

Penerapan Data Mining Dalam Klasifikasi Data Anggota Kopdit Sejahtera Bandar Lampung Dengan Algoritma *Naïve Bayes*

Rohmat Indra Borman¹, Mina Wati²

Fakultas Teknik dan Ilmu Komputer (FTIK), Universitas Teknokrat Indonesia^{1,2}

Jl. Zainal Abidin Pagar Alam No. 9-11, Bandar Lampung

¹rohmat_indra@teknokrat.ac.id, ²minawati04@gmail.com

Abstrak

Data mining adalah teknik yang memanfaatkan data dalam jumlah yang besar untuk memperoleh informasi berharga yang sebelumnya tidak diketahui dan dapat dimanfaatkan untuk pengambilan keputusan penting. Pada penelitian ini, dilakukan data mining pada anggota sebuah perusahaan Koperasi Kredit Sejahtera untuk mengetahui lancar atau tidak lancarnya anggota tersebut. Data yang ada dianalisis menggunakan algoritma *Naïve Bayes*. *Naïve Bayes* merupakan salah satu metode pada *probabilistic reasoning*. Algoritma *Naïve Bayes* bertujuan untuk melakukan klasifikasi data pada kelas tertentu, kemudian tingkat akurasi dapat digunakan untuk memperkirakan nasabah yang bergabung, sehingga perusahaan bisa mengambil keputusan yang terbaik untuk para anggota. Hasil dari penelitian ini adalah Algoritma *Naïve Bayes* dapat digunakan sebagai salah satu pilihan untuk mengklasifikasikan data anggota kelalaian pinjaman Koperasi Kredit Sejahtera. Penelitian ini menggunakan 1064 *record* data training dan 300 *record* data testing dengan 6 fitur didalamnya untuk mengetahui hasil klasifikasi menggunakan WEKA dan Rapidminer. Dari pengujian menggunakan WEKA dan Rapidminer yang menggunakan 1064 *record* data *training* serta 300 *record* data *testing* menghasilkan tingkat akurasi sebesar 70,33%, recall 70,33% dan presisi 100% atau 1. Jadi semakin banyak data testing yang digunakan maka akan semakin mempengaruhi nilai akurasi.

Kata Kunci: *Klasifikasi, Naïve Bayes, Weka, Rapidminer*

1. Pendahuluan

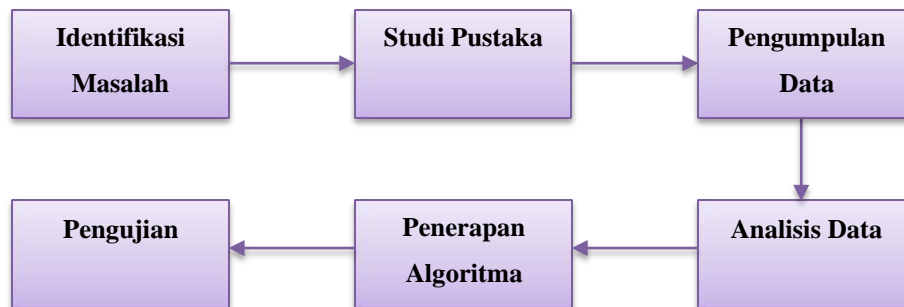
Koperasi merupakan usaha kekeluargaan dengan tujuan untuk mensejahterakan anggotanya (UUD 1945 pasal 33 ayat 1). Menurut Pasal 1 ayat 11 Undang-Undang Nomor 10 Tahun 1998, Kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan, berdasarkan persetujuan atau kesepakatan pinjam meminjam antara bank atau instansi keuangan dengan pihak lain yang mewajibkan pihak peminjam untuk melunasi utangnya setelah jangka waktu tertentu dengan jumlah bunga. Koperasi memiliki kebijakan yang berbeda-beda dalam pemberian kredit. Tetapi pada umumnya pemberian kredit dipengaruhi beberapa faktor seperti kepercayaan, kesepakatan, jangka waktu, risiko dan balas jasa [1]. Analisis kredit perlu mengidentifikasi dan menilai faktor-faktor yang dapat mempengaruhi nasabah dalam pengembalian kredit [2]. Koperasi Kredit Sejahtera merupakan salah satu Badan Usaha milik P3RI (Persatuan Purna Karyawan Perkebunan Republik Indonesia). Koperasi Kredit Sejahtera juga merupakan KOPDIT Primer yang berorientasi pada kepentingan anggota dengan meningkatkan pelaksanaan 5 (lima) wajib yaitu mengikuti program simpan pinjam, pendidikan, Daperma (Dana Perlindungan Bersama), audit dan membayar iuran solidaritas. KOPDIT Sejahtera ini berada dibawah naungan PUSKOPDIT Caraka Utama. Dimana PUSKOPDIT Caraka Utama merupakan koperasi kredit sekunder di Lampung. Lembaga koperasi memiliki banyak bidang usaha yaitu salah satunya kredit simpan pinjam. Koperasi simpan pinjam mendapatkan modal dari hasil pemupukan simpanan dari anggotanya, dan nantinya akan dipinjamkan kembali kepada anggota yang membutuhkan modal [3].

