

Prediksi Calon Nasabah Gadai Potensial pada PT. Pegadaian (Persero) Menggunakan *Support Vector Machine* dengan Algoritma Genetika



Oleh:

Muhammad Abdul Aziz

24050119130045

Dosen Pembimbing I

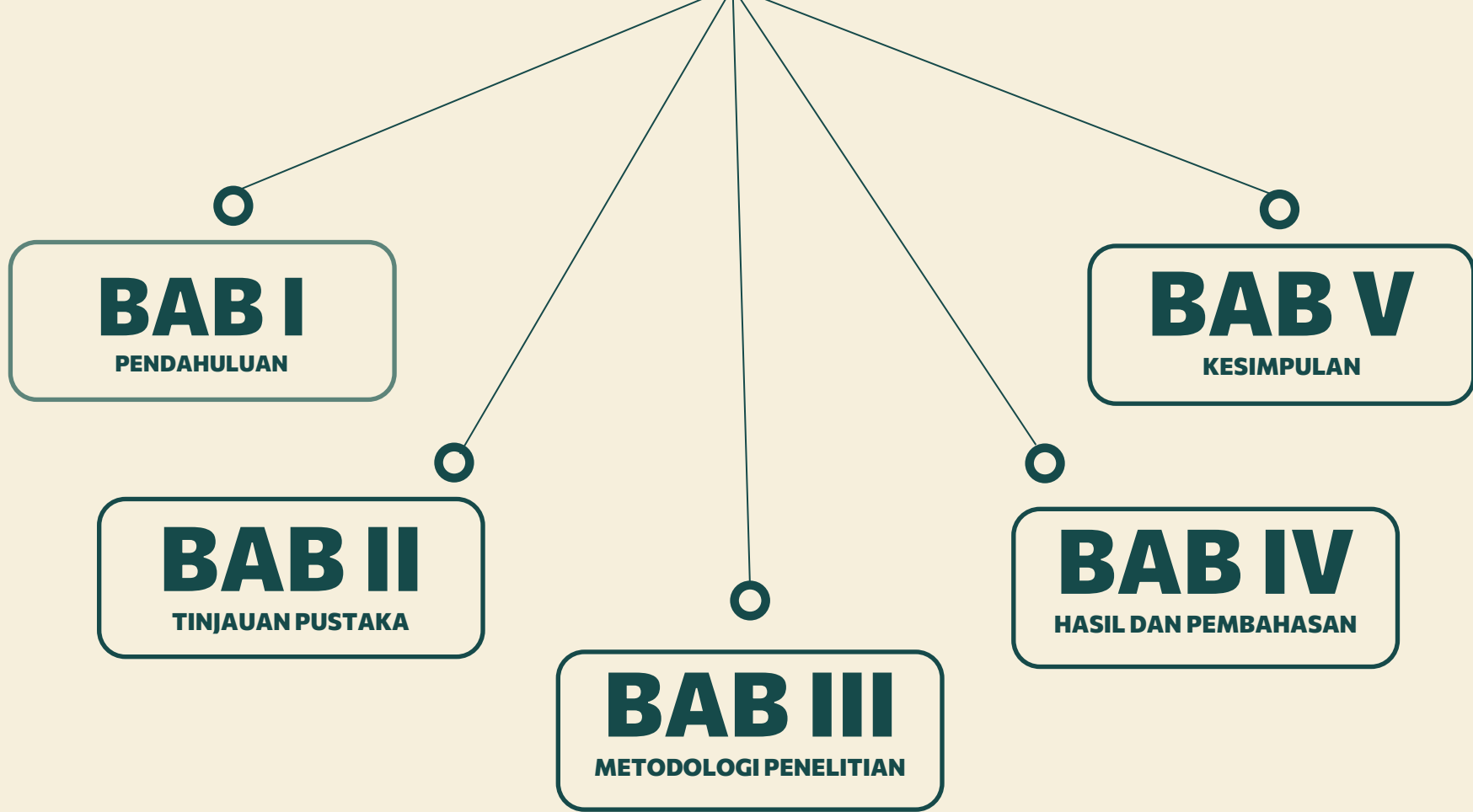
Dr. Drs. Rukun Santoso, M.Si.

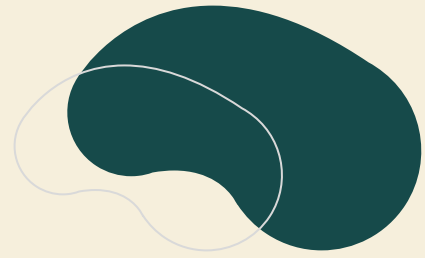
NIP. 196502251992011001

Dosen Pembimbing II

Dr. Budi Warsito, S.Si., M.Si.

