

KLASIFIKASI ANGGOTA KOPERASI AMANAH INSANI MAKMUR MENGUNAKAN METODE NAIVE BAYES

David Fathurrahman¹⁾, Dwi Safiroh Utsalina²⁾, Eka Yuniar³⁾

Program Studi S1-Sistem Informasi, STMIK PPKIA Pradnya Paramita¹⁾

email: david_20510003@stimata.ac.id

Program Studi S1-Sistem Informasi, STMIK PPKIA Pradnya Paramita²⁾

email: dwi@stimata.ac.id

Program Studi S1-Sistem Informasi, STMIK PPKIA Pradnya Paramita³⁾

email: eka@stimata.ac.id

Abstract

Cooperatives have long been one of the key pillars supporting the community economy in Indonesia, with the principles of family values and mutual cooperation as their foundation. However, in the current digital era, optimizing cooperative data management has become crucial to ensure effective management. This study aims to optimize credit data management at the Amanah Insani Makmur (AIM) Cooperative in Bantul, Yogyakarta, using the Naive Bayes classification method. By classifying credit data, it is hoped that potential, special, and non-potential customers can be identified more accurately, which can aid in decision-making related to credit facilities and promotions. The Naive Bayes method was chosen because it has been proven effective in previous studies, with high accuracy levels. The data used in this study consists of customer data from AIM Cooperative, totaling 1004 records, which were divided into 70% training data and 30% testing data. The results of the study indicate that the Naive Bayes method can provide a classification accuracy of 86.42%, demonstrating its potential to enhance the efficiency of credit data management at AIM Cooperative.

Keywords: *Cooperatives, Credit Data, and Naive Bayes.*

1. PENDAHULUAN

Koperasi adalah organisasi yang didirikan untuk mencapai tujuan bersama, berlandaskan asas kekeluargaan dan gotong royong. Di Indonesia, koperasi berperan penting dalam perekonomian, terutama setelah masa penjajahan. Koperasi Amanah Insani Makmur (KSPPS AIM) yang berdiri pada tahun 2012 di Bantul, Yogyakarta, bergerak dalam layanan simpan pinjam, tabungan, dan pembiayaan kredit dengan prinsip syariah.

Namun, KSPPS AIM menghadapi masalah kredit macet yang signifikan, dengan 230 dari 1008 transaksi kredit mengalami kemacetan antara tahun 2020 hingga 2023. Hal ini disebabkan oleh kurangnya kontrol manajemen dalam penyaluran kredit, yang masih dilakukan

secara subjektif tanpa mempertimbangkan riwayat kredit nasabah.

Untuk itu, penelitian ini bertujuan untuk mengoptimalkan pengelolaan kredit di KSPPS AIM melalui klasifikasi nasabah menggunakan metode Naive Bayes. Dengan mengklasifikasikan nasabah menjadi Potensial, Khusus, dan Non-potensial, diharapkan mengurangi risiko kredit macet pada nasabah KSPPS AIM.

2. KAJIAN LITERATUR

Data Mining adalah proses pengambilan dan pengidentifikasian informasi berguna dan pengetahuan terikat dari database besar menggunakan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin (Utomo & Mesran, 2020). Dalam data mining terdapat beberapa metode antara lain

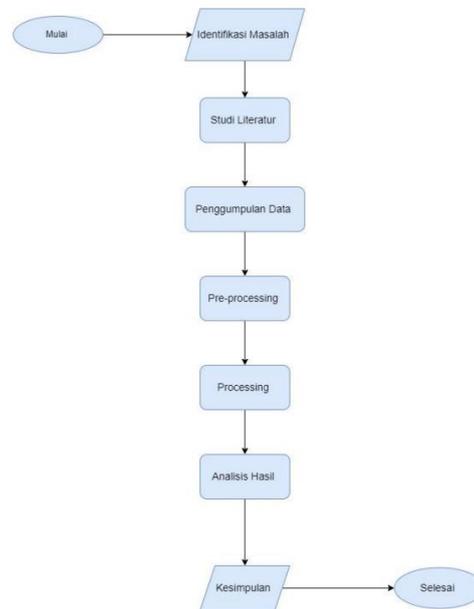
Association, Clustering, dan Klasifikasi. Salah satu metode klasifikasi yang sering digunakan adalah Naive Bayes, yang berbasis pada probabilitas. Metode ini dikenal efektif dan sederhana, sehingga sering digunakan untuk memprediksi kategori data baru berdasarkan data historis. Metode Naïve Bayes merupakan salah satu metode yang digunakan untuk melakukan klasifikasi berdasarkan probabilitas atau kemungkinan dari data sebelumnya, selain pendekatannya sederhana metode tersebut juga dapat melakukan klasifikasi secara baik (Bianto, et al., 2019).