Berdasarkan data Koperasi Kredit Sejahtera saat ini, dana kelalaian pinjaman mengalami peningkatan. Dana kelalaian pinjaman tahun 2016 meningkat sebesar 55,73% dibandingkan tahun sebelumnya (Lampiran Posisi Keuangan KOPDIT SEJAHTERA, 2017). Meningkatnya dana kelalaian pinjaman mempengaruhi kelancaran pinjaman di KOPDIT Sejahtera. Untuk meminimalisir terjadinya peningkatan kelalaian pinjaman maka diperlukan sebuah metode yang untuk mengklasifikasikan kelalaian pinjaman anggota KOPDIT Sejahtera. Sehingga nanti ada dua klasifikasi yaitu lancar dan tidak lancar, dari hasil klasifikasi tersebut pengambil keputusan dapat memberikan keputusan – keputusan yang terbaik dalam permasalahan kelalaian pinjaman anggota KOPDIT Sejahtera. Atribut yang digunakan antara lain adalah nama, jenis kelamin, umur, pekerjaan, gaji, jumlah pinjaman serta jangka waktu. Salah satu metode yang dapat diterapkan dalam permasalahan ini adalah *Naïve Bayes*. *Naïve Bayes* merupakan salah satu metode yang dapat digunakan untuk mengklasifikasikan data. *Bayesian classification* merupakan pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class* [4]. *Naïve Bayes* merupakan suatu kelas keputusan, dengan menggunakan perhitungan probabilitas matematika dengan syarat bahwa nilai keputusan adalah benar, berdasarkan informasi obyek [5]. Hasil klasifikasi yang dilakukan nantinya dapat membantu para staff di KOPDIT Sejahtera untuk meminimalisir tingkat kelalaian pinjaman dan yang

mengambil keputusan dapat memberikan keputusan yang terbaik sejak dini dengan mengetahui hasil dari klasifikasi lancar atau tidak lancarnya serta tingkat akurasi untuk mencapai hasil yang optimal.

2. Metodologi Penelitian

Penelitian merupakan kegiatan penelitian yang dilakukan secara terencana, teratur dan sistematis untuk mencapai tujuan tertentu. Tahapan penelitian ini juga merupakan pengembangan dari kerangka penelitian, dan terbagi lagi menjadi beberapa sub menu, tahapan penelitian dapat dilihat pada 1.



