESAR YANG SELALU MENDOAKAN DAN MENDUKUNG TANPA BATAS, SERTA SAUDARA-SAUDARAKU ATAS SEGALA SEMANGAT YANG DIBERIKAN. TERIMA KASIH PULA KEPADA SAHABAT DAN TEMAN-TEMAN SEPERJUANGAN YANG TELAH MENJADI TEMPAT BERKELUH KESAH, SALING MENGUATKAN, DAN MENEMANI DI TENGAH KEPENATAN HINGGA GARIS AKHIR. TERAKHIR, TERIMA KASIH KEPADA DIRI SENDIRI YANG TELAH MEMILIH UNTUK BERTAHAN, TIDAK MENYERAH PADA KEADAAN, DAN TERUS MELANGKAH MAJU SAMPAI TITIK INI.

"MATEMATIKA ADALAH BAHASA, DAN PERJALANAN INI ADALAH CARAKU MEMAHAMINYA."

"MATEMATIKA"

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

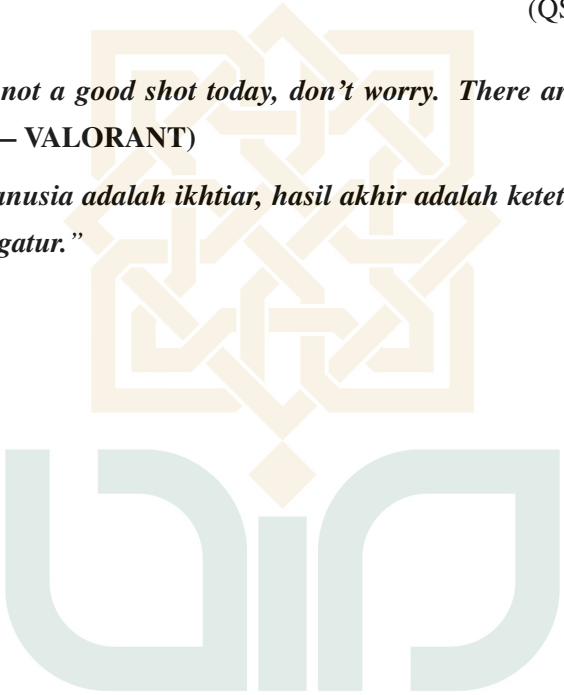
MOTTO

"Dan bahwasanya seorang manusia tidak memperoleh selain apa yang telah diusahakannya, dan bahwasanya usaha itu kelak akan diperlihatkan (kepadanya)."

(QS. An-Najm: 39–40)

"If you're not a good shot today, don't worry. There are other ways to be useful." (Sova — VALORANT)

"Usaha manusia adalah ikhtiar, hasil akhir adalah ketetapan terindah dari Sang Maha Pengatur."



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

KATA PENGANTAR

Alhamdulillah, segala puji dan syukur senantiasa penulis panjatkan ke hadirat Allah SWT atas segala limpahan rahmat, hidayah, serta kekuatan-Nya sehingga penyusunan skripsi ini dapat diselesaikan. Shalawat beserta salam semoga selalu tercurahkan kepada teladan kita, Nabi Muhammad SAW, beserta keluarga dan para sahabatnya. Atas rida dan kemudahan dari Allah SWT, penulis berhasil menyelesaikan skripsi yang berjudul “Pemilihan Model Regresi Logistik LASSO Menggunakan Akaike Information Criterion (AIC) Studi Kasus: Prediksi Default Kredit” sebagai salah satu syarat penyelesaian studi.

Selama proses penelitian dan penulisan skripsi ini, penulis menghadapi berbagai dinamika dan tantangan. Namun, berkat doa, bimbingan, arahan, serta dorongan dari banyak pihak, hambatan tersebut dapat dilalui dengan baik. Oleh karena itu, melalui lembar ini, penulis ingin menyampaikan rasa terima kasih dan penghargaan yang setinggi-tingginya kepada:

1. Bapak Prof. Noorhaidi Hasan, S.Ag., M.A., M.Phil., Ph.D., selaku Rektor Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
2. Ibu Prof. Dr. Dra. Hj. Khurul Wardati, M.Si., selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
3. Ibu Dr. Epha Diana Supandi, S.Si., M.Sc., selaku Ketua Program Studi Matematika.
4. Ibu Sri Utami Zuliana, S.Si., M.Sc., Ph.D., selaku dosen pembimbing. Terima kasih atas kesabaran, kelapangan waktu, dan ketelitian Ibu dalam memberikan bimbingan, masukan, serta evaluasi yang sangat berharga sejak awal hingga selesainya skripsi ini. Penulis juga memohon maaf yang sebesar-besarnya atas segala kekurangan dan kekhilafan selama masa bimbingan berlangsung.
5. Segenap bapak dan ibu dosen Program Studi Matematika serta tenaga kependidikan Fakultas Sains dan Teknologi yang telah membekali penulis dengan berbagai keilmuan dan memberikan fasilitas pelayanan terbaik selama masa perkuliahan.