Pada proses data mining, terdapat beberapa tahap penting, di antaranya adalah Pre-Processing yang merupakan langkah untuk membersihkan data dengan tujuan untuk menyeragamkan bentuk kata dan mengurangi volume kata (Singgalen, 2023). Pelabelan yang merupakan proses untuk memberikan tag atau label data yang masih belum diolah sehingga dapat menentukan kelas tertentu yang dimiliki oleh data tersebut dan dapat membantu sistem mengidentifikasi kelas dalam data yang tidak memiliki label (Septianingrum, et al., n.d.). Kemudian, Processing yang mencakup pembagian data menjadi data training dan testing. Data Training digunakan untuk pengolahan data awal dimana proses ini melatih sistem (komputer) yang diberikan pengetahuan dengan sekumpulan data sampai sistem cukup cerdas. Hasil dari data Training ini adalah sebuah model yang digunakan untuk melakukan proses testing atau pengujian (Siregar, et al., 2020; Utomo & Mesran, 2020).

Untuk mendukung proses ini, *Google Colab* digunakan sebagai platform komputasi berbasis web yang memungkinkan penulisan dan eksekusi kode *Python* secara interaktif, dengan akses ke GPU/TPU untuk meningkatkan kinerja, serta mendukung kolaborasi real-time. Kajian literatur ini memberikan dasar teori yang kuat bagi metodologi penelitian yang akan dilaksanakan.

3. METODE PENELITIAN

Pada metode penelitian diperlukan adanya kerangka kerja yang sistematis untuk membantu dalam penyusunan suatu penelitian (Gambar 1).



Gambar 1 Kerangka Kerja

a. Identifikasi Masalah

Berdasarkan hasil observasi yang sudah dilakukan pada KSPPS AIM, diketahui bahwa permasalahan yang terjadi adalah belum optimalnya pengolahan data riwayat kredit nasabah. Hal ini menyebabkan pihak koperasi belum bisa menentukan mana saja yang termasuk nasabah potensial, nasabah non-potensial, dan nasabah khusus.

b. Studi Literatur

Penelitian ini menggunakan sumber dari beberapa jurnal untuk mendukung penelitian dengan cara membaca, mengidentifikasi, mengolah jurnal, data *Mining*, Klasifikasi data, *Pre-Processing*, *Processing* dan Confusion Matrix sesuai dengan kebutuhan yang diinginkan.

c. Pengumpulan Data

Data yang digunakan untuk penelitian ini yaitu data riwayat kredit nasabah yang berjumlah 1008 tersebut mulai Januari 2020 sampai September 2023. Pada

penelitian ini data yang akan digunakan untuk mengkasifikasi nasabah yaitu pada kolom keterangan, ketepatan waktu bayar, dan persentase margin.

d. Pre-Processing

Pada fase Pre-Processing terdapat dua fase yaitu Cleansing dan End Coding.

1) Fase Cleansing

Fase cleansing pada penelitian ini digunakan untuk menghilangkan tanda baca titik (.) dan angka yang tidak diperlukan (Tabel 1) menjadi data yang siap pakai (Tabel 2).

Tabel 1 Data Sebelum Fase Cleansing

No	ID Anggota	Keterangan	Ketepatan Waktu Bayar	Persentase Margin
1	'N0524 94945 91020	Lunas	Terlambat 6 .	27%
2	'N0031 11150 11313	Lunas	Tepat Waktu	23%
3	'N9018 02019 00841	Lunas	Maju 9 .	33%
4	'N1109 09019 20509	Lunas	Terlambat 5 .	25%
5	'N0002 85945 10151	Lunas	Tepat Waktu	23%
6	'N0281 91111 12911	Lunas	Maju 3 .	23%
7	'N0020 40821 01041	Lunas	Tepat Waktu	23%
8	'N1010 11010 11302	Lunas	Terlambat 1 .	23%
9	'N0144 09696 00909	Belum Lunas	Terlambat 1 .	23%
10	'N0017 70947 00110	Lunas	Tepat Waktu	33%
11	'N0011 00525 21000	Lunas	Maju 6 .	23%
12	'N0022 10093 07722	Lunas	Maju 1 .	23%
13	'N0189 44300 03401	Lunas	Tepat Waktu	23%
14	'N0010 24528 11123	Lunas	Terlambat 2 .	23%
15	'N0112 21291 22222	Lunas	Maju 1 Tepat Waktu	23%
16	'N0524 94945 91020	Belum Lunas	Terlambat 7 .	23%