Gambar 1. Tahapan Penelitian

Identifikasi Masalah

Berdasarkan data Koperasi Kredit Sejahtera pada tahun 2016 dana kelalaian pinjaman meningkat sebesar 55,73%, dibandingkan tahun sebelumnya. Penulis mengidentifikasi masalah peningkatan dana kelalaian pinjaman di Koperasi Kredit Sejahtera untuk meminimalisir terjadinya peningkatan kelalaian pinjaman, dimana penulis mengklasifikasikan kelalaian pinjaman yaitu lancar dan tidak lancar.

Studi Pustaka

Untuk menunjang hasil penelitian, kegiatan pengumpulan data dilakukan melalui aktivitas tinjauan pustaka [6]. Pada tahapan ini dilakukan pengumpulan berbagai informasi terkait referensi dari jurnal-jurnal, skripsi, buku dan data-data koperasi kredit yang berhubungan dengan penelitian yang penulis lakukan. Studi pustaka dilakukan untuk menunjang penelitian.

Pengumpulan Data

Pada tahapan pengumpulan data ini, penulis menggunakan data anggota pinjaman dari Koperasi Kredit Sejahtera. Yang berada di Jalan Teuku Umar Nomor 300 Kedaton Kota Bandar Lampung. Data yang diambil yaitu data anggota pinjaman pada tahun 2015, 2016 dan 2017.

Analisis Data

Pada tahapan ini dilakukan analisis data dengan :

- 1) Menghilangkan noise (data yang tidak konsisten atau data tidak relevan).
Menghilangkan data *noise* (data yang tidak relevan / berhubungan langsung dengan tujuan akhir proses data mining). Antara lain membuang redundansi data , memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak.
- 2) Pengelompokkan data.
Dalam Pemetaan atau pengelompokkan Klasifikasi Kelalaian Pinjaman terdapat beberapa fitur yang menjadi variabel dalam perhitungan metode Klasifikasi *Naive Bayes* yaitu:
 - a. Nama
Merupakan variabel identitas nama anggota.
 - b. Jenis Kelamin
Merupakan variabel jenis kelamin anggota yang dikelompokkan dalam dua kategori yaitu laki – laki dan perempuan.
 - c. Umur
Merupakan variabel umur anggota yang di kelompokkan dalam lima kategori yaitu 20-29 Tahun, 30-39 Tahun, 40-49 Tahun, 50- 59 Tahun dan diatas 60 Tahun.
 - d. Pekerjaan
Merupakan variabel pekerjaan anggota yang di kelompokkan dalam tiga kategori yaitu Karyawan, Wiraswasta dan Pensiun.
 - e. Gaji
Merupakan variabel gaji dari anggota yang di kelompokkan dalam lima kategori yaitu 0-1 Juta, 2-3 Juta, 4-5 Juta, 6-7 Juta dan diatas 7 Juta.

- f. Jumlah Pinjaman
Merupakan variabel jumlah pinjaman anggota yang di dikelompokkan dalam lima kategori yaitu 0-5 Juta, 6–10 Juta, 11-15 Juta, 16-20 Juta dan diatas 20 Juta.
- g. Jangka Waktu
Merupakan variabel jangka waktu anggota yang di kelompokkan dalam lima kategori yaitu 1-10 Bulan, 11-20 Bulan, 21-30 Bulan, 31-40 Bulan dan diatas 40 Bulan.

Penerapan Algoritma

Untuk penerapan algoritma Naive Bayes data yang akan digunakan adalah data nama, jenis kelamin, umur, pekerjaan, gaji, jumlah pinjaman, dan jangka waktu.

