

## **PENERAPAN ALGORITMA NAÏVE BAYES UNTUK PENENTUAN RESIKO KREDIT BANK**

Harun Deppalallo<sup>1\*</sup>, Jullia Titaley<sup>2</sup>, Djoni Hatidja<sup>1</sup>

<sup>1</sup>*Program Studi Matematika, Jurusan Matematika, Universitas Sam Ratulangi\**

<sup>2</sup>*Program Studi Sistem Informasi, Jurusan Matematika, Universitas Sam Ratulangi\*\**

*\*Email: harundepalallo@gmail.com*

### **ABSTRACT**

*This study aims to classify in determining the value of bank credit risk. The data is obtained from a data bank with a population of 900 data where this data is divided into 800 training data and 100 testing data. The attributes used are 5 namely, annual income, Active KPR, Loan Duration, Number of Dependents and Average Overdue. The method used in this research is the naive Bayes algorithm which is processed using R Studio software. For manual calculation of training data used 100 data from 800 training data, and testing data taken 5 data from 100 testing data. Meanwhile, the data processed by R Studio has an accuracy of 0.997 or 99%. Sensitivity class 1 100%, Sensitivity class 2 98%, Sensitivity class 3 100%, Sensitivity class 4 100% and Sensitivity class 5 100%. Specificity class 1 100%, Specificity class 2 100%, Specificity class 3 100%, Specificity class 4 100% and Specificity class 5 99%.*

**Keywords:** *Naïve Bayes, Credit Risk*

### **ABSTRAK**

Penelitian ini bertujuan untuk melakukan klasifikasi dalam menentukan nilai resiko kredit bank. Data didapat dari data bank dengan populasi 900 data dimana data ini dibagi menjadi 800 data training dan 100 data testing. Atribut yang digunakan sebanyak 5 yaitu, pendapatan setahun, KPR Aktif, Durasi Pinjaman, Jumlah Tanggungan dan Rata-rata Overdue. Metode yang digunakan dalam penelitian ini adalah algoritma naïve bayes yang di proses menggunakan software R Studio. Untuk perhitungan manual data training yang digunakan 100 data dari 800 data training, dan data testing diambil 5 data dari 100 data testing. Sedangkan data yang di proses dengan R Studio memiliki accuracy 0.997 atau 99%. Sensitivity class 1 100%, Sensitivity class 2 98%, Sensitivity class 3 100%, Sensitivity class 4 100% dan Sensitivity class 5 100%. Specificity class 1 100%, Specificity class 2 100%, Specificity class 3 100%, Specificity class 4 100% dan Specificity class 5 99%.

**Kata Kunci:** *Naïve Bayes, Resiko Kredit*

### **PENDAHULUAN**

Bank adalah salah satu perusahaan yang dapat memberikan jasa kredit kepada nasabah [1]. Bank adalah badan usaha yang menghimpun dana dari masyarakat dalam bentuk simpanan dan menyalurkannya kepada masyarakat dalam bentuk kredit dan atau bentuk-bentuk lainnya dalam rangka meningkatkan taraf hidup rakyat banyak [2]. Fungsi utama bank diatur dalam Pasal 3 Undang-Undang Nomor 10 Tahun 1998 yaitu fungsi utama perbankan Indonesia adalah sebagai penghimpun dan penyalur dana masyarakat. Bahwa bank dapat berfungsi sebagai penerima kredit, menyalurkan kredit, melakukan pembiayaan, investasi, menerima deposito, menciptakan uang dan jasa-jasa lainnya seperti tempat penyimpanan barang berharga.

Kata Kredit berasal dari kata “credere” yang berarti : percaya, atau *to believe/to trust*. Maksud dari kata tersebut bahwa kredit mengandung unsur kepercayaan dari bank kepada nasabah untuk dapat menggunakan kredit sebaik mungkin [3]. Fungsi kredit secara luas sebagaimana yang dikemukakan oleh Kasmir dalam [4] yaitu : (1) Untuk meningkatkan daya guna uang (2) Untuk meningkatkan peredaran dan lalu lintas uang (3) Untuk meningkatkan daya guna barang (4) Untuk meningkatkan peredaran barang (5) Sebagai stabilitas ekonomi (6) Untuk meningkatkan kegairahan berusaha (7) Untuk meningkatkan pemerataan pendapatan nasional (8) Untuk meningkatkan hubungan internasional. Pengguna jasa kredit di Indonesia sangat banyak, hal ini dikarenakan semakin meningkatnya kebutuhan belanja masyarakat dan meningkatnya kebutuhan

modal untuk memulai atau mengembangkan usahanya. Adanya kebutuhan tersebut membuat masyarakat melakukan pinjaman atau kredit di Bank untuk memenuhinya.