Tabel 2 Data Setelah Fase Cleansing

Keterangan	Ketepatan_Waktu_baya r	Persentase_Margi n
Lunas	Terlambat	27%
Lunas	Tepat Waktu	23%
Lunas	Tepat Waktu	33%
Lunas	Terlambat	25%
Lunas	Tepat Waktu	23%
Lunas	Tepat Waktu	23%
Lunas	Tepat Waktu	23%
Lunas	Terlambat	23%
Belum Lunas	Terlambat	23%
Lunas	Tepat Waktu	33%
Lunas	Tepat Waktu	23%
Lunas	Tepat Waktu	23%
Lunas	Tepat Waktu	23%
Lunas	Terlambat	23%

2) Fase End Coding

Fase End Coding yaitu mengubah text menjadi angka atau numeric. Nilai yang di rubah pada fase End Coding dalam penelitian ini yaitu Pada kolom Keterangan, Ketepatan Waktu Bayar, dan Persentase Margin. Pada kolom keterangan lunas diubah menjadi 1 dan belum lunas menjadi 2. Pada kolom Ketepatan Waktu Bayar, keterangan tepat waktu diubah menjadi 1, terlambat menjadi 2 dan belum terbayar menjadi 3. Pada kolom Persentase Margin, persentase margin di atas 30% akan diubah menjadi 1 dan persentase margin bawah 30% diubah menjadi 2 (Tabel 3).

Tabel 3 Perubahan End Coding Pada Data Riwayat Kredit

Kolom Keterangan		Kolom Ketepatan Waktu Bayar		Kolom Persentase Margin	
Text	End Coding	Text	End Coding		End Coding
Lunas	1	Tepat Waktu	1	> 30%	1
Belum Lunas	2	Terlambat	2	< 30 %	2
		Belum Terbayar	3		

Fase *End Coding* pada data riwayat kredit nasabah tersebut akan dirubah menjadi angka atau numerik (Tabel 4).

Tabel 4 Hasil Fase End Coding

Keterangan	Ketepatan Waktu Bayar	Persentase Margin
1	2	2
1	1	2
1	1	1
1	2	2
1	1	2
1	1	2
1	1	2
1	2	2
2	3	2
1	1	1
1	1	2
1	1	2
1	1	2
1	1	2

e. Pelabelan

Kolom yang digunakan untuk menentukan pelabelan yaitu kolom keterangan, ketepatan waktu bayar, dan persentase margin. Label yang digunakan yaitu Nasabah Potensial, Nasabah Khusus, dan Nasabah Non-Potensial. Terdapat dua Kreteria untuk pemberian label Potensial yaitu jika keterangan “lunas”, ketepatan waktu bayar “tepat waktu”, persentase margin “besar” dan keterangan “belum lunas”, ketepatan waktu bayar “tepat waktu”, persentase margin “besar”.

Pada label Khusus terdapat empat kriteria pada kreteria ke-1 jika keterangan “lunas”, ketepatan waktu bayar “tepat waktu”, persentase margin “kecil”, kriteria ke-2 jika keterangan “lunas”, ketepatan waktu bayar “terlambat”, persentase margin “besar”, kriteria ke-3 jika keterangan “lunas”, ketepatan waktu bayar “terlambat”, persentase margin “kecil”, kriteria ke-4 jika keterangan “belum lunas”, ketepatan waktu bayar “tepat waktu”, persentase margin “kecil”.

Pada label Non-Potensial terdapat empat kriteria pada kriteria ke-1 jika keterangan “belum lunas”, ketepatan waktu bayar “terlambat”, persentase margin “besar”, kriteria ke-2 jika keterangan “belum lunas”, ketepatan waktu bayar “terlambat”, persentase margin “kecil”, kriteria ke-3 jika keterangan “belum lunas”, ketepatan waktu bayar “belum terbayar”, persentase margin “besar”, ke-4 jika keterangan “belum lunas”, ketepatan waktu bayar

“belum terbayar”, persentase margin “kecil” (Tabel 5).

