

Penerapan Algoritma Naive Bayes untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor

Habibah Jayanti Damanik¹, Eka Irawan², Irfan Sudahri Damanik³, Anjar Wanto⁴
STIKOM Tunas Bangsa Pematangsiantar
Jln. Jendral Sudirman Blok A No. 1,2,3 Pematangsiantar
Habibahjayanti104@gmail.com

Abstract - Credit companies provide rentals for the community to help obtain transportation vehicles such as motorbikes, but also require the approval received by companies such as bad payments in the payment of loans issued to companies that make money or financing. With the many credit risks, companies must be selective in choosing consumers who will be given credit so as not to make the company suffer losses. The method used in this study is the Naïve Bayes Algorithm and processed using Rapidminer Studio 5.3 software. The data used consisted of 55 training data and 10 test data. There are 9 variables used in this study, namely marital status, number of children, home ownership, housing conditions, employment, tenure, age, income, and down payment. The level of accuracy obtained from testing using Rapidminer Studio 5.3 is equal to 90%. By using the Naive Bayes algorithm in classifying consumers, it is hoped that it can assist companies in carrying out the mining process on the previous data and make the right decisions in determining new consumers from previous consumer data.

Keywords: Credit Risk, Classification, Naïve Bayes, and Data Mining.

Abstrak -Perusahaan kredit/leasing memberikan jasa kepada masyarakat untuk membantu mendapatkan kendaraan transportasi seperti sepeda motor, namun terdapat kendala-kendala yang diterima perusahaan seperti terjadinya macet dalam pembayaran kredit yang mengakibatkan kerugian yang berdampak bagi perusahaan yang menimbulkan kerugian atau resiko. Dengan banyaknya resiko kredit tersebut, perusahaan harus selektif dalam memilih konsumen yang akan diberi kredit agar tidak membuat perusahaan mengalami kerugian. Metode yang digunakan dalam penelitian ini adalah Algoritma Naïve Bayes dan diproses menggunakan software Rapidminer Studio 5.3. Data yang digunakan terdiri dari 55 data latih dan 10 data uji. Terdapat 9 variabel yang digunakan dalam penelitian ini yaitu status perkawinan, jumlah anak, kepemilikan rumah, kondisi rumah, pekerjaan, tenor, usia, penghasilan, dan uang muka. Tingkat akurasi yang didapatkan dari pengujian menggunakan Rapidminer Studio 5.3 yaitu sebesar 90%. Dengan menggunakan algoritma Naive Bayes dalam mengklasifikasikan konsumen diharapkan dapat membantu perusahaan dalam melakukan proses mining terhadap data sebelumnya dan mengambil keputusan dengan tepat dalam menentukan konsumen baru dari data-data konsumen sebelumnya.

Kata kunci: Resiko Kredit, Klasifikasi, Naïve Bayes, dan Data Mining.

1. PENDAHULUAN

Jumlah kendaraan bermotor di Indonesia meningkat setiap tahunnya, hal ini dikarenakan peningkatan kebutuhan akan mobilitas masyarakat yang meningkat, serta kualitas kendaraan umum yang rendah juga menambah daya beli masyarakat

secara keseluruhan. Hampir setiap rumah tangga memiliki paling tidak satu atau lebih kendaraan bermotor baik kendaraan bermotor roda dua maupun roda empat. Peningkatan atas kendaraan bermotor dijadikan peluang oleh lembaga pembiayaan untuk menawarkan produk konsumtif kendaraan bermotor yang banyak dikenal dengan pembiayaan kendaraan bermotor. Dalam menyalurkan dana kredit perusahaan leasing mendapati ada beberapa kredit yang dikatakan kurang lancar atau kredit macet yang kemudian akan berpengaruh kepada pemberian kredit selanjutnya atau juga bisa mempengaruhi kemampuan perusahaan dalam menyalurkan kredit. Banyak calon nasabah yang ingin melakukan kredit kendaraan bermotor dengan menggunakan jasa perusahaan leasing. Semakin banyak nasabah semakin tinggi juga tingkat resiko kredit yang diterima oleh perusahaan. Dengan banyaknya resiko kredit dalam bentuk macetnya atau ketidak lancarnya pembayaran kredit membuat perusahaan ini harus selektif dalam memilih konsumen yang akan didanai agar tidak membuat perusahaan mengalami kerugian. Salah satu cara yang mungkin bisa dilakukan untuk mencegah terjadinya kredit macet adalah dengan mengklasifikasikan calon konsumen yang good customer dan bad customer sehingga dapat mengurangi resiko yang terjadi .