NIP. 197508241999031003





BAB I

PENDAHULUAN

LATAR BELAKANG



Digitalisasi



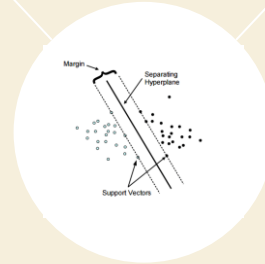
Prediksi



Algoritma
Genetika



PT Pegadaian



Support Vector Machine

LATAR BELAKANG



Digitalisasi membawa perubahan yang sangat besar khususnya perekonomian.

Program usaha mikro kecil menengah.



PT Pegadaian menyediakan pinjaman modal usaha dengan sistem gadai.

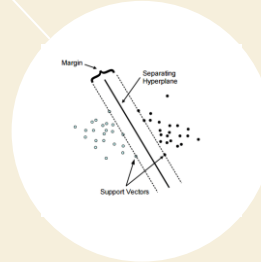
Nasabah berkewajibannya dalam mencicil kredit, memperpanjang, maupun melunasi pinjaman.

LATAR BELAKANG

Analisis prediksi diperlukan untuk mengetahui potensi nasabah.



Support Vector Machine mampu mengolah data dengan dimensi tinggi sehingga tingkat akurasi yang dihasilkan lebih baik (Siringoringo & Jamaludin, 2019)



LATAR BELAKANG



Algoritma genetika digunakan untuk mengoptimalkan fungsi yang dihasilkan dari suatu klasifikasi.

RUMUSAN MASALAH

1. Bagaimana cara penerapan metode *Support Vector Machine* dengan Algoritma Genetika (SVM-GA) untuk memprediksi nasabah potensial dan tidak potensial?

Bagaimana hasil prediksi nasabah gadai menggunakan metode *Support Vector Machine* dengan Algoritma Genetika (SVM-GA)?

2. *Machine*



BATASAN MASALAH

Database Nasabah

📍 CP Kebumen



Surat Bukti Gadai

3 Januari 2022 – 29 November 2022



TUJUAN PENELITIAN

1. Menerapkan metode *Support Vector Machine* dengan Algoritma Genetika untuk memprediksi nasabah potensial dan tidak potensial.
2. Mendapatkan hasil prediksi nasabah gadai potensial menggunakan metode *Support Vector Machine* dengan Algoritma Genetika.



BAB II

TINJAUAN PUSTAKA

PT PEGADAIAN (PERSERO)

sebuah lembaga kredit yang memperluas sistem gadai

Bank Van Leening

PT Pegadaian

1746

1901

► BUMN

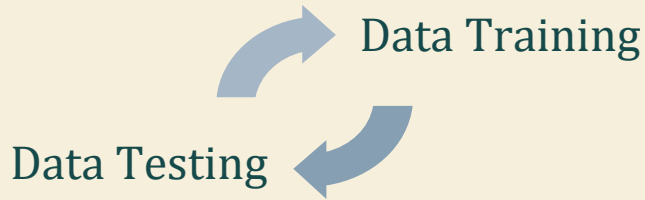
Merupakan salah satu perusahaan yang bergerak di bidang keuangan non-bank

SURAT BUKTI GADAI (SBG)

surat perjanjian utang piutang



Klasifikasi merupakan proses dari pembangunan terhadap suatu model yang mengklasifikasi suatu objek sesuai dengan atribut – atributnya (Elly, et al., 2015)



Normalisasi merupakan proses merubah nilai atribut dalam sebuah dataset menjadi skala yang berstruktur baik

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

>>Oversampling

Menambahkan data dari kelas minoritas sehingga jumlah data antar kelas sama.

>>Undersampling

Menurunkan jumlah kelas mayoritas sebanyak jumlah kelas minoritas



SUPPORT VECTOR MACHINE

SVM secara umum dilakukan dengan tujuan untuk memperoleh *hyperplane* (fungsi pemisah) terbaik sebagai pemisah dua buah kelas pada ruang input.

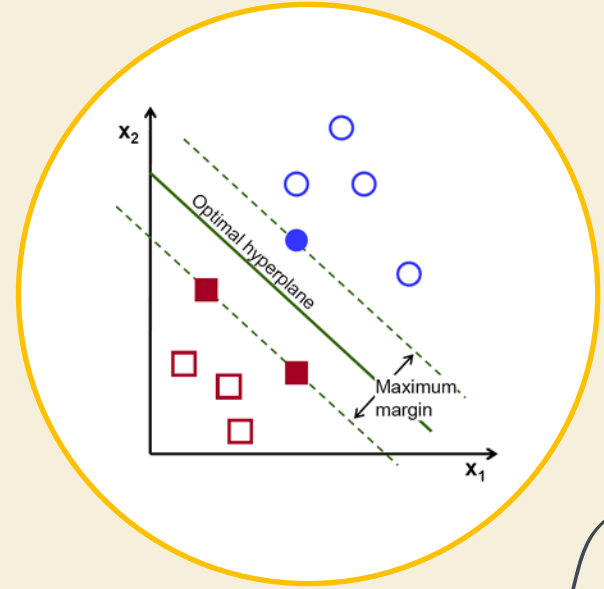
Linier Support Vector Machine

Himpunan data training dari dua kelas:

$$D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_l, y_l)\}$$

Hyperplane klasifikasi linier SVM dinotasikan:

$$f(\mathbf{x}) = \mathbf{w}^T \cdot \mathbf{x}_i + b$$



SUPPORT VECTOR MACHINE

Klasifikasi kelas data:

$$g(x) = \begin{cases} [(\mathbf{w}^T \cdot \mathbf{x}_i) + b] \geq +1, \text{ untuk } y_i = +1 \\ [(\mathbf{w}^T \cdot \mathbf{x}_i) + b] \leq -1, \text{ untuk } y_i = -1 \end{cases}$$

(Cortes & Vapnik, 1995)

atau:

$$y_i [(\mathbf{w}^T \cdot \mathbf{x}_i) + b] \geq 1, \quad \text{dengan } i = 1, 2, 3, \dots, l$$

Nilai margin:

$$\frac{1 - b - (-1 - b)}{\|\mathbf{w}\|} = \frac{2}{\|\mathbf{w}\|^2}$$

Hyperplane terbaik didapatkan dengan memaksimalkan margin atau sama dengan meminimumkan nilai $\|\mathbf{w}\|^2$

Quadratic Programming (QP) Problem

$$\min \tau(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2$$

dengan syarat

$$y_i ((\mathbf{w}^T \cdot \mathbf{x}_i) + b) - 1 \geq 0, \quad \text{dengan } i = 1, 2, 3, \dots, l$$



SUPPORT VECTOR MACHINE

Lagrange multiplier

$$\min_{\mathbf{w}, b} L(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^l \alpha_i [y_i (\mathbf{w}^T \cdot \mathbf{x}_i + b) - 1]$$

Keterangan:

α_i = Pengganda fungsi *Lagrange* ; $i = 1, 2, \dots, l$

Syarat optimal:

- Diturunkan terhadap \mathbf{w} dan b
- Gradien $L=0$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^l \alpha_i y_i = 0$$

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^l \alpha_i \mathbf{x}_i y_i = 0$$

$$L(\mathbf{w}, b, a) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^l \alpha_i y_i (\mathbf{w}^T \cdot \mathbf{x}_i) - b \sum_{i=1}^l \alpha_i y_i + \sum_{i=1}^l \alpha_i$$

Penjabaran:

$$\|\mathbf{w}\|^2 = \mathbf{w}^T \mathbf{w} = \sum_{i=1}^l \alpha_i y_i (\mathbf{w}^T \cdot \mathbf{x}_i) = \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j (\mathbf{x}_i^T \cdot \mathbf{x}_j)$$

$$L_d = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j (\mathbf{x}_i^T \cdot \mathbf{x}_j)$$



SUPPORT VECTOR MACHINE

Soft Margin

Soft margin dapat dibangun dengan menambahkan variabel kendur atau *slack variable*

$$\xi_i > 0$$

Klasifikasi kelas data:

$$y_i[(\mathbf{w}^T \cdot \mathbf{x}_i) + b] \geq 1 - \xi_i \quad ; i = 1, 2, \dots, l$$

$$\min \tau(\mathbf{w}, \xi_i) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i$$

SUPPORT VECTOR MACHINE

$$\begin{aligned} f(\mathbf{x}_j) &= \text{sign}(\mathbf{w}^T \cdot \mathbf{x}_j + b) \\ &= \text{sign}\left(\sum_{i=1}^m a_i y_i (\mathbf{x}_i \cdot \mathbf{x}_j) + b\right) \end{aligned}$$

m = Banyaknya data *Support Vector*

\mathbf{x}_i = *Support Vector*

\mathbf{x}_j = Data *testing* yang akan diprediksi

$\mathbf{x}_i \cdot \mathbf{x}_j$ = *Dot-product* antara \mathbf{x}_i dan \mathbf{x}_j

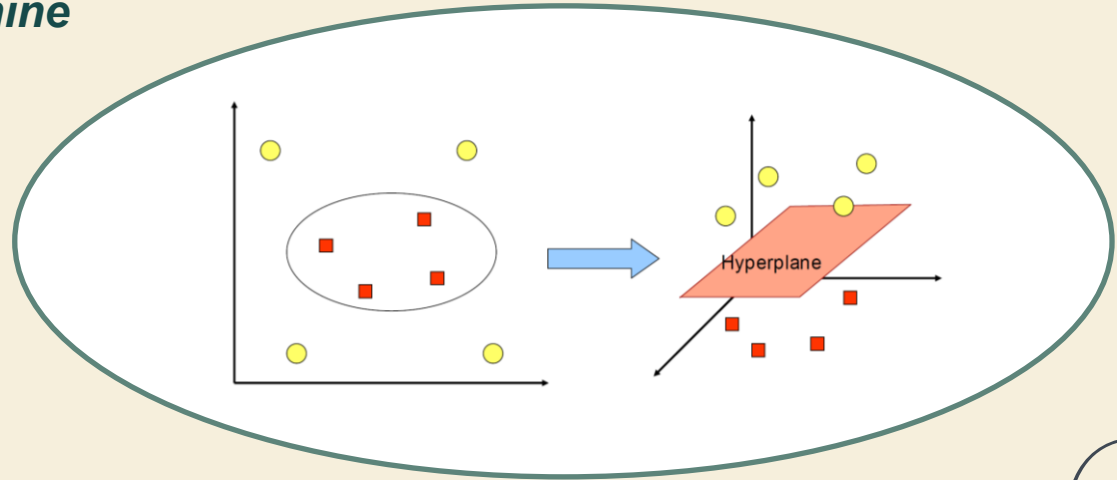


SUPPORT VECTOR MACHINE

Non-Linear Support Vector Machine

>>Kernel

Kernel merupakan fungsi yang memetakan data menjadi fungsi baru yang memiliki ruang dimensi lebih tinggi.



$$f(\mathbf{x}_j) = \text{sign} \left(\sum_{i=1}^m a_i y_i K(\mathbf{x}_i, \mathbf{x}_j) + b \right)$$

SUPPORT VECTOR MACHINE

Jenis Fungsi Kernel

>>Linier

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \cdot \mathbf{x}_j$$

>>Sigmoid

$$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(r(\mathbf{x}_i^T \mathbf{x}_j + r))$$

>>Polynomial

$$K(\mathbf{x}_i, \mathbf{x}_j) = [(\mathbf{x}_i^T \cdot \mathbf{x}_j) + 1]^d$$

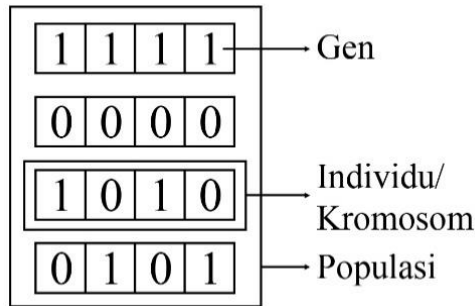
>>Radial Basis Function

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$$

ALGORITMA GENETIKA

TEORI EVOLUSI DARWIN

Kelangsungan hidup suatu makhluk dipengaruhi aturan bahwa individu yang bernilai fitness tinggi yang akan bertahan hidup sedangkan individu yang bernilai fitness rendah akan mati (Goldberg, 1989)



ALGORITMA GENETIKA

SELEKSI

Memilih kromosom untuk proses crossover dan mutasi

MUTASI

Mengganti satu gen dalam kromosom

CROSSOVER

Penyilangan pasangan kromosom

EVALUASI

Menilai kromosom

EVALUASI KINERJA KLASIFIKASI

Confusion Matrix

		Prediksi	
		Positif	Negatif
Aktual	Positif	TP	FN
	Negatif	FP	TN

TP : *True Positive* adalah proporsi positif yang ada dalam dataset yang diklasifikasi positif

FN : *False Negative* adalah proporsi positif yang ada dalam dataset yang diklasifikasi negatif

FP : *True Positive* adalah proporsi negatif yang ada dalam dataset yang diklasifikasi positif

TN : *True Negative* adalah proporsi negatif yang ada dalam dataset yang diklasifikasi negatif



EVALUASI KINERJA KLASIFIKASI

Precision

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

Accuracy

Akurasi klasifikasi
menunjukkan
performansi model
klasifikasi secara
keseluruhan

Recall

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

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



BAB III

METODOLOGI

PENELITIAN

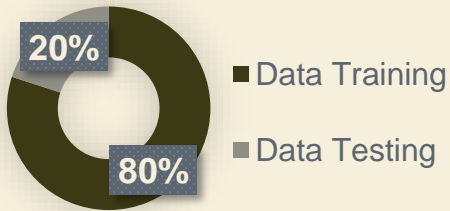


Jenis dan Sumber Data

Sekunder

Surat Bukti
Gadai (SBG)

Database
Nasabah



Variabel Penelitian

VARIABEL TERIKAT

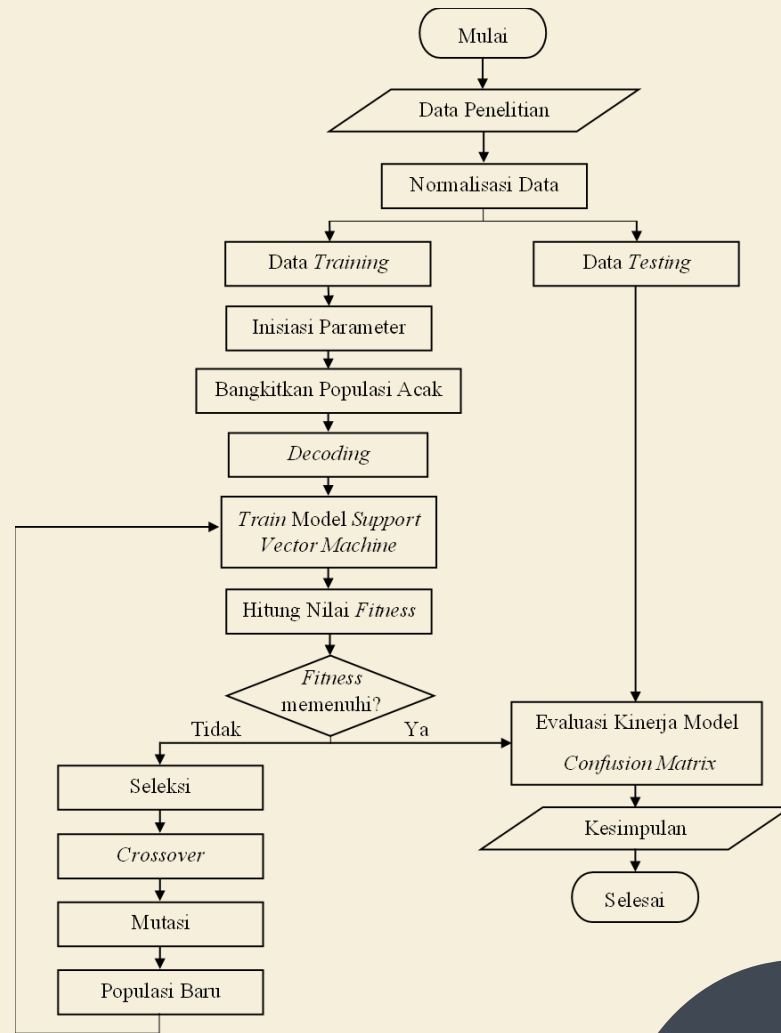
Label Kelas ◀

VARIABEL BEBAS

- ▶ Jumlah Uang Pinjaman
- ▶ Jumlah Kredit Aktif
- ▶ Jumlah Kredit Tidak Aktif
- ▶ Tingkat Pendidikan



Diagram Alir Analisis Data



BAB IV

HASIL DAN

PEMBAHASAN



DESKRIPSI DATA

Data didapat melalui akses terhadap database perusahaan, sehingga diperoleh sebanyak 3.548 data

Data yang digunakan adalah data dokumen surat bukti gadai dan database nasabah dari PT Pegadaian Kantor Cabang Kebumen

Data surat bukti gadai yang digunakan merupakan data kredit dalam kurun waktu 3 Januari 2022 – 29 November 2022

Variabel yang digunakan dalam penelitian ini adalah uang pinjaman (x_1), kredit aktif (x_2), kredit tidak aktif (x_3), dan tingkat Pendidikan (x_4)

STATISTIK DESKRIPTIF

	x_1	x_2	x_3
Min	60.000	0,000	0,000
1 st Qu.	1.140.000	0,000	0,000
Median	2.290.000	1,000	1,000
Mean	4.577.032	1,161	1,329
3 rd Qu.	4.500.000	1,000	2,000
Max	362.000.000	23,000	22,000

Variabel x_1 , x_2 , dan x_3 menyatakan bahwa uang Pinjaman, kredit aktif dan kredit tidak aktif sangat bervariasi karena memiliki rentang yang cukup jauh.



NORMALISASI DATA

x_1	x_2	x_3	x_4
0,001685362	0	1	0
0,004586396	0,217391304	0,954545455	2
0,031054871	0,130434783	0,909090909	0
0,05196994	0,652173913	0,772727273	0
0,002182682	0,652173913	0,772727273	0
...
0,008316296	0,043478261	0	3

Perhitungan:

$$x_1' = \frac{22.700.000 - 60.000}{362.000.000 - 60.000}$$

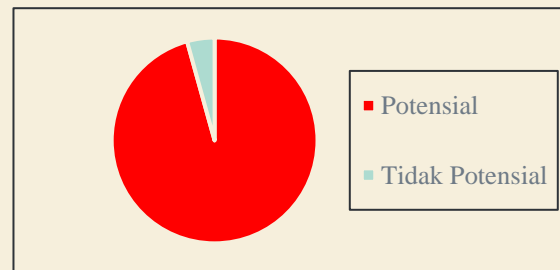
$$x_1' = \frac{22.640.000}{361.940.000} = 0,062551804$$

DATA TRAIN DAN DATA TESTING

Perbandingan terbaik untuk data *training* dan data *testing* yang digunakan adalah sebesar 80%: 20%.

RESAMPLING

Proses *resampling* digunakan karena menunjukkan adanya *imbalanced class data*



SUPPORT VECTOR MACHINE

Nilai *confusion matrix* yang dihasilkan menggunakan SVM *kernel* RBF:

		Prediksi	
		Positif	Negatif
Aktual	Positif	548	51
	Negatif	131	628

Hasil perhitungan nilai evaluasi kinerja dari model klasifikasi SVM dengan *kernel* RBF:

Ukuran Performa Klasifikasi	Nilai (%)
Accuracy	86,60
Precision	80,70
Recall	91,48

ALGORITMA GENETIKA

Hasil pencarian menggunakan Algoritma Genetika menghasilkan parameter optimal:

Cost	gamma
40,537	19,775

SUPPORT VECTOR MACHINE **ALGORITMA GENETIKA**

Nilai *confusion matrix* yang dihasilkan menggunakan SVM *kernel* RBF dengan Algoritma Genetika:

		Prediksi	
		Positif	Negatif
Aktual	Positif	624	24
	Negatif	55	655

Hasil perhitungan nilai evaluasi kinerja dari model klasifikasi SVM dengan *kernel* RBF:

Ukuran Performa Klasifikasi	Nilai (%)
Accuracy	94,20
Precision	96,29
Recall	91,89



BAB V

KESIMPULAN



KESIMPULAN

1.

Prediksi menggunakan algoritma Support Vector Machine tanpa optimasi menghasilkan nilai akurasi sebesar 86,6%.

2.

Prediksi menggunakan algoritma Support Vector Machine dengan Algoritma Genetika (SVM-GA) menghasilkan nilai akurasi sebesar 94,2%.

3.

Akurasi yang dihasilkan oleh algoritma Support Vector Machine dengan Algoritma Genetika (SVM-GA) lebih besar dibandingkan dengan Support Vector Machine tanpa optimasi sehingga dapat dinyatakan bahwa algoritma Support Vector Machine dengan Algoritma Genetika (SVM-GA) lebih baik dalam memprediksi nasabah gadai potensial pada PT Pegadaian.



SARAN

1.

Penelitian selanjutnya diharapkan menggunakan beberapa kombinasi perbandingan data training dan data testing yang lebih bervariasi sehingga dapat memberikan akurasi yang baik.

2.

Penelitian selanjutnya yang menggunakan Support Vector Machine dengan Algoritma Genetika (SVM-GA) diharapkan menggunakan beberapa kombinasi parameter Algoritma Genetika agar memberikan nilai akurasi yang lebih baik.





**THANK
YOU**