Tabel 5 Hasil Proses Pelabelan

Keterangan	Ketepatan Waktu Bayar	Persentase Margin	Label
Lunas	Tepat Waktu	Besar	potensial
Lunas	Tepat Waktu	Kecil	khusus
Lunas	Terlambat	Besar	khusus
Lunas	Terlambat	Kecil	khusus
Belum Lunas	Tepat Waktu	Besar	potensial
Belum Lunas	Tepat Waktu	Kecil	khusus
Belum Lunas	Terlambat	Besar	non-potensial
Belum Lunas	Terlambat	Kecil	non-potensial
Belum Lunas	Belum Terbayar	Besar	non-potensial
Belum Lunas	Belum Terbayar	Kecil	non-potensial

f. Processing

Pada fase *Processing* yang perlu dilakukan adalah *Splitting Data*, *Data Training Data Testing* dan *Confusion Matrix*.

1) *Splitting Data*

Splitting Data yaitu data akan di bagi menjadi data *Training* (Tabel 6) dan data *Testing* (Tabel 7). Pada penelitian ini data *Training* yang digunakan sebesar 70% dan data *Testing* 30% dari data set.

Tabel 6 Data Training

Keterangan	Ketepatan Waktu Bayar	Persentase Margin	Label
Lunas	Tepat Waktu	Besar	Potensial
Lunas	Tepat Waktu	Kecil	Khusus
Lunas	Terlambat	Besar	Khusus
Belum Lunas	Belum Terbayar	Kecil	Non-potensial
Belum Lunas	Tepat Waktu	Besar	Potensial
Belum Lunas	Terlambat	Besar	Non-potensial
Lunas	Tepat Waktu	Besar	Potensial

Tabel 7 Data Testing

Keterangan	Ketepatan Waktu Bayar	Persentase Margin	Label
Belum Lunas	Belum Terbayar	Kecil	nonpotensial
Lunas	Tepat Waktu	Besar	Potensial

Keterangan	Ketepatan Waktu Bayar	Presentase Margin	Label
Lunas	Terlambat	Kecil	khusus

2) Fase Training

Sebelum masuk ke fase *testing* model akan di latih terlebih dahulu agar bisa mengklasifikasi dengan baik. Dalam penelitian ini pelatihan menggunakan algoritma *Naive Bayes* untuk mengklasifikasi Nasabah Potensial, Khusus, dan Non-Potensial. Data Training akan digunakan untuk menghitung nilai *Term Frequency (TF)* dari kata dalam kategori Potensial, Khusus, dan Non-Potensial. Selanjutnya, akan dihitung nilai $P(X)$ atau probabilitas kemunculan kata tersebut dalam seluruh data. Kemudian akan dihitung nilai $P(H)$ atau probabilitas kemunculan data dalam kategori Potensial, Khusus, dan Non-Potensial (Tabel 8).

Tabel 8 Perhitungan Probabilitas

No	Kata	TF (Potensial)	TF (Khusus)	TF (NonPotensial)	Total	$P(X)$
1	lunas	1	2	0	3	0,15
2	belum lunas	0	1	3	4	0,2
3	tepat waktu	1	1	0	2	0,1
4	terlambat	0	1	2	3	0,15
5	belum terbayar	0	0	1	1	0,05
6	Besar	1	1	2	4	0,2
7	Kecil	0	2	1	3	0,15
Jumlah Term		3	8	9	20	
P(H)		0,15	0,4	0,45		

Setelah mengetahui nilai TF , $P(X)$, dan $P(H)$ kemudian akan di hitung nilai $P(X/H)$ dan $P(H/X)$ (Tabel 9).

Tabel 9 Perhitungan $P(X/H)$ dan $P(H/X)$

$P(X/H)$ Potensial	$P(X/H)$ Khusus	$P(X/H)$ Non-Potensial	$P(H/X)$ Potensial	$P(H/X)$ Khusus	$P(H/X)$ Non-Potensial
0,333	0,25	0	0,25	0,666	0
0	0,125	0,333	0	0	0,75
0,333	0,125	0	0,5	0	0
0	0,125	0,222	0	0,333	0

$P(X/H)$ Potensial	$P(X/H)$ Khusus	$P(X/H)$ Non-Potensial	$P(H/X)$ Potensial	$P(H/X)$ Khusus	$P(H/X)$ Non-Potensial
0	0	0,111	0	0	1
0,333	0,125	0,222	0,25	0	0
0	0,25	0,111	0	0,666	0,333

Setelah mengetahui nilai dari $P(X/H)$ dan $P(H/X)$ pada data *Training* kemudian model akan diuji menggunakan data *Training* menggunakan nilai $P(H/X)$ (Tabel 10).

