

# IMPLEMENTASI KLASIFIKASI *NAIVE BAYES* UNTUK PREDIKSI KELAYAKAN PEMBERIAN PINJAMAN PADA KOPERASI ANUGERAH BINTANG CEMERLANG

Siti Lestari<sup>1</sup>, Akmaludin<sup>2</sup>, Mohammad Badrul<sup>3</sup>

<sup>1,2,3</sup>Program Studi Sistem Informasi, STMIK Nusa Mandiri Jakarta

<sup>1</sup>[dilestha.astuaria.daffostl@gmail.com](mailto:dilestha.astuaria.daffostl@gmail.com), <sup>2</sup>[akmaludiin.akm@nusamandiri.ac.id](mailto:akmaludiin.akm@nusamandiri.ac.id),

<sup>3</sup>[mohammad.mbl@nusamandiri.ac.id](mailto:mohammad.mbl@nusamandiri.ac.id)

**Abstrak** - Pendapatan utama yang diperoleh koperasi ditentukan oleh besarnya jumlah angsuran yang dibayar oleh nasabah, sedangkan permasalahan yang sering dihadapi adalah banyaknya nasabah yang menunggak dalam membayar angsuran bahkan tidak jarang nasabah yang kabur sehingga menjadi piutang tak tertagih yang pada akhirnya menyebabkan kerugian. Hal tersebut terjadi akibat kurang akuratnya manajemen dalam menentukan pemohon mana yang layak dan tidak layak diberikan pinjaman. Oleh karena itu, penulis menerapkan metode *data mining* untuk mengklasifikasikan kelayakan nasabah dalam kategori layak dan tidak layak berdasarkan data historis nasabah di masa sebelumnya, kemudian digunakan dalam memprediksi kelayakan nasabah di masa depan, yaitu dengan algoritma *Naive Bayes*. Penulis menggunakan aplikasi *RapidMiner* 8.1 untuk menguji tingkat akurasi algoritma tersebut. Pengujian dilakukan dengan menyiapkan *data training* sebanyak 438 dan *data testing* sebanyak 146 data yang diambil dari *database* nasabah di tahun 2015. Hasil pengujian akurasi yang didapat menggunakan aplikasi *RapidMiner* maupun perhitungan manual dengan empat kali proses diperoleh akurasi tertinggi yaitu 78.08% dengan presentase eror 21.92 %. Jadi, dapat disimpulkan bahwa Algoritma *Naive Bayes* dapat diterapkan untuk memprediksi kelayakan pemberian pinjaman pada Koperasi Anugerah Bintang Cemerlang.

**Kata Kunci:** Data Mining, Kelayakan Pinjaman, Klasifikasi, *Naive Bayes*

## I. PENDAHULUAN

Koperasi merupakan salah satu lembaga keuangan yang melakukan transaksi kredit. Dalam rangka mengembangkan usahanya dalam bentuk pinjaman anggota, koperasi memiliki prinsip kehati-hatian sebagaimana yang diterapkan pada perbankan dengan melakukan analisis pemberian kredit kepada calon anggotanya (Pandie, 2018). Adapun Koperasi Anugerah Bintang Cemerlang ini merupakan koperasi simpan pinjam yang bergerak dalam bidang pembiayaan atau pemberian fasilitas pinjaman kepada nasabah atau anggotanya, dengan sistem pembayaran dicicil berdasarkan tenor angsuran yang diambil.

Sebagai lembaga keuangan, pendapatan utama yang diperoleh koperasi ditentukan oleh besarnya jumlah angsuran yang dibayar oleh nasabah. Sedangkan permasalahan yang sering dihadapi oleh koperasi adalah banyaknya nasabah yang menunggak dalam membayar angsuran bahkan tidak jarang nasabah yang kabur sehingga menjadi piutang tak tertagih yang pada akhirnya menyebabkan kerugian. Hal ini timbul karena pada tahap evaluasi dalam proses pemberian pinjaman, pihak koperasi masih kesulitan untuk menentukan pemohon mana yang layak mendapatkan fasilitas pinjaman dan tidak beresiko menyebabkan pinjaman macet. Banyaknya pemohon kredit yang mengajukan kredit dengan kondisi ekonomi yang berbeda-beda menuntut kejelian *Credit Analyst*

dalam pengambilan keputusan. Kasus kredit macet dapat diminimalisir tergantung dari kinerja *Credit Analyst* dalam proses menentukan nasabah kredit (Elyana, 2017).