2. METODOLOGI PENELITIAN

2.1 Analisa Data

Analisis data yang penulis gunakan pada penelitian ini menggunakan analisis kuantitatif. Teknik analisis data yang digunakan menggunakan jenis statistik inferensial. Statistik inferensial adalah statistik yang berkaitan dengan analisis data (sampel), kemudian diambil kesimpulan yang digeneralisasikan kepada seluruh populasi dan digunakan untuk menarik inferensi dari sampel ke populasi [1] Data yang diperoleh kemudian diolah dengan RapidMiner menggunakan Performance yang berfungsi sebagai validasi dan reabilitas data untuk mencari keakuratan data.

Tabel 1. Data Training

No	Nama	Usia	Status Perkawinan	Jlh Anak	Kepemilikan Rumah	Kondisi	Pekerjaan	Penghasilan	Uang muka	Tenor	Status
1	Sansuryadi	58	Menikah	4	Milik Sendiri	Permanen	PNS	5000000	0.38	1 thn	Good Cust
2	Eko Prianto	37	Menikah	1	Keluarga	Permanen	Karyawan	4000000	0.50	1 thn	Good Cust
3	Neni Sri W.	35	Menikah	1	Milik Sendiri	Permanen	Wiraswasta	3500000	0.50	1 thn	Good Cust
4	Nanang M.	41	Menikah	4	Kontrak	Permanen	Wiraswasta	3000000	0.55	2 thn	Bad Cust
5	Marice D.S.	39	Janda	2	Milik Sendiri	Semi Permanen	Wiraswasta	3500000	0.37	1 thn	Bad Cust
6	Dewi Hartati	40	Menikah	4	Milik Sendiri	Permanen	Wiraswasta	6000000	0.44	1 thn	Good Cust
7	Farida Nst	44	Menikah	3	Milik Sendiri	Permanen	PNS	5500000	0.36	2 thn	Good Cust
8	Deswani D.	45	Menikah	2	Milik Sendiri	Permanen	Wiraswasta	6000000	0.38	2 thn	Good Cust
9	Dewi Astuti	31	Menikah	1	Dinas	Permanen	Karyawan	4000000	0.44	1 thn	Good Cust
10	Purnama	33	Menikah	2	Milik Sendiri	Permanen	Wiraswasta	5000000	0.40	3 thn	Good

No	Nama	Usia	Status Perkawinan	Jlh Anak	Kepemilikan Rumah	Kondisi	Pekerjaan	Penghasilan	Uang muka	Tenor	Status
	Sari										Cust
...
50	Masrul Afandi	31	Menikah	3	Milik Sendiri	Permanen	Karyawan	3500000	0.39	1 thn	Good Cust
51	Sampetua Htb	51	Menikah	3	Milik Sendiri	Permanen	Wiraswasta	5000000	0.38	2 thn	Good Cust
52	Udur Lina B	59	Menikah	5	Milik Sendiri	Semi Permanen	Petani	5000000	0.81	1 thn	Good Cust
53	Zulhidayah B	28	Belum Menikah	0	Keluarga	Permanen	Karyawan	3500000	0.33	2 thn	Good Cust
54	Lasmawati S	44	Menikah	2	Kontrak	Permanen	Wiraswasta	5500000	0.38	1 thn	Bad Cust
55	Nani Susanti	30	Menikah	2	Kontrak	Permanen	Wiraswasta	3000000	0.30	2 thn	Bad Cust

2.2 Kredit dan Resiko Kredit

Kredit merupakan suatu fasilitas keuangan yang memungkinkan seseorang atau badan usaha untuk meminjam uang untuk membeli produk dan membayarkannya kembali dalam jangka waktu yang ditentukan” [2]

Resiko kredit adalah suatu kerugian yang dapat berpotensi untuk menimbulkan penolakan atau ketidakmampuan konsumen kredit untuk membayar hutangnya secara penuh dan tepat waktu [3]

2.3 Naïve Bayes

Naïve Bayes adalah sebuah pengelompokan statistik yang bisa di dipakai untuk memprediksi probabilitas anggota suatu class. Naïve Bayes juga mempunyai akurasi dan kecepatan yang sangat kuat ketika diaplikasikan pada database dengan big data [4]. Naive Bayes merupakan salah satu algoritma Data mining seperti halnya K-Means [5]–[8].