Tabel 10 Hasil Pelatihan Data

Data Latih	Label
Data 1	Potensial
Data 2	Khusus
Data 3	Khusus
Data 4	Khusus
Data 5	Non-Potensial
Data 6	Non-Potensial
Data 7	Non-Potensial

3) Fase Testing

Data *Testing* pada fase ini yang diuji adalah kata yang memiliki nilai $P(H/X)$ (Tabel 11).

Tabel 11 Hasil Klasifikasi Menggunakan Metode Naive Bayes

Data	Label
Data ke-1	Potensial
Data ke-2	Khusus
Data ke-3	Non-Potensial

4) Confusion Matrix

Setelah data berhasil di klasifikasi menggunakan metode *Naive Bayes* selanjutnya hasil dari prediksi akan di lakukan pengujian untuk mengetahui seberapa tepat hasil prediksi untuk kelas Potensial, Khusus dan Non-Potensial dengan menggunakan *Confusion Metrik* (Tabel 12).

Tabel 12 *Confusion Matrix*

Kelas	TP	FN	FP	TN
Potensial	1	0	0	2
Khusus	1	0	0	2
Non-Potensial	1	0	0	2

Setelah hasil prediksi sudah diuji menggunakan *Confusion Matrix* kemudian akan dihitung nilai *Recall*, *Precision*, dan *Accuracy*.

1) *Recall*

Recall merupakan tingkat keberhasilan sistem dalam menemukan Kembali sebuah informasi (Damuri, et al., 2021).

Recall pada Label Potensial

$$Recall = \frac{1}{1 + 1} : 0$$

Recall pada Label Khusus

$$Recall = \frac{1}{1 + 0} : 1$$

Recall pada Label Non-Potensial

$$Recall = \frac{1}{1 + 0} : 1$$

2) *Precision*

Precision yaitu untuk seberapa tepat atau akurat model yang di prediksi sebagai positif (Septianingrum, et al., n.d.).

Precision pada Label Potensial

$$Recall = \frac{1}{1 + 0} : 0$$

Precision pada Label Khusus

$$Recall = \frac{1}{1 + 0} : 1$$

Precision pada Label Non-Potensial

$$Recall = \frac{1}{1 + 0} : 1$$

3) *Accuracy*

Accuracy tingkat kedekatan antara nilai prediksi dengan nilai actual (Septianingrum, et al., n.d.).

Accuracy pada Kelas Potensial

$$Accuracy = \frac{1 + 2}{1 + 0 + 0 + 2} : \frac{3}{3} : 1$$

Accuracy pada Kelas Khusus

$$Accuracy = \frac{1 + 2}{1 + 0 + 0 + 2} : \frac{3}{3} : 1$$

Accuracy pada Kelas Non-Potensial

$$Accuracy = \frac{1 + 2}{1 + 0 + 0 + 2} : \frac{3}{3} : 1$$

Nilai *Accuracy* dari Kelas

$$Accuracy = \frac{2 + 6}{2 + 0 + 0 + 6} : \frac{8}{8} : 1$$

Kesimpulan dari perhitungan Naive Bayes dari hasil *Recall* dengan nilai 1, *Precision* dengan nilai 1 dan akurasi dengan nilai 1. Hal ini berarti kemampuan model dalam mengklasifikasikan data riwayat kredit dengan benar adalah 100%.

4. HASIL DAN PEMBAHASAN

Pada bagian ini menjelaskan tentang pengujian dan hasil pengklasifikasian menggunakan algoritma Naive Bayes untuk mengklasifikasi data kredit koperasi Amanah Insani Makmur.

a. Pre-Processing

1) Fase *Cleansing* bertujuan untuk menghilangkan tanda baca atau kalimat yang tidak di perlukan. Fase *Cleansing* dilakukan dengan menggunakan *Microsoft Excel* (Tabel 13).