Dalam upaya membantu menganalisis kelayakan pemberian pinjaman kepada nasabah, diperlukan analisis kredit menggunakan *data mining* yang dapat mengklasifikasikan nasabah mana yang termasuk dalam kategori layak dan tidak layak. Untuk menemukan pola tersebut dari *database* koperasi, dibutuhkan suatu metode *data mining*, salah satunya adalah algoritma *naive bayes*. *Naive Bayes* berjalan baik dibanding dengan model klasifikasi lainnya, dibuktikan dengan penelitian berjudul “*Comparative Study Of Data mining Model For Credit Card Application Scoring In Bank*”, dimana klasifikasi *naive bayes* memiliki tingkat akurasi yang lebih baik daripada klasifikasi ID3, yaitu dengan tingkat keakuratan *NBC* adalah 82 %, dan *ID3* adalah 76% (Madyatmadja & Aryuni, 2014). Penelitian lainnya yang berjudul “Sistem Pendukung Keputusan Penentuan Kelayakan Calon Tenaga Kerja Indonesia Menggunakan Metode *Naive Bayes*”, memberikan hasil bahwa metode *Naive Bayes* dapat digunakan untuk menentukan kelayakan calon tenaga kerja Indonesia dengan tingkat akurasi 73.89 % (Wasiati & Wijayanti, 2014).

## II. KAJIAN PUSTAKA

### A. Data Mining

Data mining adalah proses penggalian data dari tumpukan *database* yang berukuran besar yang digunakan untuk menemukan *knowledge* berupa informasi penting dan bermanfaat (Wahyuni, S, & Perangin-Angin, 2017). Definisi lain mengatakan *Data mining* adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar (Saleh, 2015).

Tahapan dari proses *Knowledge Discovery in Database(KDD)* adalah (Bustami, 2014):

1. *Selection*
2. *Pre-Processing / Cleaning*
3. *Transformation*
4. *Data Mining*
5. *Interpretation / Evaluation.*

Secara umum kegunaan *data mining* dapat dibagi menjadi dua, yaitu deskriptif dan prediktif. Deskriptif berarti *data mining* digunakan untuk mencari pola-pola yang dapat dipahami manusia yang menjelaskan karakteristik data. Sedangkan prediktif berarti *data mining* digunakan untuk membentuk sebuah model pengetahuan yang akan digunakan untuk melakukan prediksi (Suyanto, 2017).

Beberapa teknik dan sifat *data mining* adalah(Hermawati, 2013):

1. *Classification [Predictive]*
2. *Clustering [Descriptive]*
3. *Association Rule Discovery [Descriptive]*
4. *Sequential Pattern Discovery [Descriptive]*
5. *Regression [Predictive]*
6. *Deviation Detection [Predictive]*

### B. Klasifikasi

Klasifikasi adalah menentukan sebuah *record* data baru ke salah satu dari beberapa kategori (atau klas) yang telah didefinisikan sebelumnya (Hermawati, 2013). Klasifikasi merupakan peran dalam *data mining* yang menggunakan metode pendekatan prediktif (Muflikhah, Ratnawati, & Putri, 2018).

Secara umum proses klasifikasi dapat dilakukan dalam dua tahap, yaitu proses belajar dari data pelatihan dan klasifikasi kasus. Pada proses belajar, algoritma klasifikasi mengolah *data training* untuk menghasilkan sebuah model. Setelah model diuji dan dapat diterima, pada tahap klasifikasi, model tersebut digunakan untuk memprediksi kelas dari kasus baru untuk membantu proses pengambilan keputusan (Wahyuni et al., 2017).

Teknik pembagian data latih dan data uji yang umum digunakan untuk menguji algoritma klasifikasi adalah *cross validation*. Salah satu teknik *cross validation* adalah *k-fold cross validation* (Faisal, 2017), dimana k adalah bilangan bulat yang

digunakan untuk membagi data. Jika nilai k = 5 maka data akan dibagi 5, dan proses pelatihan dan pengujian dilakukan sebanyak lima kali. Pada pelatihan dan pengujian pertama, data bagian pertama akan digunakan sebagai data uji, dan sisanya menjadi data latih, sedangkan pada pelatihan dan pengujian kedua, data bagian kedua akan digunakan sebagai data uji, dan 4 bagian data lainnya sebagai data latih dan seterusnya.

### C. Algoritma Naïve Bayes

*Naïve Bayes* merupakan metode yang membagi permasalahan ke dalam sebuah kelas-kelas berdasarkan ciri-ciri persamaan dan perbedaan dengan menggunakan statistik yang bisa memprediksi probabilitas sebuah kelas. Dikutip dari (Dahri, Agus, & Khairina, 2016) persamaan Teori Bayes adalah:

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

Dimana:

- X : Data dengan class yang belum diketahui
- C : Hipotesis data X merupakan suatu class spesifik
- P(C|X) : Probabilitas hipotesis C berdasar kondisi X (probabilitas posteriori)
- P(C) : Probabilitas hipotesis C (probabilitas prior)
- P(X|C) : Probabilitas X berdasarkan kondisi pada hipotesis C
- (X) : Probabilitas X

Untuk menjelaskan teorema *Naive Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, teorema bayes di atas disesuaikan sebagai berikut:

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

Dimana Variabel C merepresentasikan kelas, sementara variabel X1 ... Xn merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi atau kriteria. Maka 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 juga *likelihood*), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga *evidence*). Karena itu, rumus di atas dapat pula ditulis secara sederhana sebagai berikut:

$$\text{Posterior} = \frac{\text{Prior} \times \text{likelihood}}{\text{Evidence}} \quad 3$$

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 diklasifikasikan. Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan  $(C|X_1 \dots X_n)$  menggunakan aturan perkalian sebagai berikut:

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

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor – faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisis satu persatu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan. Disinilah digunakan asumsi independensi yang sangat tinggi (naif), bahwa masing – masing petunjuk  $(X_1, \dots, X_n)$  saling bebas (independen) satu sama lain. Dengan asumsi tersebut, maka berlaku suatu kesamaan sebagai berikut:

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

Untuk  $i \neq j$ , sehingga:  
 $P(X_i|C, X_j) = P(X_i|C) \quad 5$

Dari persamaan di atas dapat disimpulkan bahwa asumsi independensi naif tersebut membuat syarat peluang menjadi sederhana, sehingga perhitungan menjadi mungkin untuk dilakukan. Selanjutnya penjabaran  $P(C|X_1, \dots, X_n)$  dapat disederhanakan menjadi persamaan berikut:

$$\begin{aligned} P(C|X_1, \dots, X_n) &= P(X_i|C) \\ &= \prod_{i=1}^n P(X_i|C) \quad 6 \end{aligned}$$

Keterangan:

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

Persamaan-persamaan diatas merupakan model dari teorema Naive Bayes yang selanjutnya akan digunakan dalam proses klasifikasi. Untuk klasifikasi dengan data kontinyu atau numerik digunakan rumus Densitas Gauss:

$$P(X_i=x_i|C=c_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(f_i-\mu_i)^2}{2\sigma_{ij}^2}} \quad 7$$

Keterangan:

P : Peluang

Xi : Atribut ke-i

xi : Nilai atribut ke-i

C : Kelas yang dicari

cj : Sub kelas C yang dicari

μ : Mean, menyatakan rata-rata dari seluruh atribut

σ :Deviasi standart, menyatakan varian dari seluruh atribut.

π : 3.14159

e : 2.71828

#### D. Pinjaman

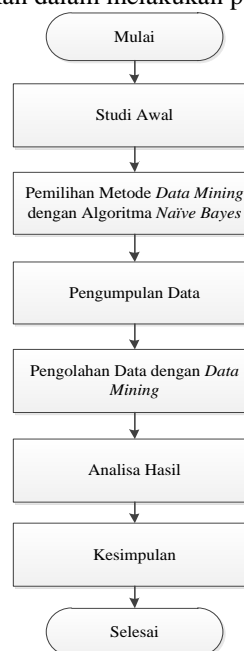
Pengertian pinjaman (kredit) menurut Undang Undang No. 10 Tahun 1998 tentang Perbankan dalam (Arifin, 2018) adalah “penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan atau kesepakatan pinjam meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam untuk melunasi utangnya setelah jangka waktu tertentu dengan pemberian bunga”.

Nilai manfaatnya lebih besar daripada nilai pinjamannya, karena dikelola dengan baik. Dengan kata lain, pinjaman baik harus bisa memberi nilai tambah dan aliran kas masuk ke rekening usaha, sehingga pada akhirnya akan membuat nilai atau jumlah asset semakin bertambah. Sedangkan pinjaman buruk menyebabkan nilai aset dari peminjam menurun atau merosot. Aset tidak bertambah, justru sebaliknya semakin menurun. Bahkan, kewajiban peminjam semakin bertambah. Umumnya, pinjaman buruk lebih banyak ditujukan untuk kebutuhan konsumtif dan bukan kebutuhan produktif (Arifin, 2018).