Untuk menyelesaikan metode Naive Bayes dapat dilakukan dengan persamaan-persamaan sebagai berikut:

1. Hitung Jumlah dan probabilitas

Hitung nilai probabilitas tiap kategori yang sama, dengan cara jumlah data yang sesuai dari kategori yang sama lalu dibagi dengan jumlah data pada kategori tersebut untuk menemukan nilai probabilitistik. Jika terdapat data numerik, maka temukan nilai mean dan standar deviasi dari masing-masing parameter yang menggambarkan data angka. Rumus yang digunakan untuk menghitung nilai rata – rata hitung (mean) dapat dilihat sebagai berikut :

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

dimana :

μ : rata – rata hitung (mean)

x_i : nilai sample ke –i

n : jumlah sampel

Dan persamaan untuk menghitung nilai simpangan baku (standar deviasi) dapat dilihat dibawah ini:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n-1}}$$

Dimana :

σ : standar deviasi

x_i : nilai x ke $-i$

μ : rata-rata hitung

n : jumlah sampel

2. Nilai Probabilitas Fitur Setiap Kelas

Setelah menghitung nilai mean dan standar deviasi untuk fitur dengan data angka, maka selanjutnya mengetahui nilai probabilitas setiap fitur pada setiap kelasnya. Untuk mengetahui nilai probabilitas setiap fitur pada kelas kita harus hitung jumlah data yang sesuai dari kategori yang sama lalu dibagi dengan jumlah data pada kategori tersebut.

3. Nilai Distribusi Gaussian

Langkah selanjutnya adalah menghitung nilai probabilitas untuk fitur data uji yang mempunyai data numerik / angka. Berikut adalah persamaan dalam mencari nilai distribusi gaussian.

$$P = (X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} \times e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}}$$

Dimana :

P : Peluang

X_i : Atribut ke i

X_j : Nilai atribut ke i

C : Kelas yang dicari

C_i : Sub kelas Y yang dicari

μ : Menyatakan rata-rata dari seluruh atribut

σ : Deviasi standar, menyatakan varian dari seluruh atribut.

4. Probabilitas Akhir Setiap Kelas

Menghitung probabilitas akhir untuk setiap kelas artinya memasukkan semua data nilai distribusi gaussian yang ada ke dalam satu kelas yang sama.

$$P(X|Kelas) = P(V1|Kelas) \times P(V2|Kelas) \times P(V3|Kelas) \times P(V4|Kelas) \times P(V5|Kelas) \times P(V6|Kelas) \times P(V7|Kelas) \times P(V8|Kelas)$$

5. Probabilitas Akhir

Probabilitas akhir didapat melalui perhitungan nilai probabilitas akhir kelas ke dalam rumus Naïve Bayes Classifier. Perhitungan probabilitas akhir adalah sebagai berikut.

$$P(Kelas|X) = P(Kelas) * P(X)$$

3. HASIL DAN PEMBAHASAN

3.1. Analisa Algoritma Naïve Bayes

a. Menghitung Jumlah dan Probabilitas

Terdapat 9 variabel yang digunakan penulis yaitu status perkawinan, jumlah anak, kepemilikan rumah, kondisi rumah, pekerjaan, tenor, usia, penghasilan, dan uang muka. Hasil Perhitungan probabilitas masing-masing kriteria dapat dilihat pada beberapa tabel-tabel berikut.