Tabel 13 Hasil dari Fase *Cleansing*

Keterangan	Ketepatan Waktu Bayar	Persentase Margin
Lunas	Terlambat	27%
Lunas	Tepat Waktu	23%
Lunas	Tepat Waktu	33%

Keterangan	Ketepatan Waktu Bayar	Persentase Margin
Lunas	Terlambat	25%
Lunas	Tepat Waktu	23%
Lunas	Tepat Waktu	23%
Lunas	Tepat Waktu	23%
Lunas	Terlambat	23%
Belum Lunas	Belum Terbayar	23%
Lunas	Tepat Waktu	33%
Lunas	Tepat Waktu	23%
Lunas	Tepat Waktu	23%
Lunas	Tepat Waktu	23%
Lunas	Terlambat	23%
Lunas	Tepat Waktu	23%
Belum Lunas	Belum Terbayar	23%
Lunas	Terlambat	28%

2) End Coding digunakan untuk mengubah nilai dari setiap kolom menjadi angka (Gambar 2).

Keterangan	Ketepatan_Waktu_Bayar	Persentase_Margin
0	1	2
1	1	1
2	1	1
3	1	2
4	1	1
...
1002	2	1
1003	2	1
1004	2	1
1005	2	1
1006	2	1

1004 rows x 4 columns

Gambar 2 Hasil dari Tahapan Pre-Processing

b. Pelabelan

Setelah melalui tahap *Pre-Processing* data akan di berikan label sesuai dengan yang sudah ditentukan (Gambar 3).

Keterangan	Ketepatan_Waktu_Bayar	Persentase_Margin	Label
0	1	2	2 khusus
1	1	1	2 khusus
2	1	1	1 potensial
3	1	2	2 khusus
4	1	1	2 khusus
5	1	1	2 khusus
6	1	1	2 khusus
7	1	2	2 khusus
8	2	3	2 non-potensial
9	1	1	1 potensial

Gambar 3 Hasil dari Fase Pelabelan

Setelah data di berikan label sesuai dengan kriteria yang di tentukan hasil dari pelabelan tersebut akan di lakukan End Coding untuk merubah hasil label potensial menjadi 1, khusus menjadi 2, non-potensial menjadi 3 (Gambar 4).

Keterangan	Ketepatan_Waktu_Bayar	Persentase_Margin	Label
0	1	2	2 2
1	1	1	2 2
2	1	1	1 1
3	1	2	2 2
4	1	1	2 2
...
1002	2	1	2 2
1003	2	1	2 2
1004	2	1	2 2
1005	2	1	1 1
1006	2	1	1 1

1004 rows x 4 columns

Gambar 4 Tampilan pada End Coding Pelabelan

c. Processing

1) Pada fase *Splitting Data*, data akan dibagi menjadi dua bagian yaitu Data Training dan Data Testing. Data yang di jadikan data training sebesar 70% dan data testing sebesar 30%. Hasil pembagian data Training sebanyak 702 data (Gambar 5) dan data Testing sebanyak 302 data (Gambar 6).

	Keterangan	Ketepatan_Waktu_Bayar	Persentase_Margin	Label
543	1	1	1	1
442	1	1	1	1
820	2	3	2	3
398	1	1	2	2
961	2	3	1	3
...
106	1	1	2	2
270	2	3	2	3
861	2	3	2	3
435	1	1	2	2
102	2	3	2	3

702 rows x 4 columns

Gambar 5 Tampilan Hasil Data Training

	Keterangan	Ketepatan_Waktu_Bayar	Persentase_Margin	Label
929	2	1	1	1
630	1	2	2	2
682	1	2	2	2
514	1	1	2	2
365	1	2	2	2
...
208	1	2	2	2
778	1	1	1	1
334	2	3	1	3
210	1	1	2	2
350	1	1	1	1

201 rows x 4 columns

Gambar 6 Tampilan Hasil Data Testing

2) Setelah data dipisahkan menjadi data Training dan data Testing, mesin akan dilatih menggunakan metode Naive Bayes dengan data pelatihan. Selanjutnya, tingkat akurasi model akan dievaluasi menggunakan data pengujian untuk memastikan keefektifan klasifikasi (Gambar 7).

```

from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import classification_report

# Inisialisasi model NBC
clf = MultinomialNB()

# Melatih model NBC menggunakan data pelatihan
clf.fit(X_train.drop('Label', axis=1), X_train['Label'])

# Melakukan prediksi pada data pengujian
predicted = clf.predict(X_test.drop('Label', axis=1))

# Menampilkan hasil klasifikasi
print(classification_report(X_test['Label'], predicted))

```

Gambar 7 Source Code pada fase Training

3) Setelah model dilatih menggunakan data Training kemudian model di uji dengan Data Testing. Hasil dari klasifikasi model menggunakan metode Naive Bayes divisualisasikan menggunakan Confusion Matrix (Gambar 8).