### III. METODE PENELITIAN

#### 1. Tahapan Penelitian

Berikut ini adalah kerangka pemikiran yang penulis gunakan dalam melakukan penelitian:



Gambar 1. Tahapan Penelitian

A. Studi Awal

Langkah awal dari penelitian ini adalah mencari dan mempelajari masalah yang ada pada Koperasi Anugerah Bintang Cemerlang. Masalah yang teridentifikasi adalah bagaimana menentukan kelayakan yang tepat dalam memberikan pinjaman kepada nasabah. Kemudian, mempelajari beberapa literatur yang berkaitan dengan permasalahan dan bagaimana mencari solusi dari masalah tersebut.

B. Pemilihan Metode *Data mining* dengan Algoritma *Naïve Bayes*

Pada tahap ini, penulis memilih metode *data mining* yang tepat untuk menyelesaikan permasalahan yang ada. Metode yang digunakan untuk menganalisis dan mengukur atau menentukan kelayakan pengaju pinjaman adalah metode klasifikasi dalam *data mining* yaitu *Naïve Bayes*.

C. Pengumpulan Data

Pada tahap ini, penulis melakukan wawancara, observasi, dan dokumentasi di Koperasi Anugerah Bintang Cemerlang. Untuk mengetahui informasi yang dibutuhkan, penulis mengumpulkan data-data historis nasabah yang dapat memberikan informasi mengenai kelayakan pemberian pinjaman.

D. Pengolahan data dengan *Data Mining*

Pada tahap ini, data-data nasabah yang didapatkan dari Koperasi Anugerah Bintang Cemerlang diolah dengan cara mengklasifikasikan data-data tersebut ke dalam kategori layak dan tidak layak. Selanjutnya, melakukan pengolahan data dengan membentuk *data training* dan *data testing*, dianalisis menggunakan metode Klasifikasi *Naïve Bayes* untuk menentukan kelayakan pengaju pinjaman.

E. Analisis Hasil

Pada tahap ini, penulis melakukan analisis dari hasil pengolahan data dengan menggunakan aplikasi *data mining* yaitu *RapidMiner*. Aplikasi tersebut digunakan untuk menganalisis dan mengukur atau menentukan kelayakan pemberian pinjaman kepada nasabah. Kemudian, mengukur tingkat akurasi dari hasil pengolahan data tersebut untuk menentukan layak atau tidaknya pengajuan pinjaman dari nasabah.

F. Kesimpulan

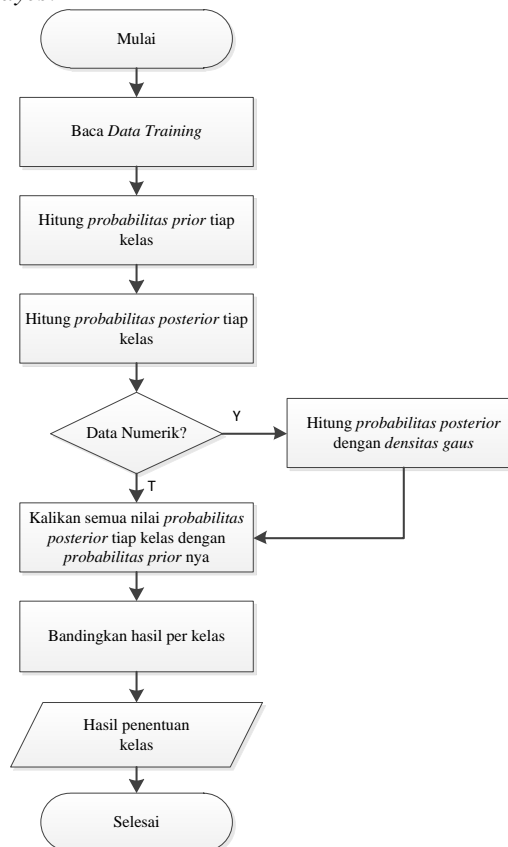
Pada pola perhitungan dengan menggunakan metode klasifikasi *Naïve Bayes* akan menunjukkan hasil akhir berupa kriteria-kriteria nasabah mana saja yang layak diberikan pinjaman pada proses pengajuan pinjaman oleh nasabah di masa yang akan datang berdasarkan data historis nasabah di masa lalu.

2. Metode Analisis Data

Metode analisis data menggunakan *Naïve Bayes Classifier (NBC)* yang merupakan sebuah pengklasifikasian probabilitas sederhana yang

mengaplikasikan Teorema Bayes dengan asumsi ketidaktergantungan yang tinggi. Keuntungan menggunakan NBC adalah metode ini hanya membutuhkan jumlah data pelatihan (*data training*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian.

Berikut ini adalah alur kerja dari metode *Naïve Bayes*:



Gambar 2. Alur Metode Algoritma *Naïve Bayes*

IV. HASIL DAN PEMBAHASAN

Kriteria Nasabah

Berdasarkan wawancara dengan bagian Analisis Kredit diperoleh kriteria-kriteria yang mempengaruhi keputusan manajemen dalam menerima atau menolak suatu pinjaman sebagai berikut:

1. Usia:
  - 17 s.d 22 tahun, 23 s.d 28 tahun, 29 s.d 34 tahun, 35 s.d 40 tahun, dan > 40 tahun
2. Jenis kelamin:
  - Laki-laki dan Perempuan
3. Status Perkawinan:
  - Menikah dan Belum Menikah
3. Profesi:
  - Kontraktor, PNS, Swasta Karyawan, dan Wiraswasta
5. Penghasilan:
  - 2.000.000 – 2.749.999, 2.750.000 – 3.499.999, 3.500.000 – 4.249.999, 4.250.000 – 4.999.999, dan ≥ 5.000.000

6. Status Pinjaman:

*New Order* atau orang yang baru pertama kali meminjam dan *Repeat Order* atau orang yang sudah pernah meminjam kemudian meminjam lagi.

G. Nilai Pinjam:

1.000.000 – 2.499.999, 2.500.000 – 3.999.999, 4.000.000 – 5.499.999, 5.500.000 – 6.999.999, dan  $\geq 7.000.000$

H. Tenor (Lama Mengangsur):

3 s.d 4 bulan, 5 s.d 6 bulan, 7 s.d 8 bulan, 9 s.d 10 bulan, dan 11 s.d 12 bulan.

**Pembagian Dataset**

Teknik yang digunakan dalam membagi *dataset* adalah *k-fold cross validation*, dengan nilai  $k=4$  yang berarti proses pelatihan dan pengujian dilakukan sebanyak 4 kali terhadap 584 *record* data yang digunakan. Berikut ini adalah ilustrasinya:

1-146	147-292	293-438	439-584
-------	---------	---------	---------

**Gambar 3. Teknik pembagian dataset**

Pada proses pertama, data ke-1 s.d data ke-146 digunakan sebagai *data testing*, dan sisanya adalah *data training*. Pada proses kedua, data ke-147 s.d data ke-292 digunakan sebagai *data testing*, dan sisanya adalah *data training*. Berlaku seterusnya sampai dengan proses keempat.

**Pelatihan dan Pengujian Pertama (1-Fold Cross Validation)**

1. Hasil Perhitungan Probabilitas *Prior* tiap Kelas

**Tabel 1. Probabilitas Prior pada 1-Fold Cross Validation**

No.	Kelas	Jumlah	Total Data	Probabilitas
1	Layak	331	438	0.7557
2	Tidak Layak	107	438	0.2443

Dalam penelitian ini, ditentukan dua kelas yaitu kelas layak dan tidak layak. Tabel diatas menjelaskan bahwa terdapat 331 *record* data layak dari 438 data yang digunakan sebagai data latih sehingga nilai probabilitasnya adalah 0.7557, dan 107 *record* data tidak layak dari 438 data sehingga nilai probabilitasnya adalah 0.2443.

2. Hasil Perhitungan Probabilitas *Posterior* tiap Atribut per Kelas

Terdapat 8 atribut yang digunakan dalam penelitian ini, berikut adalah hasil perhitungan probabilitas *posterior* masing-masing atribut:

**Tabel 2. Probabilitas Posterior Atribut Usia pada 1-Fold Cross Validation**

No.	Usia	Probabilitas	
		Layak	Tidak Layak
1	17-22	0.0876	0.0748
2	23-28	0.1390	0.1869
3	29-34	0.2356	0.1682
4	35-40	0.1873	0.2430
5	>40	0.3505	0.3271

Tabel di atas menjelaskan bahwa, pada kriteria usia 17-22 yang dinyatakan layak terdapat 29 data dari 331 data, sehingga probabilitas layaknya diperoleh 0.0876, sedangkan pada usia tersebut yang dinyatakan tidak layak terdapat 8 data dari 107 data, sehingga probabilitas tidak layaknya adalah 0.0748. Cara perhitungannya adalah dengan membagi total data kriteria usia 17-22 yang layak dengan total data layak untuk mencari probabilitas layaknya, dan untuk mencari probabilitas tidak layaknya adalah membagi total data kriteria usia 17-22 yang tidak layak dengan total data tidak layak. Berlaku untuk menghitung atribut usia lainnya sehingga didapatkan nilai probabilitas atribut usia dengan beberapa kelas di dalamnya seperti pada tabel di atas.

Perhitungan yang sama juga berlaku untuk 7 atribut lainnya yaitu jenis kelamin, status perkawinan, profesi, penghasilan, status pinjaman, nilai pinjam dan tenor, yang hasilnya ditampilkan pada tabel-tabel di bawah ini.

**Tabel 3. Probabilitas Posterior Atribut Jenis Kelamin pada 1-Fold Cross Validation**

No.	Jenis Kelamin	Probabilitas	
		Layak	Tidak Layak
1	L	0.6586	0.6636
2	P	0.3414	0.3364

**Tabel 4. Probabilitas Posterior Atribut Status Perkawinan pada 1-Fold Cross Validation**

No.	Status Perkawinan	Probabilitas	
		Layak	Tidak Layak
1	Menikah	0.7734	0.6822
2	Belum Menikah	0.2266	0.3178



**Tabel 5. Probabilitas *Posterior* Atribut Profesi pada 1-Fold Cross Validation**

No.	Profesi	Probabilitas	
		Layak	Tidak Layak
1	Wiraswasta	0.5227	0.4393
2	Swasta Karyawan	0.4622	0.5421
3	PNS	0.0030	0.0093
4	Kontraktor	0.0121	0.0093

**Tabel 6. Probabilitas *Posterior* Atribut Penghasilan pada 1-Fold Cross Validation**

No.	Penghasilan	Probabilitas	
		Layak	Tidak Layak
1	2.000.000-2.749.999	0.1450	0.0841
2	2.750.000-3.499.999	0.2779	0.2617
3	3.500.000-4.249.999	0.3172	0.3738
4	4.250.000-4.999.999	0.0544	0.0935
5	>= 5.000.000	0.2054	0.1869

**Tabel 7. Probabilitas *Posterior* Atribut Status Pinjaman pada 1-Fold Cross Validation**

No.	Status Pinjaman	Probabilitas	
		Layak	Tidak Layak
1	Repeat Order	0.3565	0.1869
2	New Order	0.6435	0.8131

**Tabel 8. Probabilitas *Posterior* Atribut Nilai Pinjam pada 1-Fold Cross Validation**

No.	Nilai Pinjaman	Probabilitas	
		Layak	Tidak Layak
1	1.000.000-2.499.999	0.0695	0.0935
2	2.500.000-3.999.999	0.3384	0.2056
3	4.000.000-5.499.999	0.3474	0.4393
4	5.500.000-6.999.999	0.1994	0.2056
5	>= 7.000.000	0.0453	0.0561

**Tabel 9. Probabilitas *Posterior* Atribut Tenor pada 1-Fold Cross Validation**

No.	Nilai Pinjaman	Probabilitas	
		Layak	Tidak Layak
1	3-4	0.0302	0.0187
2	5-6	0.0302	0.0187
3	7-8	0.2236	0.1495
4	9-10	0.1299	0.1028
5	11-12	0.5861	0.7103

3. Pengujian terhadap *Data Testing*

Semua nilai probabilitas *posterior* setiap atribut dikalikan, kemudian dikalikan dengan probabilitas *prior* kelasnya. Setelah itu, ditentukan kelasnya.

4. Hasil Perhitungan Akurasi

**Tabel 10. Confusion Matrix pada 1-Fold Cross Validation**

No.	Nama Pengukuran	Rumus	Hasil
1	Accuracy	$\frac{(TP + TN)}{n}$	62.33%
2	Error Rate	$\frac{(FP + FN)}{n}$	37.67%
3	Sensitivity/Recall	$\frac{TP}{TP + FN}$	93.62%
4	False Positive Rate	$\frac{FP}{TN + FP}$	94.23%
5	Spesificity	$\frac{TN}{TN + FP}$	5.77%
6	Precision	$\frac{TP}{FP + TP}$	64.23%
7	Prevalence	$\frac{Actual\ Positive}{n}$	64.38%

Tabel *confusion matrix* digunakan untuk mengukur kinerja algoritma klasifikasi. Pada tabel di atas berdasarkan perhitungan manual diperoleh nilai akurasi sebesar 62.33 %. Nilai tersebut nantinya akan diuji menggunakan aplikasi *RapidMiner*, untuk menunjukkan apakah perhitungan manual akurasi *naïve bayes* menunjukkan nilai yang sama jika diinput ke dalam aplikasi *RapidMiner* tersebut.

Berikut ini adalah hasil implementasi ke dalam *RapidMiner*, dimana pada gambar dibawah menunjukkan bahwa nilai akurasi yang dihasilkan oleh aplikasi sama dengan perhitungan manual.

accuracy: 62.33%

	true LAYAK	true TIDAK LAYAK
pred. LAYAK	88	49
pred. TIDAK LAYAK	6	3
class recall	93.62%	5.77%

**Gambar 4. Hasil Akurasi 1-Fold Cross Validation pada RapidMiner**

**Pelatihan dan Pengujian Kedua (2-Fold Cross Validation)**

Pelatihan dan pengujian kedua langkah-langkahnya sama dengan proses pelatihan dan pengujian pertama. Yang membedakan hanya pada penggunaan datanya. Pada proses kedua, yang dijadikan *data testing* adalah data ke-147 s.d data ke-292.

Berikut ini adalah hasil perhitungan akurasinya:

**Tabel 11. Confusion Matrix pada 2-Fold Cross Validation**

No.	Nama Pengukuran	Rumus	Hasil
1	Accuracy	$\frac{(TP + TN)}{n}$	66.44%
2	Error Rate	$\frac{(FP + FN)}{n}$	33.56%
3	Sensitivity/Recall	$\frac{TP}{TP + FN}$	100.00%
4	False Positive Rate	$\frac{FP}{TN + FP}$	100.00%
5	Spesificity	$\frac{TN}{TN + FP}$	0.00%
6	Precision	$\frac{TP}{FP + TP}$	66.44%
7	Prevalence	$\frac{Actual\ Positive}{n}$	66.44%

Berikut ini adalah hasil implementasi ke dalam *RapidMiner*, dimana pada gambar dibawah menunjukkan bahwa nilai akurasi yang dihasilkan oleh aplikasi sama dengan perhitungan manual.

accuracy: 66.44%

	true TIDAK LAYAK	true LAYAK
pred. TIDAK LAYAK	0	0
pred. LAYAK	49	97
class recall	0.00%	100.00%

**Gambar 5. Hasil Akurasi 2-Fold Cross Validation pada RapidMiner**

**Pelatihan dan Pengujian Ketiga (3-Fold Cross Validation)**

Pada proses ketiga, yang dijadikan *data testing* adalah data ke-293 s.d data ke-438.

Berikut ini adalah hasil perhitungan akurasinya:

**Tabel 12. Confusion Matrix pada 3-Fold Cross Validation**

No.	Nama Pengukuran	Rumus	Hasil
1	Accuracy	$\frac{(TP + TN)}{n}$	69.18%
2	Error Rate	$\frac{(FP + FN)}{n}$	30.82%
3	Sensitivity/Recall	$\frac{TP}{TP + FN}$	92.52%
4	False Positive Rate	$\frac{FP}{TN + FP}$	94.87%
5	Spesificity	$\frac{TN}{TN + FP}$	5.13%
6	Precision	$\frac{TP}{FP + TP}$	72.79%
7	Prevalence	$\frac{Actual\ Positive}{n}$	73.29%

Berikut ini adalah hasil implementasi ke dalam *RapidMiner*, dimana pada gambar dibawah menunjukkan bahwa nilai akurasi yang dihasilkan oleh aplikasi sama dengan perhitungan manual.

accuracy: 69.18%

	true TIDAK LAYAK	true LAYAK
pred. TIDAK LAYAK	2	8
pred. LAYAK	37	99
class recall	5.13%	92.52%

**Gambar 6. Hasil Akurasi 3-Fold Cross Validation pada RapidMiner**

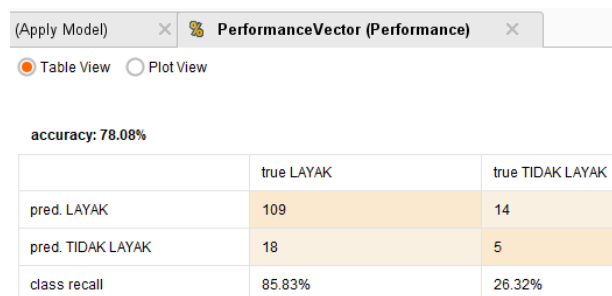
**Pelatihan dan Pengujian Keempat (4-Fold Cross Validation)**

Pada proses keempat, yang dijadikan *data testing* adalah data ke-439 s.d data ke-584. Berikut ini adalah hasil perhitungan akurasi:

**Tabel 13. Confusion Matrix pada 4-Fold Cross Validation**

No.	Nama Pengukuran	Rumus	Hasil
1	Accuracy	$\frac{TP + TN}{n}$	78.08%
2	Error Rate	$\frac{FP + FN}{n}$	21.92%
3	Sensitivity/Recall	$\frac{TP}{TP + FN}$	85.83%
4	False Positive Rate	$\frac{FP}{TN + FP}$	73.68%
5	Spesificity	$\frac{TN}{TN + FP}$	26.32%
6	Precision	$\frac{TP}{FP + TP}$	88.62%
7	Prevalence	$\frac{Actual\ Positive}{n}$	86.99%

Berikut ini adalah hasil implementasi ke dalam *RapidMiner*, dimana pada gambar di bawah menunjukkan bahwa nilai akurasi yang dihasilkan oleh aplikasi sama dengan perhitungan manual.



**Gambar 7. Hasil Akurasi 4-Fold Cross Validation pada RapidMiner**

**V. KESIMPULAN**

Berdasarkan hasil penelitian yang telah dilakukan pada Koperasi Anugerah Bintang Cemerlang mengenai penggunaan klasifikasi *Naive Bayes* untuk memprediksi kelayakan pemberian pinjaman, maka dapat disimpulkan sebagai berikut:

1. Prediksi kelayakan pemberian pinjaman dapat dilakukan dengan mempertimbangkan dataset nasabah di masa lalu menggunakan algoritma klasifikasi yaitu *naive bayes*.
2. Algoritma klasifikasi *Naive Bayes* dapat diterapkan untuk membantu Analisis Kredit dalam memprediksi kelayakan pemberian pinjaman terhadap nasabah, dimana dalam penelitian ini akurasi data yang diperoleh

dengan 4 kali proses pengujian menggunakan teknik *k-fold cross validation* mendapatkan nilai sebesar 62.33 %, 66.44 %, 69.18 %, dan pengujian terakhir mendapatkan tingkat akurasi tertinggi yaitu 78.08% dengan presentase eror sebesar 21.92 %.

**DAFTAR PUSTAKA**

Arifin, T. (2018). *Berani Jadi Pengusaha Sukses Usaha dan Raih Pinjaman* (I). Jakarta: PT. Gramedia Pustaka Utama.

Bustami. (2014). Penerapan Algoritma Naive Bayes Untuk Nasabah Asuransi. *Jurnal Informatika*, 8(1), 884–898.

Dahri, D., Agus, F., & Khairina, D. M. (2016). Metode Naive Bayes Untuk Penentuan Penerima Beasiswa Bidikmisi Universitas Mulawarman. *Jurnal Informatika Mulawarman*, 11(2), 29–35.

Elyana, I. (2017). Decision Support System untuk Kelayakan Pemberian Kredit Motor dengan Menggunakan Metode Simple Additive Weighting pada Perusahaan Leasing. *Jurnal Pilar Nusa Mandiri*, 13(1), 85–91. Retrieved from <http://ejournal.nusamandiri.ac.id/ejournal/index.php/pilar/article/view/341/273>

Faisal, M. R. (2017). *Seri Belajar Data Science Klasifikasi dengan Bahasa Pemrograman R*. Banjarmasin: Indonesia Net Developer Community. Retrieved from <https://play.google.com/books/reader?id=svXUDQAAQBAJ&hl=id&printsec=frontcover&pg=GBS.PA66>

Hermawati, F. A. (2013). *Data Mining*. (P. Christian, Ed.) (1st ed.). Yogyakarta: CV ANDI OFFSET.

Madyatmadja, E. D., & Aryuni, M. (2014). Comparative Study Of Data Mining Model For Credit Card Application Scoring In Bank. *Journal of Theoretical and Applied Information Technology*, 59(2), 269–274. Retrieved from <http://pdfs.semanticscholar.org>

Muflikhah, L., Ratnawati, D. E., & Putri, R. R. Ma. (2018). *Data Mining* (1st ed.). Malang: UB Press.

Pandie, E. S. Y. (2018). Implementasi Algoritma Data Mining Naive Bayes Pada Koperasi. *J-ICON*, 6(1), 15–20.

Saleh, A. (2015). Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Citec Journal*, 2(3), 207–217.

Suyanto. (2017). *Data Mining Untuk Klasifikasi Dan Klasterisasi Data* (1st ed.). Bandung: INFORMATIKA.

Wahyuni, S., S, K. S., & Perangin-Angin, M. I. (2017). Implementasi Rapidminer Dalam



Menganalisis Data Mahasiswa Drop Out,  
*10(2)*, 1899–1902.  
<https://doi.org/10.1002/pmic.201000579>

Wasiati, H., & Wijayanti, D. (2014). Sistem Pendukung Keputusan Penentuan Kelayakan Calon Tenaga Kerja Indonesia Menggunakan Metode Naive Bayes. *IJINS - Indonesian Journal on Networking and Security*, *3(2)*, 45–51.