Tabel 2. Probabilitas Status Perkawinan

Status Pernikahan	Jumlah Kejadian "Dipilih"		Probabilitas	
	Good Cust	Bad Cust	Good Cust	Bad Cust
Menikah	31	12	0.794872	0.750000
Belum	7	2	0.179487	0.125000
Janda/Duda	1	2	0.025641	0.125000
Jumlah	39	16	1	1

Tabel 3. Probabilitas Jumlah Anak

Jumlah Anak	Jumlah Kejadian "Dipilih"		Probabilitas	
	Good Cust	Bad Cust	Good Cust	Bad Cust
0	7	2	0.179487	0.125000
1	7	3	0.179487	0.187500
2	8	4	0.205128	0.250000
3	12	3	0.307692	0.187500
4	4	2	0.102564	0.125000
5	1	2	0.025641	0.125000
>5	0	0	0.000000	0.000000
Jumlah	39	16	1	1

Tabel 4. Probabilitas Kepemilikan Rumah

Kepemilikan Rumah	Jumlah Kejadian "Dipilih"		Probabilitas	
	Good Cust	Bad Cust	Good Cust	Bad Cust
Milik Sendiri	22	6	0.564103	0.375000
Keluarga	11	2	0.282051	0.125000
Kontrak	3	8	0.076923	0.500000
Dinas	3	0	0.076923	0.000000
Jumlah	39	16	1	1

Tabel 5. Probabilitas Kondisi Rumah

Kondisi rumah	Jumlah Kejadian "Dipilih"		Probabilitas	
	Good Cust	Bad Cust	Good Cust	Bad Cust
Permanen	31	9	0.794872	0.562500
Semi Permanen	8	7	0.205128	0.437500
Jumlah	39	16	1	1

Tabel 6. Probabilitas Pekerjaan

Pekerjaan	Jumlah Kejadian "Dipilih"		Probabilitas	
	Good Cust	Bad Cust	Good Cust	Bad Cust
PNS	6	0	0.153846	0.000000
Karyawan	10	3	0.256410	0.187500
wiraswasta	20	11	0.512821	0.687500
Guru	1	1	0.025641	0.062500
Petani	2	0	0.051282	0.000000
Pensiunan	0	1	0.000000	0.062500
Jumlah	39	16	1	1

Tabel 7. Probabilitas Tenor

Tenor	Jumlah Kejadian "Dipilih"		Probabilitas	
	Good Cust	Bad Cust	Good Cust	Bad Cust
6 bulan	0	0	0.000000	0.000000
1 tahun	18	5	0.461538	0.312500

Tenor	Jumlah Kejadian "Dipilih"		Probabilitas	
	Good Cust	Bad Cust	Good Cust	Bad Cust
18 bulan	4	4	0.102564	0.250000
2 tahun	16	7	0.410256	0.437500
3 tahun	1	0	0.025641	0.000000
Jumlah	39	16	1	1

b. Menghitung Nilai *Mean* dan Standar Deviasi.

Dalam penelitian ini digunakan dua kelas yaitu *Good Cust* dan *Bad Cust*. Sebelum mencari nilai mean dan standar deviasi, kita harus mengelompokkan terlebih dahulu jumlah kelas *Good Cust* dan *Bad Cust* berdasarkan data training yang digunakan. Untuk menghitung nilai mean di gunakan persamaan (5). Perhitungannya dapat dilihat sebagai berikut :

$$\text{Good Cust} = 39$$

$$\text{Bad Cust} = 16$$

Menghitung Nilai Mean

$$\mu(\text{Good Cust}) = (58+37+35+40+44+45+31+33+45+37+52+55+58+71+48+48+49+35+29+30+40+38+34+27+44+29+22+58+26+39+45+39+27+31+27+31+51+59+28)/39$$

$$\mu(\text{Good Cust}) = (1575)/39$$

$$\mu(\text{Good Cust}) = 40,3846$$

$$\mu(\text{Bad Cust}) = (41+39+27+38+28+49+34+45+28+58+38+46+0+21+44+30)/16$$

$$\mu(\text{Bad Cust}) = (606)/16$$

$$\mu(\text{Bad Cust}) = 37,8750$$

Data diatas diambil dari data pada variabel usia dengan hasil di kelas *good cust*, sehingga didapatkan nilai μ nya. Untuk perhitungan nilai μ dengan hasil *bad cust* pada variabel usia langkah-langkah nya sama. Sedangkan untuk menghitung nilai μ pada variabel lainnya langkah-langkah yang digunakan juga sama seperti menghitung nilai mean pada variabel usia.