Tabel 1. Data Pinjaman Anggota KOPDIT Sejahtera

| No | Nama | Jenis Kelamin | Umur | Pekerjaan | Gaji | Jumlah Pinjaman | Jangka Waktu | Klasifikasi |
|----|-----------------------|---------------|-------------|------------|----------|-----------------|--------------|--------------|
| 1 | Ata | Laki - Laki | 50-59 Tahun | Karyawan | 2-3 Juta | 16-20 Juta | 21-30 Bulan | Tidak Lancar |
| 2 | Dewi Purbayani | Perempuan | 30-39 Tahun | Karyawan | 0-1 Juta | 0-5 Juta | 21-30 Bulan | Tidak Lancar |
| 3 | Gusti Nilai | Perempuan | 40-49 Tahun | Karyawan | 2-3 Juta | >20 Juta | 21-30 Bulan | Tidak Lancar |
| 4 | Nari Aryani | Perempuan | 40-49 Tahun | Karyawan | 0-1 Juta | 16-20 Juta | 21-30 Bulan | Tidak Lancar |
| 5 | ST Talkasim | Laki - Laki | 30-39 Tahun | Karyawan | 6-7 Juta | 0-5 Juta | 1-10 Bulan | Lancar |
| 6 | Sugiono | Laki - Laki | 50-59 Tahun | Karyawan | 0-1 Juta | 6-10 Juta | 21-30 Bulan | Lancar |
| 7 | Sumardi | Laki - Laki | >60 Tahun | Wiraswasta | 0-1 Juta | 6-10 Juta | 21-30 Bulan | Lancar |
| 8 | Teguh S | Laki - Laki | 50-59 Tahun | Karyawan | 0-1 Juta | 6-10 Juta | 21-30 Bulan | Tidak Lancar |
| 9 | Tiswanto | Laki - Laki | >60 Tahun | Pensiun | 0-1 Juta | 0-5 Juta | 11-20 Bulan | Lancar |
| 10 | Totok Widiharto | Laki - Laki | 40-49 Tahun | Wiraswasta | 2-3 Juta | 0-5 Juta | 21-30 Bulan | Tidak Lancar |
| 11 | Bagus Satria utama | Laki - Laki | 20-29 Tahun | Wiraswasta | 0-1 Juta | 11-15 Juta | 21-30 Bulan | Tidak Lancar |
| 12 | Deden Hermawan | Laki - Laki | 50-59 Tahun | Pensiun | 0-1 Juta | 16-20 Juta | 31-40 Bulan | Lancar |
| 13 | Dra. Yulismalinda | Perempuan | 50-59 Tahun | Karyawan | 4-5 Juta | 6-10 Juta | 31-40 Bulan | Lancar |

| No | Nama | Jenis Kelamin | Umur | Pekerjaan | Gaji | Jumlah Pinjaman | Jangka Waktu | Klasifikasi |
|----|---------------------|---------------|-------------|------------|----------|-----------------|--------------|--------------|
| 14 | Joni Arga Hutabarat | Laki - Laki | 50-59 Tahun | Pensiun | 0-1 Juta | 16-20 Juta | >40 Bulan | Lancar |
| 15 | Kosasi | Laki - Laki | 30-39 Tahun | Karyawan | 2-3 Juta | 16-20 Juta | 31-40 Bulan | Lancar |
| 16 | Kustini | Perempuan | 50-59 Tahun | Karyawan | 2-3 Juta | 6-10 Juta | 31-40 Bulan | Tidak Lancar |
| 17 | M. Rosik | Laki - Laki | >60 Tahun | Pensiun | 0-1 Juta | 0-5 Juta | 31-40 Bulan | Lancar |
| 18 | Martinah Supo | Perempuan | 40-49 Tahun | karyawan | 0-1 Juta | 6-10 Juta | >40 Bulan | Tidak Lancar |
| 19 | Ny. Rohmi | Perempuan | 50-59 Tahun | Pensiun | 0-1 Juta | 0-5 Juta | >40 Bulan | Lancar |
| 20 | Paino II | Laki - Laki | 50-59 Tahun | Wiraswasta | 0-1 Juta | 0-5 Juta | >40 Bulan | Lancar |
| 21 | Sondang Siahaan | Perempuan | 30-39 Tahun | Karyawan | 0-1 Juta | 0-5 Juta | 1-10 Bulan | Lancar |
| 22 | Sukirno | Laki - Laki | 40-49 Tahun | Karyawan | 0-1 Juta | 11-15 Juta | 31-40 Bulan | ?? |

Berdasarkan tabel diatas dapat dihitung klasifikasi data pinjaman anggota Koperasi Kredit Sejahtera apabila atribut berupa nama, jenis kelamin, umur, pekerjaan, gaji, jumlah pinjaman, dan jangka waktu menggunakan algoritma *naive bayes*.

Adapun langkah – langkahnya sebagai berikut:

- 1) Menghitung jumlah class/label

$P(Y = \text{Lancar}) = 12/21 = 0.5714$ "Jumlah data Lancar pada data anggota yang melakukan pinjaman dibagi dengan jumlah keseluruhan data".

$P(Y = \text{Tidak Lancar}) = 9/21 = 0.4286$ "Jumlah data Tidak Lancar pada data anggota yang melakukan pinjaman dibagi dengan jumlah keseluruhan data".

- 2) Menghitung jumlah kasus yang sama dengan class yang sama $P(\text{Jenis Kelamin} = \text{Laki-laki} | Y = \text{Lancar}) = 9/12 = 0.7500$

$P(\text{Jenis Kelamin} = \text{Laki-laki} | Y = \text{Tidak Lancar}) = 4/9 = 0.4444$

$P(\text{Umur} = 40-49 \text{ Tahun} | Y = \text{Lancar}) = 0/12 = 0$

$P(\text{Umur} = 40-49 \text{ Tahun} | Y = \text{Tidak Lancar}) = 4/9 = 0.3333$

$P(\text{Pekerjaan} = \text{Karyawan} | Y = \text{Lancar}) = 5/12 = 0.4167$.

$P(\text{Pekerjaan} = \text{Karyawan} | Y = \text{Tidak Lancar}) = 7/9 = 0.7778$

$P(\text{Gaji} = 0-1 \text{ Juta} | Y = \text{Lancar}) = 9/12 = 0.7500$

$P(\text{Gaji} = 0-1 \text{ Juta} | Y = \text{Tidak Lancar}) = 5/9 = 0.5556$

$P(\text{Jumlah Pinjaman} = 11-15 \text{ juta} | Y = \text{Lancar}) = 0/12 = 0$

$P(\text{Jumlah Pinjaman} = 11-15 \text{ juta} | Y = \text{Tidak Lancar}) = 1/9 = 0.1111$

$P(\text{Jangka Waktu} = 31-40 \text{ Bulan} | Y = \text{Lancar}) = 4/12 = 0.3333$

$P(\text{Jangka Waktu} = 31-40 \text{ Bulan} | Y = \text{Tidak Lancar}) = 1/9 = 0.1111$

- 3) Kalikan semua hasil variabel Lancar dan Tidak Lancar

$P(\text{Lancar}) * P(\text{Laki-laki} | \text{Lancar}) * P(40-49 \text{ Tahun} | \text{Lancar}) * P(\text{Karyawan} | \text{Lancar}) * P(0-1 \text{ Juta} | \text{Lancar}) * P(11-15 \text{ Juta} | \text{Lancar}) * P(31-40 \text{ Bulan} | \text{Lancar})$

$$\begin{aligned}
 &= \frac{12}{21} \times \frac{9}{12} \times \frac{0}{12} \times \frac{5}{12} \times \frac{9}{12} \times \frac{0}{12} \times \frac{4}{12} \\
 &= 0.5714 \times 0.7500 \times 0 \times 0.4167 \times 0.7500 \times 0 \times 0.3333 \\
 &= 0
 \end{aligned}$$

P (Tidak Lancar) * P (Laki-laki | Tidak Lancar) * P (40-49 Tahun | Tidak Lancar) * P (Karyawan | Tidak Lancar) * P (0-1 Juta | Tidak Lancar) * P (11-15 Juta | Tidak Lancar) * P (31-40 Bulan | Tidak Lancar)

4) Bandingkan hasil class Lancar dan Tidak Lancar

$$\begin{aligned}
 &= \frac{9}{21} \times \frac{4}{9} \times \frac{4}{9} \times \frac{7}{9} \times \frac{5}{9} \times \frac{1}{9} \times \frac{1}{9} \\
 &= 0.4286 \times 0.4444 \times 0.3333 \times 0.7778 \times 0.5556 \times 0.1111 \times 0.1111 \\
 &= 0.00034
 \end{aligned}$$

Dari hasil diatas, terlihat bahwa nilai probabilitas tertinggi ada pada kelas (P | Lancar), sehingga dapat disimpulkan bahwa status anggota tersebut masuk dalam klasifikasi Tidak Lancar.

Pengujian

Pengujian dilakukan untuk melihat tingkat kesalahan atau *error* dari suatu sistem [7]. Pada tahapan pengujian ini, penulis menggunakan WEKA dan RAPIDMINER dimana metode yang dipakai adalah Naive Bayes, untuk menguji tingkat klasifikasi akurasi dapat di lihat dari nilai *presisi*, *akurasi* dan *recall*.

Pengujian *precision* digunakan dalam mencari nilai proporsi kasus positif yang benar, rumus persamaan 2 :

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

Pengujian *recall* digunakan dalam mencari nilai proporsi kasus positif yang teridentifikasi benar, rumus persamaan 3 :

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

Pengujian *accuracy* digunakan dalam mencari nilai proporsi jumlah yang benar, rumus persamaan 4 :

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{4}$$

Keterangan :

- TP : True Positive
- TN : True Negative
- FP : False Positive
- FN : False Negative

3. Hasil Dan Pembahasan

Sebagai suatu rangkaian proses, data Mining dapat dibagi menjadi beberapa tahap proses. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantaraan knowledge base [8]. Untuk melakukan proses mining terdapat beberapa tahapan yang dilakukan, diantaranya : Pembersihan data mining (*data cleaning*); Integrasi data (*data intregation*); Seleksi data (*data selection*); Transformasi Data (*data transformation*); Proses Mining; Evaluasi pola (*pattern evaluation*); Presentasi Pengetahuan (*knowledge presentasion*) [9].

Pembersihan Data (Data Cleaning)

Pembersihan data bertujuan untuk menghilangkan data *noise* (data yang tidak relevan / berhubungan langsung dengan tujuan akhir proses data mining). Proses *cleaning* mencakup antara lain membuang duplikasi data , memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak. Proses *cleaning* pada data anggota pinjaman dapat dilihat pada tabel 4.3.

Tabel 2. Proses *cleaning* data anggota pinjaman KOPDIT

| Nama Anggota | Saldo Pinjaman | Pokok | Pokok | Pokok | Pokok | Saldo Pinjaman |
|--------------------------|----------------|---------|---------|---------|---------|----------------|
| Poniat Supadi | 2,739,621 | 153,472 | 157,639 | 161,806 | 165,972 | 2,100,732 |
| Kun Suherto | 1,089,000 | 170,250 | 174,750 | 179,250 | 183,750 | 381,000 |
| Sunip Aliabat | 0 | | | | | 0 |
| Elsye Toding | 0 | | | | | 0 |
| Soeparso | 5,899,996 | 422,917 | 435,417 | 447,917 | 460,417 | 4,133,328 |
| Ny. Martalis a Kamada | 0 | | | | | 0 |
| Romli Subagio | 7,500,000 | 98,958 | 105,208 | 111,458 | 117,708 | 7,066,568 |

| | | | | | | |
|---------------------|-----------|--------|--------|---------|---------|-----------|
| Ny. Sudarwati Akumo | 0 | | | | | 0 |
| Lukmansyah | 2,743,125 | 93,125 | 96,875 | 100,625 | 104,375 | 2,348,125 |
| Nazarudin.S.T | 0 | | | | | 0 |