Pengajuan kredit bisa dilakukan oleh calon nasabah dengan memenuhi semua persyaratan yang sudah ditentukan oleh Bank. Perlu diketahui bahwa pemberian kredit kepada nasabah mempunyai resiko. Resiko tersebut dapat berupa pinjaman yang mengalami penangguhan pembayaran angsuran pokok dan tunggakan bunga dikarenakan ketidakmampuan penerima kredit untuk membayarnya, sehingga pengembalian kredit tidak dilakukan tepat waktu sesuai perjanjian kredit. Semakin banyak nasabah, semakin tinggi tingkat resiko yang dapat diterima oleh Bank [5]. Kredit bermasalah merupakan kegagalan pihak debitur memenuhi kewajibannya untuk membayar angsuran (cicilan) pokok kredit beserta bunga yang telah disepakati kedua belah pihak dalam perjanjian kredit. Dengan adanya masalah tersebut, salah satu cara yang bisa dilakukan untuk mencegah terjadinya resiko kredit adalah dengan menghitung nilai resiko kredit (Risk Rating) guna mengetahui nilai resiko kredit Bank

Pada penelitian sebelumnya dengan judul penelitian “Penerapan Algoritma Naive Bayes untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor” menggunakan metode Naive Bayes untuk penentuan resiko kredit kepemilikan kendaraan bermotor dengan data uji sebanyak 10 customer dengan menggunakan dua kelas. Klasifikasi dengan kelas Good Cust sebanyak 9 customer dan kelas Bad Cust sebanyak 1 customer. Pengujian data menggunakan metode Naive Bayes berhasil menampilkan dua kelas dari hasil klasifikasi dengan persentase keakuratan sebesar sebesar 90 % [6].

Peringkat resiko (*risk rating*) kredit sangat penting bagi Bank, karena peringkat resiko dapat dijadikan bank sebagai acuan untuk memprediksi peringkat resiko kredit dari data nasabah yang mengajukan kredit. Berdasarkan latar belakang tersebut, maka penulis melakukan penelitian tentang Penerapan Algoritma Naive Bayes untuk Penentuan Resiko Kredit.

## **METODE PENELITIAN**

### **Waktu dan Tempat Penelitian**

Penelitian ini dilaksanakan dari bulan Oktober 2021, mulai dari penyusunan proposal, pengambilan data serta pengolahan data. Berhubungan keadaan masih dalam masa pandemic Covid-19 pengolahan data dilakukan di rumah atau *study from home* dan di Jurusan Matematika.

### **Metode Analisis Data**

Data yang digunakan dalam penelitian ini adalah nasabah yang mengajukan kredit UMKM tahun 2019 dengan data sebanyak 900 data. Metode yang digunakan dalam penelitian ini adalah Algoritma *Naive Bayes* untuk menentukan Resiko Peminjaman Kredit Bank. Analisis data menggunakan Software R Studio. Pada Software R Studio, ada beberapa *package* yang akan diperlukan untuk melakukan proses perhitungan algoritma Naive Bayes di antaranya adalah *package* “*e1071*” dan *package* “*Caret*”. *Package* “*e1071*” adalah *package* yang dapat digunakan untuk menjalankan metode Naive Bayes pada R Studio. *Package* “*Caret*” merupakan *package* yang digunakan untuk melihat apakah metode Naive Bayes ini baik atau tidak untuk digunakan dilihat dari akurasi, sensitifitas, dan spesifisitas [7].

### **Naive Bayes**

Naive Bayes merupakan metode pengklasifikasian probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi kemungkinan yang terjadi di masa depan berdasarkan kejadian di masa sebelumnya [8]. *Naive Bayes* bisa di dipakai untuk memprediksi probabilitas anggota suatu *class*, serta mempunyai akurasi dan kecepatan yang sangat kuat ketika diaplikasikan pada database dengan big data. Naive Bayes dapat diselesaikan dengan persamaan sebagai berikut:

$$P(C|X) = \frac{P(X|C)*P(C)}{P(X)} \quad (1)$$

Keterangan :

$X$  : Data dengan *class* yang belum diketahui

$C$  : Hipotesis data merupakan suatu data *class* spesifik

$P(C|X)$ : Probabilitas hipotesis  $C$  berdasarkan kondisi  $X$  (posteriori probabilitas)

$P(C)$  : Probabilitas hipotesis  $C$  (prior probabilitas)

$P(X|C)$ : Probabilitas  $X$  berdasarkan kondisi pada hipotesis  $C$

$P(X)$  : Probabilitas  $X$

Untuk menjelaskan teorema *Naïve bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas yang cocok bagi sampel yang dianalisis tersebut (Natalius, 2010). Karena itu Teorema Bayes di atas disesuaikan menjadi persamaan :

$$P(C|X_1, \dots, X_n) = \frac{P(X_1, \dots, X_n|C)}{P(X_1, \dots, X_n)} \cdot P(C) \quad (2)$$

Dimana variable  $C$  merepresentasikan kelas, sementara variabel  $X_1, \dots, X_n$  merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas  $C$  (*Posterior*) adalah peluang munculnya kelas  $C$  (sebelum masuknya sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas  $C$  (disebut *likelihood*), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga *evidence*). Karena itu, rumus diatas dapat pula ditulis secara sederhana pada persamaan :

$$posterior = \frac{prior \times likelihood}{evidence} \quad (3)$$

Nilai *Evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai dari posterior tersebut nantinya akan dibandingkan dengan nilai-nilai posterior kelas lainnya untuk menentukan ke kelas apa suatu sampel akan di klasifikasikan. Penjabaran lebih lanjut rumus *Naïve Bayes* tersebut dilakukan dengan menjabarkan secara terperinci ( $C|X_1, \dots, X_n$ ) menggunakan aturan perkalian sebagai berikut.

$$\begin{aligned} P(C|X_1, \dots, X_n) &= P(C)P(X_1, \dots, X_n|C) & (4) \\ &= P(C)P(X_1|C)P(X_2, \dots, X_n|C, X_1) \\ &= P(C)P(X_1|C)P(X_2|C, X_1)P(X_3, \dots, X_n|C, X_1, X_2) \\ &= P(C)P(X_1|C)P(X_2|C, X_1)P(X_3, \dots, X_n|C, X_1, X_2, X_3) \\ &= P(C)P(X_1|C)P(X_2|C, X_1)P(X_3|C, X_1, X_2) \dots P(X_n|C, X_1, X_2, \dots, X_{n-1}) \end{aligned}$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleks faktor-faktor syarat yang berpengaruh terhadap nilai probabilitas, maka semakin tidak mungkin untuk menghitung nilai tersebut satu persatu. Akibatnya proses perhitungan akan semakin susah untuk dilakukan. Disinilah digunakan asumsi independensi yang sangat tinggi (naif), bahwa masing-masing kriteria ( $X_1, X_2, \dots, X_n$ ) saling bebas (independen) satu sama lain. Dengan asumsi tersebut, maka berlaku persamaan :

$$P(X_i|X_j) = \frac{P(X_i)P(X_j)}{P(X_j)} = \frac{P(X_i \cap X_j)}{P(X_j)} = P(X_i)$$

Untuk  $i \neq j$ , sehingga

$$P(X_i|C, X_j) = P(X_i|C) \quad (5)$$

Dari persamaan (4) dan (5) dapat disimpulkan bahwa asumsi independensi naif membuat syarat perhitungan menjadi lebih sederhana. Selanjutnya penjabaran  $P(C|X_1, \dots, X_n)$  dapat disederhanakan menjadi persamaan :

$$P(X_2|C)P(X_3|C) \dots P(C|X_1, \dots, X_n) = P(X_1|C) = \prod_{i=1}^n P(X_i|C) \quad (6)$$

Keterangan :

$\prod_{i=1}^n P(X_i|C)$  : Perkalian ranting antar atribut

Untuk menyelesaikan metode *Naïve Bayes* dapat dilakukan dengan langkah-langkah sebagai berikut:

1. Menentukan data training dan data testing

Menentukan data training dan data testing yang akan di klasifikasikan. Rumus yang digunakan untuk menghitung nilai rata – rata hitung (mean) dapat dilihat sebagai berikut :

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \text{ atau } \mu = \frac{x_1+x_2+x_3+\dots+x_n}{n} \quad (7)$$

Keterangan :

$\mu$  : Rata-rata hitung (*Mean*)

$x_i$  : Nilai Sampel ke-i

$n$  : Jumlah Sampel

2. Menghitung Probabilitas Prior

Menghitung  $P(C_i)$  yang merupakan probabilitas prior untuk setiap sub kelas C yang akan dihasilkan menggunakan persamaan :

$$P(C_i) = \frac{s_i}{s} \quad (8)$$

Dimana  $S_i$  adalah jumlah data training dari kelas  $C_i$ , dan  $s$  adalah jumlah total data training.

3. Menghitung Probabilitas akhir setiap kelas

Menghitung  $P(X_i|C_i)$  yang merupakan probabilitas posterior atau probabilitas akhir  $X_i$  dengan syarat C menggunakan persamaan (6).

$$P(X|Kelas) = P(X_1|Kelas) \times P(X_2|Kelas) \times P(X_3|Kelas) \times P(X_4|Kelas) \times P(X_5|Kelas) \quad (9)$$

4. Probabilitas Akhir

Probabilitas akhir didapat dengan mengalikan  $P(X_i|C_i)$  dan  $P(C_i)$  untuk semua kemungkinan

$$P(C|X) = P(X_i|C_i) \times P(C_i) \quad (10)$$

Hasil yang ditetapkan kedalam kelas  $C_i$  adalah nilai maksimum.

## HASIL DAN PEMBAHASAN

### A. Perhitungan Manual

Untuk perhitungan manual data training yang digunakan 100 data dari 800 data training, dan data testing diambil 5 data dari 100 data testing.

Tabel 1. Data Training

No	Kode Kontrak	Pendapatan Setahun (Juta Rupiah)	KPR Aktif	Durasi Pinjaman (Bulan)	Jumlah Tanggungan	Rata-rata Overdue	Risk Rating
1	AGR-000173	167	YA	36	3	0 – 30 days	1
2	AGR-000211	166	TIDAK	24	5	31 - 45 days	2
3	AGR-000240	156	TIDAK	24	4	0 - 30 days	1
4	AGR-000275	140	TIDAK	12	0	31 – 45 days	2

No	Kode Kontrak	Pendapatan Setahun (Juta Rupiah)	KPR Aktif	Durasi Pinjaman (Bulan)	Jumlah Tanggungan	Rata-rata Overdue	Risk Rating
100	AGR-009533	193	YA	12	5	61 - 90 days	4

Perhitungan manual dilakukan dengan menggunakan 100 data dari 800 data training. Selanjutnya menghitung jumlah dan probabilitas berdasarkan data pada tabel 2. Dari 100 data training yang digunakan, diketahui Risk Rating 1 sebanyak 23 data, Risk Rating 2 sebanyak 25 data, Risk Rating 3 sebanyak 22 data, Risk Rating 4 sebanyak 14 data dan Risk Rating 5 sebanyak 14 data. Perhitungan probabilitas prior menggunakan persamaan 8 :

$$P(C_i) = \frac{s_i}{s}$$

$$P(\text{Risk Rating 1}) = \frac{23}{100} = 0.23$$

$$P(\text{Risk Rating 2}) = \frac{25}{100} = 0.25$$

$$P(\text{Risk Rating 3}) = \frac{22}{100} = 0.22$$

$$P(\text{Risk Rating 4}) = \frac{16}{100} = 0.16$$

$$P(\text{Risk Rating 5}) = \frac{14}{100} = 0.14$$

Tabel 2. Data Testing

No	Kode Kontrak	Pendapatan Setahun (Juta Rupiah)	KPR Aktif	Durasi Pinjaman (Bulan)	Jumlah Tanggungan	Rata-rata Overdue	Risk Rating
1	AGR-009970	120	TIDAK	36	0	0 - 30 days	?
2	AGR-009976	193	TIDAK	24	2	0 - 30 days	?
3	AGR-010246	175	YA	48	5	> 90 days	?
4	AGR-010415	103	YA	24	6	61 - 90 days	?
5	AGR-010498	182	TIDAK	24	2	0 - 30 days	?

Setelah probabilitas prior diketahui, selanjutnya menghitung probabilitas dari masing-masing kriteria yang digunakan. Kriteria yang digunakan adalah Pendapatan Setahun (Juta), KPR Aktif, Durasi Pinjaman, Jumlah Tanggungan dan Rata-rata Overdue. Terdapat 5 Risk Rating yang digunakan dalam pengelompokan yaitu Risk Rating 1 (R1), Risk Rating 2 (R2), Risk Rating 3 (R3), Risk Rating 4 (R4) dan Risk Rating 5 (R5), sehingga perhitungan probabilitas masing-masing kriteria dapat dilihat pada tabel berikut :

1. Menghitung Probabilitas masing-masing kriteria Risk Rating 1 (R1)

Tabel 3. Probabilitas Risk Rating 1 Pendapatan Setahun (Juta Rupiah)

Pendapatan (Juta Rupiah)	R1	Probabilitas
91	1	0.043478261
103	2	0.086956522
115	1	0.043478261
120	1	0.043478261
137	1	0.043478261

<b>Pendapatan (Juta Rupiah)</b>	<b>R1</b>	<b>Probabilitas</b>
156	1	0.043478261
157	1	0.043478261
167	1	0.043478261
175	3	0.130434783
182	1	0.043478261
190	1	0.043478261
192	1	0.043478261
193	5	0.217391304
204	1	0.043478261
213	1	0.043478261
218	1	0.043478261
Jumlah	23	1

Tabel 4. Probabilitas Risk Rating 1 KPR Aktif

<b>KPR Aktif</b>	<b>R1</b>	<b>Probabilitas</b>
ya	2	0.086956522
tidak	21	0.913043478
Jumlah	23	1

Tabel 5. Probabilitas Risk Rating 1 Durasi pinjaman

<b>Durasi Pinjaman</b>	<b>R1</b>	<b>Probabilitas</b>
12	7	0.304347826
24	7	0.304347826
36	4	0.173913043
48	5	0.217391304
Jumlah	23	1

Tabel 6. Probabilitas Risk Rating 1 Jumlah Tanggungan

<b>Jumlah Tanggungan</b>	<b>R1</b>	<b>Probabilitas</b>
0	9	0.391304348
1	6	0.260869565
2	1	0.043478261
3	1	0.043478261
4	2	0.086956522
5	3	0.130434783
6	1	0.043478261
Jumlah	23	1

Tabel 7. Probabilitas Risk Rating 1 Rata-rata *Overdue*

<b>Rata-rata <i>overdue</i></b>	<b>R1</b>	<b>Probabilitas</b>
0 - 30 days	23	1
Jumlah	23	1

2. Menghitung Probabilitas masing-masing kriteria Risk Rating 2 (R2)

Tabel 8. Probabilitas Risk Rating 2 Pendapatan Setahun (Juta Rupiah)

Penghasilan (Juta Rupiah)	R2	Probabilitas
93	1	0.04
94	1	0.04
103	4	0.16
105	1	0.04
109	1	0.04
110	1	0.04
111	1	0.04
120	1	0.04
129	1	0.04
137	4	0.16
140	1	0.04
156	1	0.04
157	1	0.04
166	1	0.04
186	1	0.04
187	2	0.08
193	2	0.08
Jumlah	25	1

Tabel 9. Probabilitas Risk Rating 2 KPR Aktif

KPR Aktif	R2	Probabilitas
ya	1	0.04
tidak	24	0.96
Jumlah	25	1

Tabel 10. Probabilitas Risk Rating 2 Durasi Pinjaman

Durasi Pinjaman	R2	Probabilitas
12	8	0.32
24	4	0.16
36	7	0.28
48	6	0.24
Jumlah	25	1

Tabel 11. Probabilitas Risk Rating 2 Jumlah Tanggungan

Jumlah Tanggungan	R2	Probabilitas
0	6	0.24
1	4	0.16
2	4	0.16
3	1	0.04
4	3	0.12
5	5	0.2
6	2	0.08
Jumlah	25	1

Tabel 12. Probabilitas Risk Rating 2 Rata-rata *Overdue*

Rata-rata <i>overdue</i>	R2	Probabilitas
31 - 45 days	25	1
Jumlah	25	1

3. Menghitung Probabilitas masing-masing kriteria Risk Rating 3 (R3)

Tabel 13. Probabilitas Risk Rating 3 Pendapatan Setahun (Juta Rupiah)

<b>Penghasilan (Juta Rupiah)</b>	<b>R3</b>	<b>Probabilitas</b>
72	1	0.045454545
81	2	0.090909091
84	1	0.045454545
94	1	0.045454545
98	1	0.045454545
103	2	0.090909091
107	1	0.045454545
108	1	0.045454545
115	1	0.045454545
120	1	0.045454545
134	1	0.045454545
137	2	0.090909091
157	1	0.045454545
175	3	0.136363636
182	1	0.045454545
193	1	0.045454545
222	1	0.045454545
Jumlah	22	1

Tabel 14. Probabilitas Risk Rating 3 KPR Aktif

<b>KPR Aktif</b>	<b>R3</b>	<b>Probabilitas</b>
ya	20	0.909090909
tidak	2	0.090909091
Jumlah	22	1

Tabel 15. Probabilitas Risk Rating 3 Durasi Pinjaman

<b>Durasi Pinjaman</b>	<b>R3</b>	<b>Probabilitas</b>
12	3	0.136363636
24	3	0.136363636
36	4	0.181818182
48	12	0.545454545
Jumlah	22	1

Tabel 16. Probabilitas Risk Rating 3 Jumlah Tanggungan

<b>Jumlah Tanggungan</b>	<b>R3</b>	<b>Probabilitas</b>
0	4	0.181818182
1	1	0.045454545
2	5	0.227272727
3	9	0.409090909
4	3	0.136363636
Jumlah	22	1

Tabel 17. Probabilitas Risk Rating 3 Rata-rata *Overdue*

<b>Rata-rata <i>overdue</i></b>	<b>R3</b>	<b>Probabilitas</b>
46 - 60 days	22	1
Jumlah	22	1



4. Menghitung Probabilitas masing-masing kriteria Risk Rating 4 (R4)

Tabel 18. Probabilitas Risk Rating 4 Pendapatan Setahun (Juta Rupiah)

<b>Penghasilan (Juta Rupiah)</b>	<b>R4</b>	<b>Probabilitas</b>
71	1	0.0625
103	1	0.0625
110	1	0.0625
130	1	0.0625
136	1	0.0625
145	1	0.0625
146	1	0.0625
157	1	0.0625
159	1	0.0625
171	1	0.0625
175	1	0.0625
182	1	0.0625
193	1	0.0625
253	1	0.0625
255	1	0.0625
256	1	0.0625
Jumlah	16	1

Tabel 19. Probabilitas Risk Rating 4 KPR Aktif

<b>KPR Aktif</b>	<b>R4</b>	<b>Probabilitas</b>
ya	16	1
tidak	0	0
Jumlah	16	1

Tabel 20. Probabilitas Risk rating 4 Durasi Pinjaman

<b>Durasi Pinjaman</b>	<b>R4</b>	<b>Probabilitas</b>
12	8	0.5
24	3	0.1875
36	1	0.0625
48	4	0.25
Jumlah	16	1

Tabel 21. Probabilitas Risk Rating 4 Jumlah Tanggungan

<b>Jumlah Tanggungan</b>	<b>R4</b>	<b>Probabilitas</b>
0	1	0.0625
5	10	0.625
6	5	0.3125
Jumlah	16	1

Tabel 22. Probabilitas Risk Rating 4 Rata-rata *Overdue*

<b>Rata-rata <i>overdue</i></b>	<b>R4</b>	<b>Probabilitas</b>
61 - 90 days	16	1
Jumlah	16	1

5. Menghitung Probabilitas masing-masing kriteria Risk Rating 5 (R5)

Tabel 23. Probabilitas Risk Rating 5 Pendapatan Setahun (Juta Rupiah)

Penghasilan (Juta Rupiah)	R5	Probabilitas
83	1	0.071428571
95	1	0.071428571
101	1	0.071428571
111	1	0.071428571
120	2	0.142857143
128	1	0.071428571
132	1	0.071428571
137	1	0.071428571
175	3	0.214285714
182	1	0.071428571
193	1	0.071428571
Jumlah	14	1

Tabel 24. Probabilitas Risk Rating 5 KPR Aktif

KPR Aktif	R5	Probabilitas
ya	12	0.857142857
tidak	2	0.142857143
Jumlah	14	1

Tabel 25. Probabilitas Risk Rating 5 Durasi Pinjaman

Durasi Pinjaman	R5	Probabilitas
12	3	0.214285714
24	3	0.214285714
36	4	0.285714286
48	4	0.285714286
Jumlah	14	1

Tabel 26. Probabilitas Risk Rating 5 Jumlah Tanggungan

Jumlah Tanggungan	R5	Probabilitas
5	8	0.571428571
6	6	0.428571429
Jumlah	14	1

Tabel 27. Probabilitas Risk Rating 5 Rata-rata *Overdue*

Rata-rata	R5	Probabilitas
> 90 days	14	1
Jumlah	14	1

Setelah masing-masing probabilitas diketahui, langkah selanjutnya adalah menghitung data testing berdasarkan data training. Berdasarkan tabel 2, data yang dihitung manual adalah 5 data. Perhitungan adalah sebagai berikut :

$$\begin{aligned}
 (X_i|R1) &= (Pendapatan\ Setahun|R1) \times (KPR\ Aktif|R1) \times (Durasi\ Pinjaman|R1) \times \\
 &\quad (Jumlah\ Tanggungan|R1) \times (Rata - rata\ Overdue|R1) \\
 (AGR - 009970|R1) &= 0.043478261 \times 0.913043478 \times 0.173913043 \times 0.391304348 \times 1 \\
 &= 0.002701534
 \end{aligned}$$

$$(X_i|R2) = (Pendapatan\ Setahun|R2) \times (KPR\ Aktif|R2) \times (Durasi\ Pinjaman|R2) \\ \times (Jumlah\ Tanggungan|R2) \times (Rata - rata\ Overdue|R2) \\ (AGR - 009970|R2) = 0.04 \times 0.96 \times 0.28 \times 0.24 \times 0 = 0$$

$$(X_i|R3) = (Pendapatan\ Setahun|R3) \times (KPR\ Aktif|R3) \times (Durasi\ Pinjaman|R3) \\ \times (Jumlah\ Tanggungan|R3) \times (Rata - rata\ Overdue|R3) \\ (AGR - 009970|R3) = 0.045454545 \times 0.090909091 \times 0.181818182 \times 0.181818182 \times 0 \\ = 0$$

$$(X_i|R4) = (Pendapatan\ Setahun|R4) \times (KPR\ Aktif|R4) \times (Durasi\ Pinjaman|R4) \\ \times (Jumlah\ Tanggungan|R4) \times (Rata - rata\ Overdue|R4) \\ (AGR - 009970|R4) = 0 \times 0 \times 0.0625 \times 0.0625 \times 0 = 0$$

$$(X_i|R5) = (Pendapatan\ Setahun|R5) \times (KPR\ Aktif|R5) \times (Durasi\ Pinjaman|R5) \\ \times (Jumlah\ Tanggungan|R5) \times (Rata - rata\ Overdue|R5) \\ (AGR - 009970|R5) = 0.142857143 \times 0.142857143 \times 0.285714286 \times 0 \times 0 = 0$$

Perhitungan diatas menggunakan data testing AGR-009970. Untuk menghitung data testing lainnya menggunakan cara yang sama. Probabilitas akhir setiap kelas untuk masing-masing data testing dapat dilihat pada tabel 28.

Tabel 28. Probabilitas Akhir setiap Kelas

No	Kode Kontrak	R1	R2	R3	R4	R5
1	AGR - 009970	0.002701534	0	0	0	0
2	AGR - 009976	0.002626491	0	0	0	0
3	AGR - 010246	0	0	0	0	0.029987505
4	AGR - 010415	0	0	0	0.003662109	0
5	AGR - 010498	0.000525298	0	0	0	0

Setelah melakukan perhitungan probabilitas akhir setiap kelas, selanjutnya adalah menghitung Probabilitas Akhir didapat melalui perhitungan berikut :

$$P(C|X) = P(X|C) \times P(C)$$

$$P(R1|AGR - 009970) = P(X|R1) \times P(Risk\ Rating\ 1) \\ P(R1|AGR - 009970) = 0.002701534 \times 0.23 = 0.000621353$$

$$P(R2|AGR - 009970) = P(X|R2) \times P(Risk\ Rating\ 2) \\ P(R2|AGR - 009970) = 0 \times 0.25 = 0$$

$$P(R3|AGR - 009970) = P(X|R3) \times P(Risk\ Rating\ 3) \\ P(R3|AGR - 009970) = 0 \times 0.22 = 0$$

$$P(R4|AGR - 009970) = P(X|R4) \times P(Risk\ Rating\ 4) \\ P(R4|AGR - 009970) = 0 \times 0.16 = 0$$

$$P(R5|AGR - 009970) = P(X|R5) \times P(Risk\ Rating\ 5) \\ P(R5|AGR - 009970) = 0 \times 0.14 = 0$$

Perhitungan diatas menggunakan data testing AGR-009970. Menghitung probabilitas akhir data testing lainnya menggunakan cara yang sama. Probabilitas akhir masing-masing data testing dapat dilihat pada tabel 29.

Tabel 29. Probabilitas Akhir

No	Kode Kontrak	R1	R2	R3	R4	R5
1	AGR – 009970	0.000621353	0	0	0	0
2	AGR – 009976	0.000604093	0	0	0	0
3	AGR – 010246	0	0	0	0	0.004198251
4	AGR – 010415	0	0	0	0.000585938	0
5	AGR – 010498	0.000120819	0	0	0	0

Setelah melakukan perhitungan probabilitas akhir, hasil yang ditetapkan ke dalam kelas  $C$  adalah yang mempunyai nilai  $P(X|C)$ .  $P(C)$  maksimum, maka dapat disimpulkan :

$$AGR - 009970 = R1 > R2; R1 > R3; R1 > R4; R1 > R5$$

Jadi, AGR-009970 maksimum di R1.

$$AGR - 009976 = R1 > R2; R1 > R3; R1 > R4; R1 > R5$$

Jadi, AGR-009976 maksimum di R1.

$$AGR - 010246 = R5 > R1; R5 > R2; R5 > R3; R5 > R4$$

Jadi, AGR-010246 maksimum di R5.

$$AGR - 010415 = R4 > R1; R4 > R2; R4 > R3; R4 > R5$$

Jadi, AGR-010415 maksimum di R4.

$$AGR - 010498 = R1 > R2; R1 > R3; R1 > R4; R1 > R5$$

Jadi, AGR-010498 maksimum di R1.

Risk Rating atau peringkat resiko dari perhitungan manual 5 data yang diambil dari 100 data testing dapat dilihat pada tabel 30.

Tabel 30. Risk rating data testing perhitungan manual

Kode Kontrak	Risk Rating
AGR – 009970	1
AGR – 009976	1
AGR – 010246	5
AGR – 010415	4
AGR – 010498	1

## B. Implementasi klasifikasi Naïve Bayes pada R Studio

Data yang di analisa dalam Software R Studio adalah 800 data training dan 100 data testing. Hasil yang ditampilkan adalah sebagai berikut :

Tabel 31. *Confusion Matrix and statistics* data Training pada R Studio

Prediksi	1	2	3	4	5
1	90	0	0	0	0
2	0	65	0	0	0
3	0	0	121	0	0
4	0	0	0	57	0
5	0	1	0	0	40

Pada tabel 31 dijelaskan bahwa kelas risk rating 1 memperdiksi risk rating 1

sebanyak 90 data, kelas risk rating 2 memperdiksi risk rating 2 sebanyak 65 data, kelas risk rating 3 memperdiksi risk rating 3 sebanyak 121 data, kelas Risk rating memperdiksi risk rating 4 sebanyak 57 data dan kelas Risk rating 5 memperdiksi risk rating 5 sebanyak 40. Dapat dilihat bahwa ada risk rating 5 memperdiksi risk rating 2 sebanyak 1 data, artinya 1 data yang gagal di prediksi pada risk rating 5.

Tabel 32. *Accuracy* dari model yang digunakan pada R Studio

Accuracy	0.9973
95% CI	(0.9852, 0.9999)
No Information Rate	0.3235
P-Value [Acc > NIR]	< 2.2e-16
Kappa	0.9965
McNemar's Test P-Value	NA

Tabel 33. Informasi *Statistics by Class* pada R Studio

	Class : 1	Class : 2	Class : 3	Class : 4	Class : 5
Sensitivity	1.0000	0.9848	1.0000	1.0000	1.0000
Specificity	1.0000	1.0000	1.0000	1.0000	0.9970
Pos Pred Value	1.0000	1.0000	1.0000	1.0000	0.9756
Neg Pred Value	1.0000	0.9968	1.0000	1.0000	1.0000
Prevalence	0.2406	0.1765	0.3235	0.1524	0.1070
Detection Rate	0.2406	0.1738	0.3235	0.1524	0.1070
Detection Prevalence	0.2406	0.1738	0.3235	0.1524	0.1096
Balanced Accuracy	1.0000	0.9924	1.0000	1.0000	0.9985

Pada tabel 32 dapat dilihat bahwa kemampuan memprediksi kelas secara keseluruhan adalah 0.997, artinya dari segi akurasi model yang digunakan sudah sangat bagus.

Hasil prediksi 100 data testing menggunakan Algoritma Nive Bayes pada *software* R Studio dapat dilihat pada tabel 33 :

Tabel 33. Hasil prediksi data testing pada Software RStudio

No	Kode Kontrak	Risk rating
1	AGR-009589	1
2	AGR-009608	2
3	AGR-009621	3
4	AGR-009634	4
5	AGR-009642	2
6	AGR-009655	3
7	AGR-009667	4
8	AGR-009676	4
9	AGR-009683	4
...	...	...
100	AGR-010790	1

## KESIMPULAN

Algoritma Naïve Bayes bisa digunakan untuk menghitung dan mengelompokkan nilai resiko kredit Bank. Dari 900 koleksi data, 800 data digunakan sebagai data training dan 100 data sebagai data testing dengan 5 kelas *Risk rating*. Dari data testing sebanyak 100 data, klasifikasi dengan risk rating 1 sebanyak 23 data, klasifikasi dengan risk rating 2 sebanyak 13 data, klasifikasi dengan risk rating 3 sebanyak 32 data, klasifikasi dengan risk rating 4 sebanyak 17 data dan klasifikasi risk rating 5 sebanyak 15 data. Pengujian

data pada *Software R Studio* dengan metode Naïve Bayes berhasil memprediksi Risk rating (nilai resiko) dengan *Accuracy* 0.9973 atau 99%. *Sensitivity class 1, class 3, class 4 dan class 5* 100%, dan *Sensitivity class 2* 98%. *Specificity class 1, class 2, class 3, class 4* 100% dan *Specificity class 5* 99%.

#### DAFTAR PUSTAKA

- [1] Septiawan, Bayu. 2016. Pelaksanaan Kredit Usaha Rakyat (KUR) ditinjau dari Undang-undang Nomor 7 Tahun 1992 Tentang Perbankan. Fakultas Hukum Universitas Muhammadiyah Jember.
- [2] Fure, Joey Allen. 2016. Fungsi Bank sebagai Lembaga Keuangan di Indonesia Menurut Undang-undang Nomor 10 Tahun 1998 Tentang Perbankan. *Jurnal Elektronik Bagian Hukum Pidana Fakultas Hukum Unsrat*. **5(4)**:117-121.
- [3] Andrianto. 2020. Manajemen Kredit. Jawa Timur : CV Penerbit Qiara Media
- [4] Widiantari, Ni, M.D. 2014. Pengaruh Penilaian Kredit terhadap Keputusan Pemberian Kredit pada BPR. *E-journal Bisma Universitas Pendidikan Ganesha Jurusan Manajemen*. **9**:74.
- [5] Dendawijaya, Lukman. 2005. Manajemen Perbankan. Bogor: Ghalia Indonesia
- [6] Damanik, Habibah Jayanti. 2019. Penerapan Algoritma Naive Bayes untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*
- [7] Fatchan, Muhamad. 2021. Analisa Terpilihnya Tri Rismaharini sebagai Menteri Sosial dengan Pendekatan Algoritma Naïve Bayes. *Journal of Practical Computer Science*. **1(2)**:52-53.
- [8] Sari, Desi Ratna. 2019. Penerapan Metode Naïve Bayes dalam Memprediksi Kepuasan Mahasiswa Terhadap Cara Pengajaran Dosen. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*