c. Menghitung Standar Deviasi

Setelah mendapatkan nilai μ pada masing-masing variabel, maka selanjutnya kita mencari nilai standar deviasi, Untuk menghitung nilai standar deviasi digunakan persamaan (2). Berikut adalah proses menghitung nilai standar deviasi :

$$\sigma(\text{Good}) = \sqrt{\frac{(58 - 40,3846)^2 + (37 - 40,3846)^2 + (35 - 40,3846)^2 + (40 - 40,3846)^2 + (44 - 40,3846)^2 + (45 - 40,3846)^2 + (37 - 40,3846)^2 + (52 - 40,3846)^2 + (55 - 40,3846)^2 + (58 - 40,3846)^2 + (71 - 40,3846)^2 + (48 - 40,3846)^2 + (48 - 40,3846)^2 + (49 - 40,3846)^2 + (35 - 40,3846)^2 + (29 - 40,3846)^2 + (30 - 40,3846)^2 + (40 - 40,3846)^2 + (38 - 40,3846)^2 + (34 - 40,3846)^2 + (27 - 40,3846)^2 + (44 - 40,3846)^2 + (29 - 40,3846)^2 + (22 - 40,3846)^2 + (58 - 40,3846)^2 + (26 - 40,3846)^2 + (39 - 40,3846)^2 + (45 - 40,3846)^2 + (39 - 40,3846)^2 + (27 - 40,3846)^2 + (31 - 40,3846)^2 + (27 - 40,3846)^2 + (31 - 40,3846)^2 + (51 - 40,3846)^2 + (59 - 40,3846)^2 + (28 - 40,3846)^2}{39 - 1}}$$

$$\sigma(\text{Good Cust}) = \sqrt{5009,2308/38}$$

$$\sigma(\text{Good Cust}) = \sqrt{131,8219}$$

$$\sigma(\text{Good Cust}) = 11,414$$

$$\sigma(\text{Bad Cust}) = \sqrt{\frac{(41 - 37,8750)^2 + (39 - 37,8750)^2 + (27 - 37,8750)^2 + (38 - 37,8750)^2 + (28 - 37,8750)^2 + (49 - 37,8750)^2 + (34 - 37,8750)^2 + (45 - 37,8750)^2 + (28 - 37,8750)^2 + (58 - 37,8750)^2 + (38 - 37,8750)^2 + (46 - 37,8750)^2 + (40 - 37,8750)^2 + (21 - 37,8750)^2 + (44 - 37,8750)^2 + (30 - 37,8750)^2}{16 - 1}}$$

$$\sigma(\text{Bad Cust}) = \sqrt{1373,7500/15}$$

$$\sigma(\text{Bad Cust}) = \sqrt{91,5833}$$

$$\sigma(\text{Good Cust}) = 9,5699$$

Data diatas diambil dari data pada variabel Usia dengan hasil dikelas *Good Cust* dan *Bad Cust*, sehingga didapatkan nilai σ nya. Untuk perhitungan nilai σ dengan hasil *Good Cust* dan *Bad Cust* pada varibel lainnya langkah-langkah nya sama.

d. Nilai Probabilitas Fitur Setiap Kelas

Setelah menghitung nilai mean dan standar deviasi untuk data angka, maka selanjutnya menentukan nilai probabilitas setiap fitur pada setiap kelasnya. Untuk mengetahui nilai probabilitas setiap fitur pada kelas kita harus hitung jumlah data yang sesuai dari kategori yang sama lalu dibagi dengan jumlah data pada kategori tersebut. Jumlah data yang digunakan adalah sebanyak 48 data. Data *Good Cust* ada 34, data *Bad Cust* ada 14. Sehingga untuk menghitung nilai probabilitas nya adalah sebagai berikut.

$$P(\text{Good Cust}) = 39/55$$

$$P(\text{Good Cust}) = 0,709$$

$$P(\text{Bad Cust}) = 16/55$$

$$P(\text{Bad Cust}) = 0,291$$

Setelah mencari nilai probabilitas tiap fitur, maka didapat tabel untuk nilai, mean, standar deviasi dan probabilitas untuk masing-masing variabel, bisa dilihat pada tabel berikut.

Tabel 8. Probabilitas Usia

Nilai	Good Cust	Bad Cust
μ Mean	40.3846	37.8750
σ Deviasi ²	131.8219	91.5833
σ Deviasi	11.4814	9.5699

Tabel 9. Probabilitas Penghasilan

Nilai	Good Cust	Bad Cust
μ Mean	4717948.7179	2937500
σ Deviasi ²	1878879892038	1129166666666.6700
σ Deviasi	1370722.3979	1062622.5420

Tabel 10. Probabilitas Uang Muka

Nilai	Good Cust	Bad Cust
μ Mean	0.4017	0.3406
σ Deviasi ²	0.0117	0.0051
σ Deviasi	0.1083	0.0714

e. Menghitung Nilai Distribusi *Gaussian*.

Langkah selanjutnya adalah menghitung nilai distribusi *gaussian* untuk *data testing* yang mempunyai data numerik / angka.

Tabel 11. Data Testing

No	Nama	Usia	Status Perkawinan	Jlh Anak	Kepemilikan Rumah	Kondisi	Pekerjaan	Penghasilan	Uang muka	Tenor	Status
1	Chicie W.	27	Belum Menikah	0	Keluarga	Permanen	Karyawan	2700000	0.33	18 bulan	Good Cust
2	Adek K.	29	Belum Menikah	0	Keluarga	Permanen	Karyawan	2500000	0.45	1 tahun	Good Cust
3	Rimmawati N.	46	Menikah	5	Milik Sendiri	Semi Permanen	Wiraswasta	3000000	0.55	2 tahun	Bad Cust
4	Susilawati	41	Menikah	3	Milik Sendiri	Permanen	Wiraswasta	4000000	0.33	2 tahun	Good Cust
5	Nurlela	29	Belum Menikah	0	Keluarga	Permanen	Wiraswasta	3000000	0.33	18 bulan	Good Cust
6	Mhd. Afandi	28	Belum Menikah	0	Keluarga	Permanen	Karyawan	2700000	0.30	2 tahun	Good Cust
7	Rusmina P.	34	Menikah	2	Keluarga	Permanen	Karyawan	2500000	0.31	2 tahun	Good Cust
8	Rina Siahaan	52	Menikah	3	Milik Sendiri	Permanen	Wiraswasta	4000000	0.31	2 tahun	Good Cust
9	Ruhayati	57	Janda	1	Milik Sendiri	Permanen	PNS	4000000	0.30	2 tahun	Good Cust
10	Ezra Teguh	24	Belum Menikah	0	Keluarga	Permanen	Karyawan	3000000	0.54	18 bulan	Good Cust

Untuk menghitung nilai distribusi gaussian, menggunakan persamaan (3). Berikut adalah proses menghitung nilai distribusi gaussian :

$$P(\text{Usia} = 27 \mid \text{Good Cust}) = \frac{1}{\sqrt{2\pi \cdot 11,4814}} * e^{-\frac{(27-40,3846)^2}{(2 \cdot 131,8219)}}$$

$$P(\text{Usia} = 27 \mid \text{Good Cust}) = \frac{1}{8,4913488352} * 2,718281^{-0,6795076545}$$

$$P(\text{Usia} = 27 \mid \text{Good Cust}) = \frac{1}{8,4913488352} * 0.5075599108$$

$$P(\text{Usia} = 27 \mid \text{Good Cust}) = 0.0597737675$$

f. Probabilitas Akhir Setiap kelas

Menghitung nilai probabilistik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut, dapat dilihat menggunakan persamaan (4).