6. Kedua orang tua, kakak, serta keluarga besar tercinta yang tak pernah lelah mendoakan, mencurahkan kasih sayang, dan memberikan dukungan, baik secara moril maupun materiil, dalam setiap langkah penulis.
7. Keluarga besar Mahasiswa Matematika angkatan 2022. Terima kasih atas solidaritas, kebersamaan, dan cerita-cerita bermakna yang telah terukir selama empat tahun menimba ilmu bersama.
8. Rekan-rekan seperbimbingan yang telah menjadi teman diskusi yang suportif dan saling menguatkan dalam perjuangan menyelesaikan tugas akhir ini.
9. Sahabat-sahabat terdekat penulis yang senantiasa hadir merangkul, sedia mendengarkan keluh kesah, serta menyuntikkan semangat yang luar biasa di kala suka maupun duka.

Semua pihak yang tidak dapat penulis sebutkan satu per satu, yang telah memberikan bantuan dan doa restu selama perjalanan akademik penulis. Semoga Allah SWT membalas segala kebaikan yang telah diberikan dengan pahala dan keberkahan yang berlipat ganda.

Penulis menyadari sepenuhnya bahwa karya tulis ini masih jauh dari kesempurnaan dan memiliki berbagai keterbatasan. Oleh karena itu, kritik dan saran yang konstruktif sangat penulis harapkan demi penyempurnaan karya ini. Akhir kata, penulis berharap semoga skripsi ini dapat memberikan kontribusi positif dan wawasan yang bermanfaat bagi pengembangan ilmu pengetahuan serta bagi para pembaca. Aamiin ya Rabbal 'Alamin.

Yogyakarta, 20 Juli 2026

Lucky Ramadhani

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN	i
HALAMAN PERSETUJUAN	iii
HALAMAN PERNYATAAN KEASLIAN	iv
HALAMAN PERSEMBAHAN	v
MOTTO	vi
PRAKATA	vii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMBANG	xiv
ABSTRAK	xvi
ABSTRACT	xvii
I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Batasan Masalah	3
1.4. Tujuan Penelitian	3
1.5. Manfaat Penelitian	4
1.6. Tinjauan Pustaka	4
1.7. Sistematika Penulisan	5
II LANDASAN TEORI	7
2.1. Konsep Dasar <i>Credit Scoring</i> dan Risiko <i>Default</i>	7
2.2. Teknik Pembagian Data (<i>Data Splitting</i>)	7
2.3. Model Linier Klasik dan Batas Keberlakuannya	8

2.4. <i>Generalized Linear Model</i> (GLM)	9
2.4.1. Distribusi Bernoulli dan Binomial	10
2.4.2. Model Regresi Logistik Biner	10
2.4.3. Pengodean Variabel <i>Dummy</i>	11
2.4.4. Estimasi Parameter via Algoritma <i>Iteratively Reweighted Least Squares</i> (IRLS)	11
2.5. Pemeriksaan Multikolinearitas dan Batasan Model Klasik	13
2.6. Estimasi Regresi Logistik LASSO	13
2.6.1. Formulasi Optimisasi Penalti L_1	13
2.6.2. Algoritma Aproksimasi Kuadratik Lokal dan <i>Cyclical Coordinate Descent</i>	14
2.7. Optimisasi Parameter λ Menggunakan <i>Akaike Information Criterion</i> (AIC)	15
2.7.1. Landasan Teori Informasi dan Penurunan Matematis AIC	15
2.7.2. Derajat Bebas Efektif (<i>Effective Degrees of Freedom</i>) pada LASSO	16
2.7.3. Formulasi Optimisasi AIC pada Pemodelan Logistik LASSO	17
2.7.4. Komparasi Teoretis AIC terhadap BIC dan <i>Cross-Validation</i>	17
2.8. Evaluasi Kinerja Model pada Data Uji	18
2.8.1. Matriks Konfusi (<i>Confusion Matrix</i>)	19
2.8.2. Kurva ROC dan <i>Area Under the Curve</i> (AUC)	19
2.8.3. Ketidakseimbangan Kelas dan Paradoks Akurasi	19
2.9. Interpretasi Parameter Model Terpilih	20
III METODE PENELITIAN	21
3.1. Jenis dan Pendekatan Penelitian	21
3.2. Sumber Data, Populasi, dan Sampel Penelitian	21
3.3. Variabel Penelitian	22
3.3.1. Variabel Respons (Y)	22
3.3.2. Variabel Prediktor Bebas (X)	22

3.4. Metode Pengumpulan dan Prapemrosesan Data	23
3.5. Perangkat Lunak (<i>Software</i>) Mandatori dan Pustaka R	23
3.6. Tahapan Analisis Data	24
3.6.1. Tahap 1: Pemuatan Data (<i>Input Data</i>)	24
3.6.2. Tahap 2: Pembagian Data (<i>Data Splitting</i>)	24
3.6.3. Tahap 3: Model Regresi Logistik Penuh	24
3.6.4. Tahap 4: Pemeriksaan Multikolinearitas (VIF)	25
3.6.5. Tahap 5: Estimasi Regresi Logistik LASSO	25
3.6.6. Tahap 6: Optimisasi Parameter λ dengan AIC	26
3.6.7. Tahap 7: Evaluasi Model pada Data Uji	26
3.6.8. Tahap 8: Interpretasi Model Terpilih	27
3.7. Diagram Alir Penelitian	27
IV HASIL DAN PEMBAHASAN	29
4.1. Analisis Deskriptif	30
4.2. Pemeriksaan Multikolinearitas (VIF)	30
4.3. Estimasi Model Regresi Logistik LASSO	32
4.3.1. Proses Seleksi Variabel	32
4.3.2. Pemilihan Parameter λ Optimal	33
4.4. Evaluasi Performa Model pada Data Uji	33
4.5. Interpretasi Koefisien dan <i>Odds Ratio</i>	35
V PENUTUP	38
5.1. Kesimpulan	38
5.2. Saran	38
DAFTAR PUSTAKA	40
LAMPIRAN	43
A DATA PENELITIAN	43
B Source Code Program R	44
BIODATA PENULIS	49

DAFTAR TABEL

2.1	Struktur <i>Confusion Matrix</i> Klasifikasi Risiko Kredit Biner	19
4.1	Distribusi Variabel Respons (<i>Y</i>)	30
4.2	Hasil Pengujian <i>Variance Inflation Factor</i> (VIF) Model Penuh	31
4.3	Ringkasan Hasil Seleksi Variabel Regresi Logistik LASSO	33
4.4	<i>Confusion Matrix</i> Prediksi <i>Default</i> Kredit pada Data Uji	34
4.5	Resume Metrik Performa Model pada Data Uji	34
4.6	Estimasi Koefisien dan <i>Odds Ratio</i> Model Terpilih	36
A.1	Skema Pengodean dan Struktur Variabel Data Kartu Kredit	43



DAFTAR GAMBAR

3.1	Diagram Alir Penelitian Prediksi <i>Default</i> Kredit	28
4.1	Plot Lintasan Koefisien Regresi Logistik LASSO	32
4.2	Plot Pemilihan Parameter λ Optimal Berdasarkan Nilai AIC	34
4.3	Kurva ROC Model Regresi Logistik LASSO-AIC ($AUC = 0,7202$)	35



DAFTAR LAMBANG

Y	= variabel respons biner (status <i>default</i> kartu kredit)
X	= variabel prediktor (atribut nasabah dan riwayat kredit)
x_{ij}	= nilai pengamatan variabel prediktor ke- j pada observasi ke- i
$\mathbf{x}_i$	= vektor kolom variabel prediktor observasi ke- i berukuran $(p + 1) \times 1$
$\mathbf{x}_i^\top$	= vektor transpos prediktor observasi ke- i berukuran $1 \times (p + 1)$
y_i	= nilai aktual variabel respons pada observasi ke- i (1 atau 0)
$\hat{Y}_i$	= nilai hasil klasifikasi prediksi biner observasi ke- i
β_0	= konstanta atau intersep model regresi
β_j	= koefisien regresi untuk variabel prediktor ke- j
$\boldsymbol{\beta}$	= vektor parameter regresi berukuran $(p + 1) \times 1$
$\hat{\boldsymbol{\beta}}$	= vektor penduga (<i>estimator</i>) parameter regresi
$\tilde{\boldsymbol{\beta}}$	= estimasi nilai parameter regresi pada iterasi berjalan
p_i	= peluang observasi ke- i masuk ke dalam kelas target ($Y = 1$)
$\hat{p}_i$	= estimasi peluang prediksi keberhasilan pada observasi ke- i
$\pi(x)$	= probabilitas sukses yang bergantung pada nilai prediktor x
μ_i	= nilai harapan (<i>mean</i>) dari respons pada pengamatan ke- i
η_i	= prediktor linier pada pengamatan ke- i ($\mathbf{x}_i^\top \boldsymbol{\beta}$)
$g(\cdot)$	= fungsi penghubung (<i>link function</i>) logit
$\exp(\cdot)$	= fungsi eksponensial natural ($e^\cdot$)
θ_i	= parameter alami (<i>canonical parameter</i>) pada keluarga eksponensial
ϕ	= parameter dispersi (<i>dispersion parameter</i>)
ϵ_i	= sisaan acak (<i>error</i>) pada pengamatan ke- i
$L(\boldsymbol{\beta})$	= fungsi <i>likelihood</i>
$\ell(\boldsymbol{\beta})$	= fungsi <i>log-likelihood</i>
$\mathbf{U}(\boldsymbol{\beta})$	= vektor gradien turunan pertama (<i>score vector</i>)
$\mathbf{H}(\boldsymbol{\beta})$	= matriks turunan parsial kedua (<i>Hessian matrix</i>)
λ	= parameter regularisasi atau <i>tuning</i> pada metode LASSO
λ_{optimal}	= nilai parameter penalti λ terbaik berdasarkan AIC minimum
$\ \boldsymbol{\beta}\ _1$	= fungsi penalti pembatas norma L_1 ($\sum_{j=1}^p \beta_j $)

$\mathcal{S}_\gamma(\cdot)$	= operator ambang batas lunak (<i>soft-thresholding operator</i>)
$\text{sign}(\cdot)$	= fungsi tanda arah bilangan ($\{+1, 0, -1\}$)
$\mathbf{X}$	= matriks variabel prediktor (matriks desain) berukuran $n \times (p + 1)$
$\mathbf{W}$	= matriks pembobot diagonal berukuran $n \times n$ pada algoritma IRLS
w_{ii}	= elemen diagonal matriks pembobot varians Bernoulli ($p_i(1 - p_i)$)
$\mathbf{z}$	= vektor respons semu (<i>working response</i>) pada algoritma IRLS
z_i	= nilai respons kerja semu pada observasi ke- i
$r_{i,-j}$	= sisaan parsial (<i>partial residual</i>) tanpa prediktor ke- j
n	= jumlah total observasi sampel
n_{train}	= jumlah observasi partisi data latih (24.000 data)
n_{test}	= jumlah observasi partisi data uji (6.000 data)
p	= jumlah variabel prediktor bebas (23 variabel)
k	= jumlah parameter yang diestimasi dalam model (termasuk intersep)
$\text{df}(\lambda)$	= derajat bebas efektif (<i>effective degrees of freedom</i>) LASSO
$I(\cdot)$	= fungsi indikator (bernilai 1 jika benar dan 0 jika salah)
D_{KL}	= jarak divergensi <i>Kullback-Leibler</i> (<i>information loss</i>)
τ	= nilai ambang batas keputusan klasifikasi (<i>cutoff threshold</i> = 0,5)
OR_j	= rasio kecenderungan (<i>Odds Ratio</i>) untuk variabel prediktor ke- j
R_j^2	= koefisien determinasi regresi prediktor X_j terhadap prediktor lainnya
AIC	= <i>Akaike Information Criterion</i>
VIF	= <i>Variance Inflation Factor</i>
TP	= <i>True Positive</i> (prediksi default benar)
TN	= <i>True Negative</i> (prediksi lancar benar)
FP	= <i>False Positive</i> (kesalahan tipe I / prediksi lancar salah)
FN	= <i>False Negative</i> (kesalahan tipe II / prediksi default salah)
ROC	= <i>Receiver Operating Characteristic</i>
AUC	= <i>Area Under the Receiver Operating Characteristic Curve</i>

ABSTRAK

Pemilihan Model Regresi Logistik LASSO Menggunakan AIC: Studi Kasus Prediksi Default Kredit

Oleh:

Lucky Ramadhani

22106010087

Risiko kredit macet (*default*) merupakan tantangan utama industri perbankan yang dapat ditekan melalui pemodelan penilaian kelayakan kredit (*credit scoring*). Penelitian ini memprediksi risiko gagal bayar tagihan kartu kredit menggunakan regresi logistik *Least Absolute Shrinkage and Selection Operator* (LASSO) yang disempurnakan dengan kriteria pemilihan model terbaik *Akaike Information Criterion* (AIC). Metode LASSO berfungsi menyaring serta membuang variabel transaksi yang saling bertumpuk (multikolinearitas), sementara AIC bertugas memilih kombinasi model yang paling sederhana dan akurat. Menggunakan 30.000 data nasabah dari *UCI Machine Learning Repository* dengan pembagian 80% data latih dan 20% data uji, algoritma berhasil memangkas dua variabel tagihan yang tidak efektif (*BILL_AMT4* dan *BILL_AMT5*) pada tingkat penalti optimal $\lambda = 0,0002013$. Model akhir yang terbentuk mampu memberikan ketepatan prediksi (akurasi) sebesar 81,16%, kemampuan menyaring nasabah lancar (spesifisitas) sebesar 97,20%, serta daya pembeda risiko (AUC) sebesar 0,7202. Hasil analisis juga membuktikan bahwa keterlambatan pembayaran pada bulan terakhir (*PAY_1*) merupakan faktor penentu utama, di mana setiap peningkatan tunggakan melipatgandakan risiko gagal bayar sebesar 1,766 kali lipat. Penelitian ini menyimpulkan bahwa integrasi LASSO dan AIC menghasilkan sistem penilaian kredit yang ringkas, stabil, dan aplikatif bagi perbankan untuk mendeteksi potensi kredit bermasalah sejak dini.

Kata Kunci: *Credit scoring*, Regresi logistik, LASSO, AIC, Risiko gagal bayar.

ABSTRACT

Selection of the LASSO Logistic Regression Model Using AIC: A Case Study on Credit Default Prediction

By:

Lucky Ramadhani

22106010087

Credit default risk poses a major challenge in the banking industry, which can be mitigated through predictive credit scoring models. This study predicts credit card default risk using Least Absolute Shrinkage and Selection Operator (LASSO) logistic regression optimized by the Akaike Information Criterion (AIC). The LASSO method eliminates redundant and highly correlated transaction variables (multicollinearity), while AIC selects the simplest and most accurate model architecture. Utilizing 30,000 customer records from the UCI Machine Learning Repository partitioned into 80% training and 20% testing sets, the algorithm successfully removed two redundant billing variables (`BILL_AMT4` and `BILL_AMT5`) at the optimal penalty level of $\lambda = 0.0002013$. The final model achieved an overall accuracy of 81.16%, a specificity of 97.20% in identifying non-default customers, and an AUC discrimination power of 0.7202. Furthermore, risk factor analysis revealed that the most recent repayment delay (`PAY_1`) is the most critical trigger, multiplying default risk by 1.766 times per delay level. This study concludes that integrating LASSO with AIC yields a concise, stable, and practical credit scoring system to assist banks in detecting high-risk borrowers early.

Keywords: Credit scoring, Default risk, Logistic regression, LASSO, AIC.

BAB I

PENDAHULUAN

1.1. Latar Belakang

Regresi logistik merupakan metode standar dalam analisis klasifikasi biner, terutama dalam industri perbankan untuk membangun sistem *credit scoring*. Kemampuannya dalam memberikan interpretasi yang transparan melalui nilai peluang dan *odds ratio* menjadikan metode ini sangat diandalkan untuk mengevaluasi kelayakan kredit nasabah (Yeh & Lien, 2009). Namun, dalam kerangka model linier umum (*Generalized Linear Model*), estimasi parameter regresi logistik tidak memiliki solusi tertutup (*closed-form solution*), sehingga memerlukan pendekatan numerik seperti pendekatan kuadratik aproksimasi yang dikombinasikan dengan algoritma *Cyclical Coordinate Descent* untuk mencapai konvergensi *Maximum Likelihood Estimation* (MLE).

Tantangan teknis muncul ketika model dihadapkan pada dimensi data yang besar (*Big Data*). Penggunaan variabel prediktor yang terlalu banyak tanpa seleksi yang tepat sering kali memicu masalah multikolinearitas yang parah. Hal ini menyebabkan matriks varians-kovarians menjadi tidak stabil, standar error membengkak, dan model kehilangan daya generalisasi atau mengalami *overfitting* (Yeh & Lien, 2009). Dalam kondisi ini, model mungkin tampak sangat akurat pada data latih namun gagal memberikan prediksi yang konsisten pada data baru.

Sebagai solusi atas permasalahan tersebut, metode *Least Absolute Shrinkage and Selection Operator* (LASSO) menawarkan teknik regularisasi melalui penalti norma L_1 pada fungsi *log-likelihood* (Tibshirani, 1996). Berbeda dengan regresi klasik, lasso mampu menyusutkan koefisien variabel yang redundan atau tidak signifikan hingga tepat bernilai nol. Proses ini memungkinkan dilakukannya estimasi parameter sekaligus seleksi variabel secara simultan, menghasilkan model yang lebih sederhana (*parsimonious*). Dalam implementasinya, efektivitas lasso sangat bergantung pada pemilihan parameter penalti (λ) yang optimal. *Akaike Information Criterion* (AIC), yang berakar pada teori informasi, digunakan sebagai kriteria pemilihan model untuk

meminimalkan kehilangan informasi dengan menyeimbangkan *goodness of fit* dan kompleksitas model (Akaike, 1974).

Penerapan metode lasso dengan mengutamakan AIC dalam penelitian ini difokuskan pada pengolahan dataset nasabah kartu kredit yang memiliki karakteristik dimensi tinggi (Yeh & Lien, 2009). Dataset ini mencakup beragam informasi, mulai dari profil demografi hingga riwayat transaksi tagihan (*bill amount*) dan catatan pembayaran (*payment status*) selama enam bulan terakhir (Yeh & Lien, 2009). Secara teknis, variabel riwayat tagihan bulanan sering kali memiliki korelasi antar-waktu yang sangat kuat, yang secara statistik terdeteksi sebagai multikolinearitas ekstrem. Jika tidak ditangani, variabel-variabel yang saling bergantung ini akan membuat model prediksi *default* menjadi bias dan tidak reliabel bagi pihak bank. Oleh karena itu, penggunaan studi kasus ini bertujuan untuk mendemonstrasikan bagaimana regularisasi lasso dapat bekerja secara cerdas dalam menyaring variabel tagihan yang redundan, sementara kriteria AIC memastikan model yang dihasilkan tetap sederhana namun memiliki tingkat akurasi yang optimal dalam membedakan nasabah yang berisiko gagal bayar (Akaike, 1974).

Penerapan kombinasi metode regresi logistik lasso dan kriteria AIC ini menjadi sangat relevan dalam studi kasus manajemen risiko perbankan, khususnya pada prediksi *default* kredit (Yeh & Lien, 2009). Lembaga keuangan secara kontinu menghadapi risiko kerugian akibat kegagalan debitur memenuhi kewajiban kontraktualnya (Yeh & Lien, 2009). Bank saat ini memiliki akses ke ribuan titik data nasabah, mulai dari informasi demografi hingga riwayat transaksi tagihan yang sangat detail. Dengan memanfaatkan dataset histori transaksi nasabah pemegang kartu kredit, penelitian ini bertujuan untuk menguji bagaimana optimisasi lasso dengan mengutamakan AIC dapat mereduksi variabel yang mengalami multikolinearitas seperti pada variabel riwayat tagihan sehingga dihasilkan model klasifikasi risiko yang stabil, efisien, dan memiliki daya interpretasi yang kuat dalam menekan angka *Non-Performing Loan* (NPL).

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana mekanisme estimasi parameter dan seleksi variabel pada regresi logistik LASSO yang dioptimasi menggunakan *Akaike Information Criterion* (AIC) dalam menangani multikolinearitas pada data prediksi *default* kredit?
2. Bagaimana arsitektur model regresi logistik LASSO terbaik yang terpilih berdasarkan nilai AIC minimum pada data prediksi *default* kredit?
3. Apa saja faktor-faktor atau variabel prediktor yang paling dominan mempengaruhi risiko *default* kredit berdasarkan nilai *Odds Ratio* dari model terpilih?

1.3. Batasan Masalah

Agar fokus pada sasaran yang diharapkan maka perlu dilakukan pembatasan masalah sebagai berikut:

1. Metode klasifikasi yang digunakan adalah regresi logistik dengan regularisasi LASSO.
2. Pemilihan parameter regularisasi (λ) menggunakan *Akaike Information Criterion* (AIC).
3. Data yang digunakan adalah data sekunder *default* kredit nasabah (`default of credit card clients.csv`).
4. Perangkat lunak (*software*) yang digunakan adalah R versi 4.5.0.

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan dari penelitian ini adalah sebagai berikut:

1. Menjelaskan mekanisme estimasi parameter dan seleksi variabel pada regresi logistik LASSO yang dioptimasi menggunakan *Akaike Information Criterion* (AIC) dalam menangani multikolinearitas pada data prediksi *default* kredit.
2. Menentukan arsitektur model regresi logistik LASSO terbaik yang terpilih berdasarkan nilai AIC minimum pada data prediksi *default* kredit.

3. Mengidentifikasi faktor-faktor atau variabel prediktor yang paling dominan mempengaruhi risiko *default* kredit berdasarkan nilai *Odds Ratio* dari model terpilih.

1.5. Manfaat Penelitian

Secara teoritis, penelitian ini berkontribusi pada pengembangan metode seleksi variabel melalui kombinasi regularisasi LASSO dan kriteria AIC, serta memberikan penjelasan matematis mengenai penanganan fungsi penalti tak-terdiferensialkan melalui operator ambang batas lunak (*soft-thresholding*). Secara praktis, penelitian ini menyediakan kerangka analisis *credit scoring* yang efisien bagi industri perbankan untuk mengevaluasi kelayakan kredit nasabah dan menekan angka kredit macet (NPL) berdasarkan model yang parsimonius dan bebas multikolinearitas.

1.6. Tinjauan Pustaka

Tibshirani (1996) memperkenalkan metode LASSO sebagai alternatif untuk seleksi variabel dan regularisasi Tibshirani (1996). Penelitian ini menjelaskan bahwa LASSO menambahkan penalti norma L_1 pada koefisien regresi, yang memiliki sifat unik dapat menghasilkan solusi *sparse* (beberapa koefisien tepat nol). Tibshirani mendemonstrasikan bahwa LASSO dapat menghasilkan model yang lebih *interpretable* dibandingkan *ridge regression* karena kemampuannya dalam melakukan seleksi variabel secara otomatis. Melalui simulasi dan aplikasi pada data riil, penelitian ini membuktikan bahwa LASSO efektif dalam mengurangi *overfitting* dan meningkatkan akurasi prediksi, terutama pada kasus dengan banyak variabel prediktor yang berkorelasi.

Akaike (1974) memperkenalkan *Akaike Information Criterion* (AIC) sebagai kriteria pemilihan model yang didasarkan pada teori informasi Kullback-Leibler Akaike (1974). AIC didefinisikan sebagai $-2 \ln(L) + 2k$, di mana k adalah jumlah parameter dalam model dan L adalah fungsi *likelihood*. Penelitian ini menjelaskan bahwa AIC mengukur *trade-off* antara *goodness of fit* dan kompleksitas model, dengan tujuan meminimalkan kehilangan informasi. Akaike mendemonstrasikan bahwa AIC memberikan pendekatan objektif dalam membandingkan model-model yang berbeda dan cenderung memilih model yang mampu menjelaskan data dengan baik tanpa terlalu kompleks.

Hosmer, Lemeshow, dan Sturdivant (2013) menyajikan panduan komprehensif mengenai regresi logistik dalam berbagai aplikasi praktis. Buku ini menjelaskan secara detail tentang pembentukan model regresi logistik biner, penggunaan fungsi taut (*link function*), interpretasi koefisien, dan evaluasi model. Penulis juga membahas permasalahan praktis seperti penanganan variabel kategorikal melalui variabel *dummy*, multikolinearitas, dan *overfitting* yang menjadi landasan penting dalam memodelkan status risiko *default* kredit yang bersifat biner.

Yeh dan Lien (2009) memaparkan analisis komparatif performa deteksi risiko finansial kartu kredit menggunakan beberapa teknik penggalan data Yeh & Lien (2009). Penelitian ini mengonfirmasi bahwa rekam jejak perilaku keterlambatan pembayaran bulanan nasabah berbanding lurus dengan pembengkakan probabilitas terjadinya kegagalan kewajiban kontraktual, yang menjadi acuan utama basis data dalam penelitian ini.

Meier, van de Geer, dan Bühlmann (2008) Meier et al. (2008) mengembangkan perluasan teori LASSO khusus untuk model regresi logistik yang menangani variabel prediktor kategori. Penelitian ini menjadi dasar penting dalam pengelolaan variabel yang memiliki banyak kategori, sehingga seleksi variabel dilakukan pada seluruh kelompok variabel *dummy* secara simultan. Perluasan ini memastikan konsistensi pemilihan model pada data demografi yang kompleks.

1.7. Sistematika Penulisan

BAB I PENDAHULUAN Berisi latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat, tinjauan pustaka, dan sistematika penulisan.

BAB II LANDASAN TEORI Menjelaskan konsep Regresi Logistik, algoritma optimisasi koordinat menurun, operator pembatasan lunak, hingga kriteria pemilihan model AIC.

BAB III METODE PENELITIAN Membahas jenis penelitian, sumber data, variabel penelitian, dan tahapan analisis data secara sistematis.

BAB IV HASIL DAN PEMBAHASAN Berisi proses komputasi, hasil seleksi variabel, pemilihan λ optimal, dan evaluasi performa model.

BAB V PENUTUP Berisi kesimpulan akhir dan saran untuk pengembangan penelitian selanjutnya.



BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah diuraikan, kesimpulan penelitian ini disusun untuk menjawab rumusan masalah secara sistematis sebagai berikut:

1. Mekanisme estimasi parameter dan seleksi variabel pada regresi logistik LASSO dilakukan dengan mengombinasikan aproksimasi kuadrat lokal (*IRLS-like step*) dan algoritma *Cyclical Coordinate Descent*. Penalti norma L_1 dengan operator *soft-thresholding* terbukti efektif mengatasi multikolinearitas ekstrem dengan menyusutkan koefisien variabel redundan (*BILL_AMT4* dan *BILL_AMT5*) tepat menjadi nol secara simultan.
2. Arsitektur model regresi logistik LASSO terbaik yang terpilih berdasarkan nilai AIC minimum diperoleh pada tingkat parameter penalti optimal $\lambda = 0,0002013$. Titik optimal ini menghasilkan arsitektur model final yang paling parsimonius (sederhana) dan stabil tanpa mengorbankan informasi penting data.
3. Berdasarkan analisis nilai *Odds Ratio* dari model terpilih, faktor atau variabel prediktor yang paling dominan mempengaruhi risiko *default* kredit adalah riwayat status pembayaran bulan terakhir (*PAY_1*) dengan nilai *Odds Ratio* sebesar 1,766267, yang mengindikasikan bahwa keterlambatan pembayaran terkini meningkatkan risiko gagal bayar secara signifikan.