Gambar 8 Hasil Confusion Matrik

Pada baris pertama dengan nilai [0, 36, 0] pada kolom ke-1 dengan nilai 0 data pada kelas potensial tidak ada yang berhasil di prediksi oleh model ke dalam kelas Potensial. Pada kolom ke-2 dengan nilai 36 pada kelas potensial, data tersebut terprediksi oleh model ke dalam kelas Khusus. Pada kolom ke-3 dengan nilai 0, data pada kelas potensial yang di prediksi oleh model ke dalam kelas Non-Potensial.

Pada baris kedua dengan nilai [0, 165, 5] pada kolom ke-1 dengan nilai 0, data kelas Khusus yang terprediksi pada kelas Potensial yaitu 0. Pada kolom ke-2 data kelas Khusus yang terprediksi ke dalam kelas Khusus sebanyak 165 data. Pada kolom ke-3 kelas Khusus yang terprediksi pada kelas Non-Potensial sebanyak 5 data.

Pada baris ketiga dengan nilai [0, 0, 95] pada kolom ke-1 dengan nilai 0 data kelas Non-Potensial yang terprediksi ke dalam kelas Potensial sebanyak 0 data. Pada kolom ke-2 data kelas Non-Potensial yang terprediksi ke dalam kelas Khusus sebanyak 0 data. Pada kolom ke-3 data dengan kelas Non-Potensial berhasil di prediksi sebanyak 95 data.

d. Setelah data di evaluasi dengan Confusion Matrik selanjutnya akan di hitung nilai Recall (Tabel 14).

Tabel 14 Hasil dari Perhitungan Recall

No	Kelas	Recall
1	Potensial	0
2	Khusus	0.97
3	Non-Potensial	1

e. Setelah data dihitung nilai Recall kemudian data akan dihitung tingkat Presicion (Tabel 15).

Tabel 15 Hasil Perhitungan Nilai Presicion

No	Kelas	Precision
1	Potensial	0
2	Khusus	0.82
3	Non-Potensial	0.95

f. Setelah model di evaluasi menggunakan Confusion, Metrix, Recall, Precision kemudian model akan di uji tingkat akurasi. Hasil dari Accuracy model Klasifikasi menggunakan metode Naive Bayes dengan pembagian Data Training sebesar 70% dan Data Testing sebesar 30 % sebesar 0,8642.

5. KESIMPULAN

Dari proses mengklasifikasi data dengan menerapkan metode Naive Bayes menggunakan data Training sebesar 70% dan data Testing 30% menghasilkan nilai akurasi sebesar 86 % meskipun di dapatkan nilai akurasi relative tinggi dari pengujian di ketahui bahwa pada kelas Potensial pada data Riwayat kredit tidak terklasifikasi dengan benar. pada nilai precicion tidak memenuhi hasil perhitungan yang di harapkan oleh pihak menejemen dimana pihak menejemen menginginkan nilai akurasi sebesar 95 %. Hal ini menunjukkan bahwa nasabah yang termasuk dalam kelas "Potensial" tidak memenuhi syarat untuk mengajukan kredit.

6. REFERENSI

Bianto, M. A., Kusriani & Sudarmawan, 2019. Perancangan Sistem Klasifikasi Penyakit Jantung Menggunakan Naive Bayes.

Damuri, A., Riyanto, U., Rusdianto, H. & Aminudin, M., 2021. Implementasi Data Mining dengan Algoritma Naive Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako

Septianingrum, F., Jaman, H. & Enri, U., n.d. Analisis Sentimen Pada Isu Vaksin Covid-19 di Indonesia dengan Metode Naive Bayes Classifier.

Siregar, N. C., Ruli, R. & Sudirman, M. Y. D., 2020. Implementasi Metode Naive Bayes Classifier (NBC) Pada Komentar Warga Sekolah Mengenai Pelaksanaan Pembelajaran Jarak Jauh (PJJ).

Utomo, D. P. & Mesran, 2020. Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung.