

**PENERAPAN MODEL REGRESI LOGISTIK *LEAST  
ABSOLUTE SHRINKAGE AND SELECTION OPERATOR*  
PADA *CREDIT SCORING***

**SKRIPSI**



**Fani Fahira  
H051181324**

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGATAHUN ALAM  
UNIVERSITAS HASANUDDIN**

**MAKASSAR**

**2022**

**PENERAPAN MODEL REGRESI LOGISTIK *LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR* PADA CREDIT SCORING**

**SKRIPSI**

**Diajukan sebagai salah satu syarat memperoleh gelar Sarjana Sains pada  
Program Studi Statistika Departemen Statistika Fakultas Matematika dan  
Ilmu Pengetahuan Alam Universitas Hasanuddin**



**FANI FAHIRA**

**H051181324**

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS HASANUDDIN**

**MAKASSAR**

**DESEMBER 2022**

## LEMBAR PERNYATAAN KEOTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

**Penerapan Model Regresi Logistik *Least Absolute Shrinkage And Selection Operator* Pada Credit Scoring**

adalah benar hasil karya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun

Makassar, 06 Desember 2022



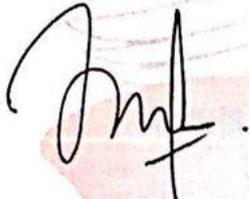
**FANI FAHIRA**

**NIM H051181324**

**PENERAPAN MODEL REGRESI LOGISTIK *LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR* PADA CREDIT SCORING**

**Disetujui Oleh:**

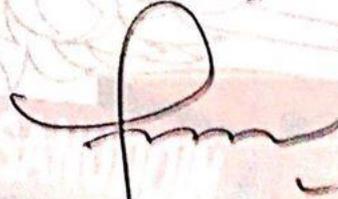
**Pembimbing Utama**



**Anisa, S.Si., M.Si.**

**NIP. 19730227 199802 2 001**

**Pembimbing Pertama**



**Dr. Nirwan, M.Si.**

**NIP. 19630306 198702 1 002**

**Ketua Program Studi**

  
**Dr. Nurfitri Sunusi, S.Si., M.Si.**  
**NIP. 19720117 199703 2 002**  


**Pada 06 Desember 2022**

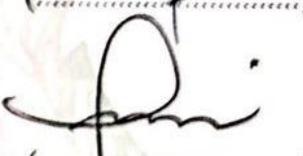
## HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Fani Fahira  
NIM : H051181324  
Program Studi : Statistika  
Judul Skripsi : Penerapan Model Regresi Logistik *Least Absolute Shrinkage And Selection Operator* Pada *Credit Scoring*

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

### DEWAN PENGUJI

1. Ketua : Anisa, S.Si., M.Si. 
2. Sekretaris : Dr. Nirwan, M.Si. 
3. Anggota : Andi Kresna Jaya, S.Si, M.Si. 
4. Anggota : Dra. Nasrah Sirajang, M.Si. 

Ditetapkan di : Makassar

Tanggal : 06 Desember 2022

## KATA PENGANTAR

### *Assalamu'alaikum Warahmatullahi Wabarakatuh*

Segala puji hanya milik Allah *Subhanallahu Wa Ta'ala* atas segala limpahan rahmat dan hidayah-Nya yang telah diberikan kepada penulis sampai saat ini. Shalawat serta salam senantiasa tercurahkan kepada baginda Rasulullah *Shallallahu 'Alaihi Wa sallam. Alhamdulillahirobbil'aalamiin*, berkat rahmat dan kemudahan yang diberikan oleh Allah *Subhanallahu Wa Ta'ala*, penulis dapat menyelesaikan tugas akhir ini.

Penulis tidak akan sampai pada titik ini tanpa dukungan dan bantuan dari pihak yang selalu ada, peduli dan menyayangi penulis. Oleh karena itu, penulis haturkan rasa terima kasih yang setulus-tulusnya serta penghargaan yang setinggi-tingginya kepada keluarga terutama orang tua penulis, Ayahanda **Aris** dan Ibunda **Nurfaedah** yang telah memberikan dukungan penuh, pengorbanan, kesabaran hati, cinta dan kasih sayang, serta dengan ikhlas telah mengiringi setiap langkah penulis dengan doa dan restunya. Teruntuk kedua kakak ku **Muh. Fadli** dan **Nurfajri**, terima kasih telah menjadi kakak yang sangat baik dan selalu ada.

Penghargaan yang tulus dan ucapan terima kasih dengan penuh keikhlasan juga penulis ucapkan kepada:

1. **Bapak Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.**, selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.
2. **Bapak Dr. Eng. Amiruddin**, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.
3. **Ibu Dr. Nurtiti Sunusi S.Si., M.Si.**, selaku Ketua Departemen Statistika, segenap Dosen Pengajar dan Staf yang telah membekali ilmu dan kemudahan kepada penulis dalam berbagai hal selama menjadi mahasiswa di Departemen Statistika.

4. **Ibu Anisa, S.Si., M.Si.**, selaku Pembimbing Utama dan **Bapak Dr. Nirwan, M.Si.**, selaku Pembimbing Pendamping yang dengan penuh kesabaran telah meluangkan waktu dan pemikirannya di tengah berbagai kesibukan dan prioritasnya untuk senantiasa memberikan arahan, dorongan, dan motivasi kepada penulis mulai dari awal hingga selesainya penulisan tugas akhir ini.
5. **Bapak Andi Kresna Jaya S.Si., M.Si.**, dan **ibu Dra. Nasrah Sirajang, M.Si.**, selaku Tim Penguji yang telah memberikan kritikan yang membangun dalam penyempurnaan penyusunan tugas akhir ini serta waktu yang telah diberikan kepada penulis.
6. **Bapak Dr. Nirwan, M.Si.**, selaku Penasehat Akademik penulis. Terima kasih atas segala bantuan, nasehat serta motivasi yang selalu diberikan kepada Penulis selama menjalani pendidikan di Departemen Statistika.
7. Teman-teman **Statistika 2018**, terima kasih untuk segala kebersamaan yang tak terlupakan selama 4 tahun, kerja sama, suka dan duka serta dukungan yang telah diberikan kepada penulis setiap kali penulis menghadapi hambatan dalam menyelesaikan tugas akhir ini.
8. Teman-teman **Dafujmin** dan **LTB** yang selalu ada setiap kali penulis butuh bantuan, terima kasih untuk dukungan dan bantuannya selama ini.
9. Teman-teman **Netijen, Rifka Yulia Sari Ifadah Latif, Claudian T Tangdilomban, Nurul Hijrah, Ika Pratiwi Haya, Reni Roihanah, Emi Astuti, Putri Aprilia S,** dan **Hafis Dwi Fernando**, yang selalu kebersamaan sejak awal perkuliahan, dan banyak membantu segala urusan perkuliahan, serta senantiasa menjadi tempat bertukar pikiran. Serta teman-teman **Koperasi**.
10. Terima kasih untuk diri sendiri, yang telah berjuang dan selalu kembali bangkit setelah menghadapi banyak kesulitan dan tantangan hidup, terima kasih karena tidak menyerah, *u did it*.
11. Semua pihak yang tidak dapat penulis sebutkan satu-persatu, semoga segala dukungan dan partisipasi yang diberikan kepada penulis bernilai ibadah disisi Allah *Subhanallahu Wa Ta'ala*.

Penulis menyadari bahwa masih banyak kekurangan dalam tugas akhir ini, untuk itu dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga tulisan ini memberikan manfaat untuk pembaca.

*Wassalamu'alaikum Warahmatullahi Wabarakatuh*

Makassar, 06 Desember 2022



Fani Fahira

**PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK  
KEPENTINGAN AKADEMIK**

---

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama : Fani Fahira  
NIM : H051181324  
Program Studi : Statistika  
Departemen : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam  
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non-eksklusif** (*Non-exclusive Royalty- Free Right*) atas tugas akhir saya yang berjudul:

**“Penerapan Model Regresi Logistik *Least Absolute Shrinkage And Selection Operator* Pada *Credit Scoring*”**

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar pada tanggal, 06 Desember 2022

Yang menyatakan

  
(Fani Fahira)

## ABSTRAK

*Credit scoring* merupakan suatu metode berbasis analisis statistika yang digunakan untuk mengukur besaran resiko kredit. Metode klasifikasi debitur yang paling populer digunakan untuk *credit scoring* adalah regresi logistik. Regresi logistik digunakan untuk memprediksi variabel respon yang biner dengan satu atau lebih variabel prediktor. Regresi logistik mempunyai keterbatasan yaitu jika terdapat korelasi yang tinggi antar variabel prediktor membuat model regresi yang didapat menjadi tidak lagi efisien karena nilai standar error koefisien regresi menjadi sangat besar atau dengan kata lain mengurangi akurasi dari estimasi. Oleh karena itu, diusulkan metode *Least Absolute Shrinkage and Selection Operator (LASSO)* untuk mengatasi hal tersebut. *LASSO* akan menyusutkan koefisien parameter menjadi nol atau mendekati nol sehingga menghasilkan model akhir yang lebih representatif. Berdasarkan model regresi logistik *LASSO* terdapat dua variabel prediktor yang terseleksi dan diperoleh lima variabel yang mempengaruhi status debitur yaitu usia, pekerjaan, jumlah pinjaman, jangka waktu pinjaman, dan jumlah tanggungan serta debitur akan memiliki skor kredit maksimum jika debitur tersebut berusia kurang dari 25 tahun, bekerja sebagai profesional, memiliki jumlah pinjaman 500-650 juta, dengan jangka waktu pinjaman lebih dari 15 tahun, serta jumlah tanggungan kurang dari 2 orang.

**Kata Kunci:** *Credit scoring* , Regresi logistik, *Least Absolute Shrinkage and Selection Operator*

***ABSTRACT***

Credit scoring is a method based on statistical analysis used to measure the amount of credit risk. The most popular debtor classification method used for credit scoring is logistic regression. Logistic regression is used to predict a response variable that is binary with a set of predictor variables. Logistic regression has limitations, namely if there is a high correlation between the independent variables it makes the regression model obtained no longer efficient because the standard error value of the regression coefficient becomes very large or in other words reduces the accuracy of the estimate. Therefore, the Least Absolute Shrinkage and Selection Operator (*LASSO*) method is proposed to overcome this. *LASSO* will shrink the parameter coefficients to zero or close to zero to produce a more representative final model. Based on the *LASSO* logistic regression model, there are two predictor variables selected and five variables are obtained that affect the debtor's status, those are age, occupation, loan amount, loan term, and number of dependents and the debtor will have a maximum credit score if the debtor is less than 25 years old, work as a professional, have a loan amount of 500-650 million, with a loan term of more than 15 years, and less than 2 people.

**Keywords:** Credit scoring , Logistic regression, Least Absolute Shrinkage and Selection Operators

## DAFTAR ISI

HALAMAN JUDUL.....	ii
LEMBAR PERNYATAAN KEOTENTIKAN.....	iii
HALAMAN PERSETUJUAN PEMBIMBING .....	iv
HALAMAN PENGESAHAN.....	v
KATA PENGANTAR .....	vi
PERNYATAAN PERSETUJUAN PUBLIKASI .....	ix
ABSTRAK .....	x
<i>ABSTRACT</i> .....	xi
DAFTAR ISI.....	xii
DAFTAR GAMBAR .....	xiv
DAFTAR TABEL.....	xv
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian .....	3
1.4 Batasan Penelitian.....	3
1.5 Manfaat Penelitian .....	3
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>4</b>
2.1 Analisis Regresi Logistik.....	4
2.2 <i>Least Absolute Shrinkage and Selection Operator (LASSO)</i> .....	5
2.3 Diskritisasi ( <i>Binning</i> ).....	6
2.4 <i>Weight of Evidence (WoE)</i> .....	6
2.5 Validasi Silang ( <i>Cross Validation</i> ).....	7
2.6 <i>Credit Scoring</i> .....	8
2.7 Kredit Kepemilikan Rumah (KPR) .....	9
<b>BAB II METODOLOGI PENELITIAN .....</b>	<b>11</b>
3.1 Jenis dan Sumber Data .....	11
3.2 Variabel Penelitian .....	11
3.3 Proses Analisis Data .....	12

<b>BAB IV PEMBAHASAN</b> .....	<b>14</b>
4.1 Estimasi Koefisien Regresi Logistik dengan Metode <i>Least Absolute Shrinkage and Selection Operator</i> .....	14
4.2 Penerapan Metode LASSO Pada <i>Credit Scoring</i> .....	17
4.2.1 Eksplorasi Data.....	17
4.2.2 WoE ( <i>Weight Of Evidence</i> ).....	20
4.2.3 Pemodelan Regresi Logistik <i>LASSO</i> .....	23
4.2.4 Interpretasi Model.....	24
4.2.5 <i>Score Credit</i> .....	26
<b>BAB V PENUTUP</b> .....	<b>28</b>
4.1 Kesimpulan.....	28
4.2 Saran.....	28
<b>DAFTAR PUSTAKA</b> .....	<b>29</b>
<b>LAMPIRAN</b> .....	<b>31</b>

## DAFTAR GAMBAR

<b>Gambar 4. 1</b> Karakteristik debitur berdasarkan status pembayaran .....	17
<b>Gambar 4. 2</b> Persentase debitur yang mengalami lunas dan gagal bayar untuk variabel prediktor usia dan pekerjaan. ....	18
<b>Gambar 4. 3</b> Persentase debitur yang mengalami lunas dan gagal bayar untuk variabel jenis agunan dan jumlah pinjaman. ....	18
<b>Gambar 4. 4</b> Persentase debitur yang mengalami lunas dan gagal bayar untuk variabel prediktor jangka waktu pinjam dan suku bunga.....	19
<b>Gambar 4. 5</b> Persentase debitur yang mengalami lunas dan gagal bayar untuk variabel prediktor jumlah tanggungan. ....	19
<b>Gambar 4. 6</b> <i>Cross validasion 10-fold</i> parameter penyusutan $\log(\lambda)$ pada model <i>LASSO</i> .....	23

**DAFTAR TABEL**

<b>Tabel 4. 1</b> Nilai <i>Odds Ratio</i> dan WoE Usia .....	20
<b>Tabel 4. 2</b> Nilai <i>Odds Ratio</i> dan WoE Pekerjaan .....	20
<b>Tabel 4. 3</b> Nilai <i>Odds Ratio</i> dan WoE Jenis Agunan .....	21
<b>Tabel 4. 4</b> Nilai <i>Odds Ratio</i> dan WoE Jumlah Pinjaman .....	21
<b>Tabel 4. 5</b> Nilai <i>Odds Ratio</i> dan WoE Jangka Waktu Pinjaman .....	21
<b>Tabel 4. 6</b> Nilai <i>Odds Ratio</i> dan WoE Suku Bunga .....	22
<b>Tabel 4. 7</b> Nilai <i>Odds Ratio</i> dan WoE Tanggungan .....	22
<b>Tabel 4. 8</b> Nilai <i>Odds Ratio</i> .....	24
<b>Tabel 4. 9</b> <i>Score Credit</i> .....	26

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Kredit adalah fasilitas keuangan yang digunakan seseorang maupun badan usaha untuk meminjam uang kemudian membayarnya kembali dalam jangka waktu yang telah ditentukan. Menurut UU No. 10 Tahun 1998, kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu berdasarkan persetujuan atau kesepakatan antara bank dan pihak lain yang mewajibkan pihak peminjam melunasi hutangnya setelah waktu dengan pemberian bunga yang telah ditentukan. Menurut hasil survei perbankan Bank Indonesia pertumbuhan kredit pada tahun 2022 akan mencapai 8,7% *year of the year* (yoy) atau meningkat dibandingkan dengan pertumbuhan pada tahun 2021 yaitu sebesar 5,2% (yoy). Optimisme tersebut antara lain didorong oleh kondisi moneter dan kondisi ekonomi yang diperkirakan semakin baik, serta relatif terjaganya risiko penyaluran kredit (BI, 2021).

Peningkatan permintaan kredit ini tak lantas membuat pihak jasa layanan kredit akan dapat mengabulkan semua permohonan yang ada. Untuk itu, perlu dilakukan suatu proses penyeleksian dalam rangka melihat calon debitur mana yang layak diberi pinjaman. Proses tersebut disebut *credit scoring* yang merupakan suatu proses penilaian atau *scoring* menggunakan data historis debitur berupa karakteristik-karakteristik kreditur yang dimiliki oleh instansi keuangan yang memberikan kreditnya untuk kemudian diklasifikasikan apakah layak diberi pinjaman atau tidak. *Credit score* adalah angka yang akan merepresentasikan kelayakan seorang debitur untuk menerima dana. Calon debitur yang lolos seleksi disebut *good* debitur, dan permohonan kreditnya dikabulkan. Sebaliknya, calon debitur yang tidak lolos seleksi disebut *bad* debitur, dan permohonan kreditnya akan ditolak (Kantardzic, 2003).

Metode yang sudah digunakan dalam analisis *credit scoring* diantaranya analisis regresi logistik yang merupakan salah satu metode yang dapat digunakan untuk mencari hubungan antara variabel respon yang bersifat dikotomis dengan satu atau lebih variabel prediktor yang bersifat kontinu atau kategorik. Berdasarkan

penelitian terdahulu, Tabagari (2015) menggunakan regresi logistik untuk analisis *credit scoring* namun menerapkan prosedur *forward selection* sehingga secara otomatis mendapatkan hasil model yang signifikan. Serta, penelitian yang dilakukan Aziz (2020) juga melakukan *credit scoring* dengan regresi logistik namun menggunakan bantuan metode penduga parameter *Newton Raphson* untuk meningkatkan kinerja regresi logistik kemudian melakukan teknik ensemble menggunakan bagging. Oleh karenanya, analisis *credit scoring* dengan menggunakan regresi logistik masih kurang efektif karena membutuhkan uji hipotesis dan uji asumsi yang harus terpenuhi. Serta, analisis regresi logistik juga akan memiliki standar error yang sangat besar jika terjadi kolinieritas antara variabel prediktor maka model yang dihasilkan menjadi tidak efisien.

Untuk mengatasi hal tersebut maka digunakan metode analisis regresi logistik *Least Absolute Shrinkage Selection and Operator (LASSO)* yang mampu menyusutkan koefisien (parameter  $\beta$ ) tepat nol atau mendekati nol sehingga menghasilkan ragam parameter  $\beta$  yang lebih kecil dan model lebih mudah diinterpretasikan dibandingkan dengan regresi linier sederhana maupun regresi logistik (Tibshirani, 1996). Kelebihan regresi *LASSO* yaitu dapat digunakan sebagai seleksi variabel prediktor pada model, sehingga hanya variabel-variabel terbaik yang masuk kedalam model. Hal ini juga bermanfaat untuk mempermudah dalam menginterpretasikan model regresi. Sehingga *LASSO* merupakan solusi dari masalah interpretasi maupun multikolinieritas yang sering terjadi pada analisis regresi. Berdasarkan penelitian yang dilakukan oleh Andana, *et.all* (2017) terkait Regresi *Least Absolute Shrinkage Selection and Operator* yaitu penelitian dengan fokus penelitian untuk mengatasi multikolinieritas menggunakan regresi *LASSO* menghasilkan penyusutan koefisien *LASSO* tepat pada nol mengakibatkan model regresi menjadi lebih sederhana sehingga model regresi mudah untuk diinterpretasikan. Oleh karena itu, dalam fokus penelitian ini akan membahas mengenai “**Penerapan Model Regresi Logistik *Least Absolute Shrinkage And Selection Operator* Pada *Credit Scoring*”.**

## 1.2 Rumusan Masalah

Berdasarkan latar belakang, rumusan masalah dari penelitian ini sebagai berikut:

1. Bagaimana membangun model regresi logistik *Least Absolute Shrinkage And Selection Operator* dalam analisis *credit scoring*?
2. Variabel-variabel apa saja yang berpengaruh dalam analisis *credit scoring* menggunakan metode regresi logistik *Least Absolute Shrinkage and Selection Operator*?

## 1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan, maka tujuan dari penelitian ini sebagai berikut:

1. Untuk memperoleh bentuk model regresi logistik *Least Absolute Shrinkage And Selection Operator* dalam analisis *credit scoring*.
2. Untuk memperoleh variabel-variabel yang berpengaruh dalam analisis *credit scoring* menggunakan metode regresi logistik *Least Absolute Shrinkage And Selection Operator*.

## 1.4 Batasan Penelitian

Penelitian ini difokuskan hanya terbatas pada pembahasan model regresi logistik *Least Absolute Shrinkage And Selection Operator* untuk analisis *credit score* dan data yang digunakan yaitu data nasabah bank mandiri untuk pengajuan Kredit Kepemilikan Rumah tahun 2022. Pada penelitian ini tidak menentukan besar batas nilai pada kartu skor karena penentuan batasan nilai ditentukan oleh besarnya keuntungan dan kerugian yang hanya diketahui oleh pihak bank.

## 1.5 Manfaat Penelitian

Diharapkan penelitian ini dapat memberikan informasi ilmiah kepada pembaca mengenai model regresi logistik *Least Absolute Shrinkage And Selection Operator* dan penerapannya dalam kehidupan nyata, serta sebagai gambar untuk pemberi kredit dalam menentukan apakah pengajuan kredit dari peminjam dapat disetujui atau tidak.

## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Analisis Regresi Logistik

Regresi Logistik adalah suatu metode analisis statistika untuk mendeskripsikan hubungan antara variabel terikat yang memiliki dua kategori atau lebih dengan satu atau lebih variabel bebas berskala kategori atau kontinu. Adapun regresi logistik dapat dibagi menjadi regresi logistik biner, regresi logistik multinomial dan regresi logistik ordinal. Menurut Tampil *et al.*, (2017) Model regresi logistik merupakan salah satu metode yang dapat digunakan untuk mencari hubungan antara variabel respon yang bersifat dikotomis dengan satu atau lebih variabel prediktor yang berskala kategorik atau numerik. Bentuk model fungsi regresi logistik dengan  $p$  variabel sebagai berikut:

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)}$$

model regresi logistik secara umum dapat dituliskan dalam bentuk:

$$n(x) = \frac{\exp(g(x))}{1 + \exp(g(x))}$$

Fungsi  $\pi(x)$  merupakan fungsi nonlinier sehingga perlu dilakukan transformasi logit untuk memperoleh fungsi linier yang dapat menggambarkan hubungan antara variabel respon dan variabel prediktor. Sehingga menjadi

$$g(x) = \ln \left[ \frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \quad (2.1)$$

dengan:

$\beta_0$  = Konstanta model regresi logistik

$\beta_k$  = Koefisien regresi logistik, untuk  $k = 1, 2, \dots, p$

$X_k$  = Variabel prediktor, untuk  $k = 1, 2, \dots, p$

$p$  = banyaknya variabel prediktor.

## 2.2 Least Absolute Shrinkage and Selection Operator (LASSO)

Metode *Least Absolute Shrinkage and Selection Operator (LASSO)* diperkenalkan pertama kali oleh Tibshirani pada tahun 1996 yang diperoleh dengan cara pemrograman kuadratik. *Least Absolute Shrinkage and Selection Operator (LASSO)* bertujuan memperbaiki pendugaan metode regresi linier sederhana. Pada pendugaan regresi linier sederhana sering ditemui pendugaan yang tidak bias tetapi mempunyai ragam yang besar sehingga mempengaruhi ketepatan prediksi dan susah interpretasinya. Adapun metode *LASSO* mampu menyusutkan koefisien (parameter  $\beta$ ) tepat nol atau mendekati nol sehingga menghasilkan ragam parameter  $\beta$  yang lebih kecil dan model lebih mudah diinterpretasikan dibandingkan dengan regresi linier sederhana. Penduga parameter dengan metode *LASSO*, sebagai berikut (Tibshirani,1996):

$$\hat{\beta} = \arg \min \left[ \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{k=1}^p \beta_k x_{ik} \right)^2 \right] \quad (2.2)$$

dengan

- $y_i$  : nilai variabel respon ke-i.
- $\beta_0$  : konstanta regresi *LASSO*.
- $\beta_k$  : koefisien regresi logistik *LASSO* variabel prediktor ke-k.
- $x_{ik}$  : pengamatan ke-i variabel prediktor ke-k.
- $n$  : banyaknya pengamatan dengan  $i = 1,2,3, \dots, n$ .
- $p$  : banyaknya variabel prediktor dengan  $k = 1,2,3, \dots, p$ .

Dalam bentuk persamaan *Lagrange*, dugaan koefisien regresi dengan menggunakan *LASSO*, sebagai berikut :

$$\hat{\beta}^{Lasso} = \arg \min \left[ \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{k=1}^p \beta_k x_{ik} \right)^2 + \lambda \sum_{k=1}^p |\hat{\beta}_k| \right] \quad (2.3)$$

$\lambda$  adalah parameter yang mengendalikan besarnya koefisien *LASSO* yang diatur dengan kendala  $\sum_{k=1}^p |\hat{\beta}_k| \leq t$ . Nilai  $t$  merupakan suatu besaran yang mengendalikan besarnya penyusutan pada pendugaan koefisien *LASSO* dengan  $t >$

0. Jika  $\widehat{\beta}_k$  merupakan penduga kuadrat terkecil dan  $t_0 = \sum_{k=1}^p \beta_k$ , maka nilai  $t < t_0$  akan menyebabkan solusi penduga kuadrat terkecil menyusut ke arah nol, dan memungkinkan beberapa koefisien tepat nol. Jika nilai  $t$  yang dipilih lebih besar atau sama dengan daripada  $t_0$ , maka penduga *LASSO* memberikan hasil yang sama dengan penduga kuadrat terkecil.

Nilai  $\lambda > 0$  juga merupakan parameter tuning yang mampu mengendalikan koefisien  $\hat{\beta}^{Lasso}$  yang dapat diperoleh dengan menggunakan proses *cross validation k-fold* (Hastie, *et.al*, 2008).

### 2.3 Diskritisasi (*Binning*)

Pada penelitian ini penulis menggunakan data numerik sehingga perlu dilakukan diskritisasi sehingga data berskala numerik dapat menjadi data kategorik. Proses diskritisasi merupakan proses transformasi data kuantitatif ke dalam data kualitatif. Teknik ini digunakan untuk mereduksi jumlah nilai suatu variabel yang berskala numerik atau kontinu dengan cara memilah-milah selang nilai data variabel ke dalam sub-sub selang nilai. Secara umum, proses diskritisasi terdiri dari 4 tahapan, yaitu:

1. Mengurutkan nilai yang akan didiskritisasi.
2. Melakukan pemisahan atau penyatuan selang nilai.
3. Menghentikan proses pada titik tertentu.

*Binning* memetakan nilai-nilai sebuah variabel ke dalam satu gugus bin. Sebuah bin bisa terdiri dari satu nilai saja, selang kontinu, data hilang, atau bahkan nilai yang tidak ada sebelumnya. Label selang nantinya digunakan untuk menggantikan nilai data aktual (Alfiansyah, 2008).

### 2.4 *Weight of Evidence* (WoE)

*Weight of Evidence* (WoE) merupakan perbandingan proporsi kategori tertentu suatu variabel untuk kelompok atribut atau variabel yang menunjukkan status kelas kredit seseorang atau nasabah. Variabel status tersebut merupakan variabel target dalam proses membangun model *credit scoring* model. Data status tersebut biasanya berbentuk biner yaitu *good* dan *bad*. Status *good* bisa didefinisikan sebagai nasabah yang tidak pernah lalai membayar sedangkan *bad* bisa

didefinisikan sebagai nasabah yang lalai atau pailit. WoE didefinisikan sebagai berikut (Hollowed, 2014):

WoE merupakan selisih atau besarnya perbedaan antara log odds tiap bin (kategori) dengan log odds total. Dalam proses membangun *credit scoring* model, WoE berfungsi menunjukkan tingkat resiko seseorang.

$$WoE(i) = \log\left(\frac{f_G(i)}{f_B(i)}\right) \quad (2.5)$$

dengan:

$$f_G(i) = 100 \left(\frac{n_{Gi}}{n_G}\right) = \text{Presentasi kategori ke-}i \text{ pada individu } good.$$

$$f_B(i) = 100 \left(\frac{n_{Bi}}{n_B}\right) = \text{Presentasi kategori ke-}i \text{ pada individu } bad.$$

## 2.5 Validasi Silang (*Cross Validation*)

Menurut Nurwahidah *et al.*, (2015) Validasi silang merupakan metode yang dapat digunakan untuk mengestimasi galat prediksi dalam meningkatkan ketepatan dari pemilihan model. Validasi silang membagi data menjadi dua bagian, yaitu *data training* dan *data test*. *Data training* digunakan untuk menentukan nilai  $\hat{\beta}$  atau untuk membangun model, sedangkan *data test* digunakan untuk menguji kebaikan prediksi model. Salah satu metode tipe validasi silang adalah *k-fold* dengan persamaan berikut:

$$CV = \frac{1}{k} \sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n} \quad (2.6)$$

dengan:

$\hat{y}_i$  : Nilai prediksi variabel respon ke-*i*.

*y* : Variabel respon ke-*i*

*n* : Jumlah data

*k* : Jumlah *k-fold*

Proses validasi *k-folds* silang akan menghasilkan prediksi sebanyak *k* dari kesalahan uji  $MSE_1, MSE_2, \dots, MSE_k$ . Keuntungan menggunakan proses validasi dengan *5-fold* atau *10-fold* validasi silang karena akan menghasilkan ragam rendah.

## 2.6 Credit Scoring

*Credit Scoring* merupakan salah satu metode yang digunakan untuk menilai kelayakan dari calon nasabah kredit. Metode *credit scoring* merupakan salah satu metode statistika yang digunakan untuk memprediksi peluang calon nasabah tidak sanggup membayar kembali pinjaman, ataupun bagi nasabah kredit yang sudah terdaftar yang berpeluang menjadi nasabah yang pembayaran pinjamannya menunggak (Hakim, 2019). Model *Credit Scoring* dibangun berdasarkan informasi data historis di masa lalu yakni mengenai informasi data diri, pengalaman kredit seperti riwayat pembayaran tagihan seseorang, telat membayar, cicilan-cicilan, dan umur rekening (*account*), dikumpulkan dari aplikasi kredit dan laporan kredit. Penggunaan metode *credit scoring* ini dapat memberikan informasi status kredit dari seorang calon kreditur hingga penentuan persetujuan personal loan, seperti pembuatan kartu kredit, KPR (Kredit Pemilikan Rumah), KTA (kredit tanpa anjungan), kredit modal usaha, dan lain sebagainya pemberi kredit atau pinjaman untuk menentukan apakah pengajuan kredit atau pinjaman dari peminjam dapat disetujui atau tidak.

Model penskoran kredit menghasilkan suatu kartu skor (*scorecard*), yang berisikan nilai skor di setiap kategori variabel prediktor. Kelebihan dari penskoran kredit tidak hanya untuk bank saja tetapi juga untuk nasabah. Penskoran kredit menghasilkan suatu hasil perhitungan statistik dari setiap kategori pada setiap variabel prediktor yang dapat digunakan untuk memisahkan kategori suatu nasabah ‘*Good*’ atau ‘*bad*’. Saat pembentukan kartu skor diperlukan adanya teknik penskalaan. Teknik ini mengacu pada jangkauan dan format skor dalam kartu skor. Teknik penskalaan ini tidak berpengaruh terhadap tingkat prediksi kartu skor namun digunakan agar pengguna lebih mudah memahami data yang bertipe diskret. Penskalaan dihitung dengan menggunakan persamaan berikut (Siddiqi, 2006):

$$Score = offset + (factor) \ln(odds) \quad (2.7)$$

Nilai *factor* dan *offset* dapat diperoleh jika telah didefinisikan :

1. Nilai skor yang diinginkan untuk odds tertentu.
2. Nilai pdo (*points to double the odds*), yaitu besarnya kenaikan skor yang menyebabkan odds-nya menjadi dua kali lipat.

sehingga

$$Score = offset + (factor) \ln(odds)$$

$$Score + pdo = offset + (factor) \ln(2 odds)$$

$$pdo = offset + (factor) \ln(2 odds) - offset + (factor) \ln(odds)$$

$$pdo = \ln(2) (factor) = 0,693 factor$$

$$factor = 1,443 pdo$$

maka

$$Offset = target skor - (faktor \times \ln(odds)) \quad (2.8)$$

Perhitungan skor untuk setiap kategori pada satu variabel prediktor, disajikan sebagai berikut (Ardita, 2014):

$$Skor kredit = \left( WOE \times \beta_p + \frac{\beta_0}{p} \right) factor + \frac{offset}{p} \quad (2.9)$$

dengan:

*WOE* = Nilai *WOE* pada setiap kategori variabel prediktor.

$\beta_k$  = Koefisien regresi *LASSO* untuk setiap variabel, untuk  $k = 1, 2, \dots, p$

$\beta_0$  = konstanta pada regresi *LASSO*.

$p$  = Banyaknya variabel bebas.

## 2.7 Kredit Kepemilikan Rumah (KPR)

Kredit menurut Undang-Undang perbakan Nomor 10 tahun 1998 adalah Penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan atau kesepakatan pinjaman meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam melunasi utangnya setelah jangka waktu tertentu dengan pemberian bunga. Sebelum kredit diberikan, bank terlebih dahulu mengadakan analisis kredit. Tujuan analisis ini adalah agar bank yakin bahwa kredit

yang diberikan benar- benar aman dalam arti uang yang disalurkan pasti kembali.Salah satu kredit yang ditawarkan oleh jasa perbankan adalah Kredit Kepemilikan Rumah (KPR) (Aziz, 2020).

Bank Indonesia memberikan definisi terkait KPR (Kredit Pemilikan Rumah) yang merupakan suatu fasilitas pembiayaan yang diberikan oleh perbankan kepada para nasabah perorangan yang akan membeli atau memperbaiki rumah. KPR atau Kredit Kepemilikan Rumah merupakan salah satu jenis pelayanan kredit yang diberikan oleh bank kepada para nasabah yang menginginkan pinjaman khusus untuk memenuhi kebutuhan dalam pembangunan rumah atau renovasi rumah. KPR sendiri muncul karena adanya kebutuhan memiliki rumah yang semakin lama semakin tinggi tanpa diimbangi daya beli yang memadai oleh masyarakat.

Seperti layaknya produk perbankan yang memiliki keanekaragaman jenis, KPR secara umum dibagi menjadi 2 jenis, yaitu:

1. KPR Subsidi adalah suatu kredit yang diperuntukkan kepada masyarakat yang memiliki kemampuan ekonomi menengah ke bawah. Adapun bentuk dari subsidi ini telah diatur oleh pemerintah, sehingga tidak semua masyarakat dapat mengajukan kredit jenis ini. Secara umum batasan yang ditetapkan oleh pemerintah dalam memberikan subsidi adalah penghasilan pemohon dan maksimum kredit yang diberikan.
2. KPR non Subsidi adalah suatu KPR yang diperuntukkan bagi seluruh masyarakat tanpa adanya campur tangan pemerintah.

Ketentuan KPR ditetapkan oleh bank itu sendiri sehingga penentuan besarnya suku bunga sesuai dengan kebijakan bank (Fatmasari,2013).