$$P(X \mid \text{Good Cust}) = P(\text{Usia} \mid \text{Good Cust}) * P(\text{Status Perkawinan} \mid \text{Good Cust}) * P(\text{Jumlah Anak} \mid \text{Good Cust}) * P(\text{Kepemilikan Rumah} \mid \text{Good Cust}) * P(\text{Kondisi Rumah} \mid \text{Good Cust}) * P(\text{Pekerjaan} \mid \text{Good Cust}) * P(\text{Penghasilan} \mid \text{Good Cust}) * P(\text{Uang Muka} \mid \text{Good Cust}) * P(\text{Tenor} \mid \text{Good Cust})$$

$$P(X \mid \text{Good Cust}) = 0.059773767 * 0.179487 * 0.179487 * 0.282051 * 0.794872 * 0.256410 * 0.0001155757 * 1.0387950234 * 0.102564$$

$$P(X \mid \text{Good Cust}) = 0.000000001363$$

Untuk menghitung probabilitas akhir setiap kelas dengan hasil *Bad Cust* menggunakan langkah penyelesaian yang sama.

$$P(X \mid \text{Bad Cust}) = P(\text{Usia} \mid \text{Bad Cust}) * P(\text{Status Perkawinan} \mid \text{Bad Cust}) * P(\text{Jumlah Anak} \mid \text{Bad Cust}) * P(\text{Kepemilikan Rumah} \mid \text{Bad Cust}) * P(\text{Kondisi Rumah} \mid \text{Bad Cust}) * P(\text{Pekerjaan} \mid \text{Bad Cust}) * P(\text{Penghasilan} \mid \text{Bad Cust}) * P(\text{Uang Muka} \mid \text{Bad Cust}) * P(\text{Tenor} \mid \text{Bad Cust})$$

$$P(X \mid \text{Bad Cust}) = 0.067720221 * 0.125000 * 0.125000 * 0.125000 * 0.562500 * 0.187500 * 0.000377577 * 1.487042598 * 0.250000$$

$$P(X \mid \text{Bad Cust}) = 0.000000001958$$

Penyelesaian diatas menggunakan satu sampel data uji yaitu alternatif Chicie Wardani. Probabilitas akhir setiap kelas untuk masing-masing alternatif, dapat dilihat pada tabel berikut.

Tabel 12. Probabilitas Akhir Setiap Kelas

No	Nama	Good Cust	Bad Cust
1	Chicie Wardani	0.000000001363	0.000000001958
2	Adek Kurniawan	0.000000006288	0.000000001069
3	Sriadi	0.000000080554	0.000000355777
4	Susilawati	0.000000837748	0.000000381433
5	Nurlela	0.00000002214	0.00000002485
6	Muhammad Afandi	0.000000004983	0.000000003399
7	Rusmina Panjaitan	0.000000037072	0.000000045492
8	Rina Siahaan	0.000000438666	0.000000125457
9	Ruhayati	0.000000001599	0.000000000000
10	Ezra Teguh Wastika	0.000000000823	0.000000000068

g. Probabilitas Akhir

Probabilitas akhir didapat melalui perhitungan menggunakan persamaan (5). Perhitungan probabilitas akhir adalah sebagai berikut.

$$P(\text{Good Cust}|X) = P(\text{Good Cust}) * P(X)$$

$$P(\text{Good Cust}|X) = 0,709 * 0.000000001363$$

$$P(\text{Good Cust}|X) = 0.000000000967$$

$$P(\text{Bad Cust}|X) = P(\text{Bad Cust}) * P(X)$$

$$P(\text{Bad Cust}|X) = 0,291 * 0.000000001958$$

$$P(\text{Bad Cust}|X) = 0.000000000570$$

Penyelesaian diatas menggunakan satu sampel data uji yaitu alternatif Suryanto. Probabilitas akhir untuk masing-masing alternatif, dapat dilihat pada tabel berikut.

Tabel 13. Probabilitas Akhir

No	Nama	Good Cust	Bad Cust
1	Chicie Wardani	0.000000000967	0.000000000570
2	Adek Kurniawan	0.000000004459	0.000000000311
3	Sriadi	0.000000571202	0.000001034989
4	Susilawati	0.000000594040	0.000000110962
5	Nurlela	0.000000001570	0.000000000723
6	Muhammad Afandi	0.000000003534	0.000000000989
7	Rusmina Panjaitan	0.000000026287	0.000000013234
8	Rina Siahaan	0.000000311054	0.000000036497
9	Ruhayati	0.000000001133	0.000000000000
10	Ezra Teguh Wastika	0.000000000584	0.000000000020

3.1 Hasil

Hasil pengujian Model Algoritma Naive Bayes Classifier ditunjukan pada gambar berikut:

accuracy: 90.00%		
	true Good Cust	true Bad Cust
pred Good Cust	9	1
pred Bad Cust	0	0
class recall	100.00%	0.00%
		class precision
		90.00%
		0.00%

Gambar 1. Nilai Accuracy Performance Data Testing

Keterangan :

1. Jumlah prediksi Good Cust dan kenyataannya benar Good Cust adalah 9 record.
2. Jumlah prediksi Bad Cust dan kenyataannya benar Good Cust adalah 0 record.
3. Jumlah prediksi Good Cust dan kenyataannya benar Bad Cust adalah 1 record.
4. Jumlah prediksi Bad Cust dan kenyataannya benar Bad Cust adalah 0 record.

Pada gambar 1. Nilai Accuracy sebesar 90 %. class precision pada prediksi Good Cust memiliki nilai 90%, sedangkan pada prediksi Bad Cust memiliki nilai 0

%. Class recall pada true Good Cust memiliki nilai 100%, sedangkan pada true Bad Cust memiliki nilai 0%.

4. KESIMPULAN

Berdasarkan pembahasan diatas dapat disimpulkan bahwa :

1. Penerapan Data Mining menggunakan algoritma naïve 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.
2. Pengujian data pada Rapiminer 5.3 dengan menggunakan metode Naive Bayes berhasil menampilkan dua kelas dari hasil klasifikasi dengan persentase keakuratan sebesar sebesar 90 %.

DAFTAR PUSTAKA

- [1] M. R. Munandar, E. S. Astuti, and M. S. Hakam, "(Studi Pada Pekerja bagian Produksi PT . SEKAWAN KARYATAMA MANDIRI Sidoarjo)," vol. 9, no. 1, pp. 1–9, 2014.
- [2] F. Zulfami, "ANALISA DAN PERANCANGAN APLIKASI DATA MINING PENENTUAN RESIKO KREDIT KEPEMILIKAN KENDARAAN BERMOTOR MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR," *Jurnal Infokar*, vol. 1, no. 1, pp. 32–39, 2017.
- [3] D. J. F. Rozi and K. Yahya, "Analisa Risiko Kredit Sepeda Motor Pada PT . X Finance (Studi Kasus Kantor Cabang Wilayah Gresik dan Lamongan)," *JURNAL SAINS DAN SENI POMITS*, vol. 2, no. 2, pp. 231–236, 2013.
- [4] R. Wijayatun and Y. Sulisty, "Prediksi Rating Film Menggunakan Metode Naïve Bayes," *Jurnal Teknik Elektro*, vol. 8, no. 2, pp. 60–63, 2016.
- [5] I. Parlina, A. P. Windarto, A. Wanto, and M. R. Lubis, "Memanfaatkan Algoritma K-Means dalam Menentukan Pegawai yang Layak Mengikuti Asessment Center untuk Clustering Program SDP," *CESS (Journal of Computer Engineering System and Science)*, vol. 3, no. 1, pp. 87–93, 2018.
- [6] M. G. Sadewo, A. P. Windarto, and A. Wanto, "Penerapan Algoritma Clustering dalam Mengelompokkan Banyaknya Desa/Kelurahan Menurut Upaya Antisipasi/ Mitigasi Bencana Alam Menurut Provinsi dengan K-Means," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 2, no. 1, pp. 311–319, 2018.
- [7] S. Sudirman, A. P. Windarto, and A. Wanto, "Data Mining Tools | RapidMiner : K-Means Method on Clustering of Rice Crops by Province as Efforts to Stabilize Food Crops In Indonesia," *IOP Conference Series: Materials Science and Engineering*, vol. 420, no. 12089, pp. 1–8, 2018.
- [8] R. W. Sari, A. Wanto, and A. P. Windarto, "Implementasi Rapidminer dengan Metode K-Means (Study Kasus : Imunisasi Campak pada Balita Berdasarkan Provinsi)," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 2, no. 1, pp. 224–230, 2018.