5.2. Saran

Berdasarkan hasil penelitian dan keterbatasan model yang telah diulas, terdapat beberapa saran yang dapat diberikan untuk pengembangan penelitian selanjutnya maupun bagi pihak praktisi:

1. Bagi lembaga perbankan atau instansi penerbit kartu kredit, disarankan untuk mengintegrasikan variabel riwayat status pembayaran bulan terakhir (*PAY_1*)

sebagai indikator peringatan utama (*red-flag*) di dalam sistem deteksi dini (*early warning system*) guna memitigasi potensi pembengkakan risiko kredit macet sejak awal.

2. Mengingat rendahnya nilai sensitivitas model (24,72%) akibat adanya kondisi ketidakseimbangan kelas (*class imbalance*) yang memicu terjadinya paradoks akurasi, penelitian selanjutnya disarankan untuk menerapkan teknik penanganan data tidak seimbang seperti *Synthetic Minority Over-sampling Technique* (SMOTE) atau mengaplikasikan fungsi pembobotan kelas (*class weight*) pada pemodelan regresi logistik guna meningkatkan sensitivitas deteksi pada kelompok nasabah minoritas yang mengalami *default*.
3. Penelitian selanjutnya disarankan untuk memperluas komparasi pemilihan parameter penalti λ optimal dengan membandingkan kriteria AIC terhadap kriteria informasi lain, seperti *Bayesian Information Criterion* (BIC) maupun metode *k-fold Cross-Validation*, untuk menguji tingkat konsistensi stabilitas variabel prediktor yang terseleksi.



DAFTAR PUSTAKA

- Agresti, A. (2002). *Categorical Data Analysis*. John Wiley & Sons, New Jersey, 2nd edition.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.
- Amalina, M. N. et al. (2026). *OPTIMISASI ANALISIS REGRESI LOGISTIK LASSO MENGGUNAKAN BAYESIAN INFORMATION CRITERION (BIC)(STUDI KASUS: FAKTOR PENYEBAB PERCERAIAN DI INDONESIA TAHUN 2024)*. PhD thesis, UIN SUNAN KALIJAGA YOGYAKARTA.
- Bolton, C. (2009). Logistic regression and its application in credit scoring. Master's thesis, University of Pretoria, South Africa.
- Dewi, A. S. (2024). Analisa variabel penting pada kasus credit scoring menggunakan permutation importance melalui random forest dan logistic regression.
- Fahira, F. (2022). Penerapan model regresi logistik least absolute shrinkage and selection operator pada credit scoring.
- Friedman, J., Hastie, T., & Tibshirani, R. (2010a). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22.
- Friedman, J., Hastie, T., & Tibshirani, R. (2010b). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media, New York, 2nd edition.
- Hurley, M. & Adebayo, J. (2016). Credit scoring in the era of big data. *Yale Journal of Law and Technology*, 18(1):148–216.
- Kuhn, M. & Johnson, K. (2013). *Applied Predictive Modeling*. Springer Science & Business Media.

- Louzada, F., Moreira, F., & Oliveira Jr, M. (2017). A zero-inflated non-default rate regression model for credit scoring data. *Communications in Statistics-Theory and Methods*, 46(24):12304–12314, DOI: 10.1080/03610926.2017.1346803.
- Meier, L., Van De Geer, S., & Bühlmann, P. (2008). The group lasso for logistic regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):53–71.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2001a). *Introduction to Linear Regression Analysis*. John Wiley & Sons, New York, 4th edition.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2001b). *Introduction to Linear Regression Analysis*. John Wiley & Sons, 3rd edition.
- Myers, R. H., Montgomery, D. C., Vining, G. G., & Robinson, T. J. (2010). *Generalized Linear Models: With Applications in Engineering and the Sciences*. John Wiley & Sons.
- Ncapai, T. H. (2019). Modelling credit scoring with data limitations: Logistic vs bayesian linear techniques. Minor dissertation, University of Johannesburg, South Africa.
- Sohn, S. Y., Kim, D. H., & Yoon, J. H. (2016). Technology credit scoring model with fuzzy logistic regression. *Applied Soft Computing*, 43:150–158.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.
- Wijayakanthi, F. A. et al. (2026). *OPTIMASI REGRESI LOGISTIK LASSO DENGAN KRITERIA BAYESIAN INFORMATION CRITERION (BIC)(STUDI KASUS: KLASIFIKASI KEMISKINAN DI INDONESIA TAHUN 2024)*. PhD thesis, UIN SUNAN KALIJAGA YOGYAKARTA.
- Yeh, I.-C. & Lien, C.-h. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications*, 36(2):2473–2480.

Zhou, X., Wang, X., & Dougherty, E. R. (2005). Gene selection using logistic regressions based on aic, bic and mdl criteria. *New Mathematics and Natural Computation*, 1(01):129–145.

Zou, H., Hastie, T., & Tibshirani, R. (2007). On the “degrees of freedom” of the lasso. *The Annals of Statistics*, 35(5):2173–2192.