Integrasi Data (*Data Integration*)

Tahapan integrasi data adalah tahapan dimana data dijadikan dalam satu kesatuan data untuk proses data mining. Hasil integrasi data dapat dilihat pada tabel 3 :

Tabel 3. Integrasi data anggota pinjaman KOPDIT Sejahtera

| Nama | Jenis Kelamin | Umur | Pekerjaan | Gaji | Jumlah Pinjaman | Jangka Waktu | Klasifikasi |
|------------------|---------------|-------------|------------|----------|-----------------|--------------|--------------|
| Suciati suparno | Perempuan | >60 Tahun | Pensiun | 0-1 Juta | 6-10 Juta | >40 Bulan | Tidak Lancar |
| Maedalena | Perempuan | 40-49 Tahun | Wiraswasta | 0-1 Juta | 6-10 Juta | 31-40 Bulan | Lancar |
| Endang Suharti | Perempuan | 40-49 Tahun | Karyawan | 2-3 Juta | >20 Juta | 11-20 Bulan | Lancar |
| Sugiri | Laki - Laki | 50-59 Tahun | Karyawan | 2-3 Juta | 6-10 Juta | 11-20 Bulan | Lancar |
| Usrek | Laki - Laki | 50-59 Tahun | Pensiun | 0-1 Juta | 6-10 Juta | >40 Bulan | Tidak Lancar |
| Rodi Ashari | Laki - Laki | 40-49 Tahun | Karyawan | 2-3 Juta | 16-20 Juta | 31-40 Bulan | Lancar |
| Sarimin | Laki - Laki | 50-59 Tahun | Karyawan | 0-1 Juta | 6-10 Juta | 11-20 Bulan | Lancar |
| M. Arsyad | Laki - Laki | 50-59 Tahun | Pensiun | 0-1 Juta | 16-20 Juta | 31-40 Bulan | Lancar |
| Suwanto | Laki - Laki | 30-39 Tahun | Karyawan | 0-1 Juta | 16-20 Juta | 31-40 Bulan | Lancar |
| Ahmad Handi | Laki - Laki | 30-39 Tahun | Wiraswasta | 2-3 Juta | >20 Juta | 11-20 Bulan | Lancar |
| Ny. Rohinah | Perempuan | >60 Tahun | Pensiun | 0-1 Juta | 0-5 Juta | 1-10 Bulan | Tidak Lancar |
| Indra Irawan | Laki - Laki | 30-39 Tahun | Wiraswasta | 0-1 Juta | 6-10 Juta | 21-30 Bulan | Lancar |
| Jatmoko Ardy | Laki - Laki | 50-59 Tahun | Karyawan | 2-3 Juta | 11-15 Juta | 31-40 Bulan | Lancar |
| Budiman | Laki - Laki | 40-49 Tahun | Karyawan | 2-3 Juta | 6-10 Juta | 21-30 Bulan | Lancar |
| Roulina Hasibuan | Laki - Laki | 50-59 Tahun | Pensiun | 0-1 Juta | 6-10 Juta | 21-30 Bulan | Lancar |
| Supeni M | Laki - Laki | 40-49 Tahun | Pensiun | 0-1 Juta | 0-5 Juta | 31-40 Bulan | Lancar |
| Yatno B | Laki - Laki | >60 Tahun | Pensiun | 0-1 Juta | 0-5 Juta | >40 Bulan | Tidak Lancar |
| Turini | Perempuan | 50-59 Tahun | Pensiun | 0-1 Juta | 0-5 Juta | >40 Bulan | Tidak Lancar |
| Mariatini Akbar | Perempuan | 50-59 Tahun | Pensiun | 0-1 Juta | 16-20 Juta | >40 Bulan | Tidak Lancar |

Seleksi Data (*Data Selection*)

Data yang akan digunakan untuk diuji dalam proses klasifikasi diambil dan dibuat dalam satu data baru, dapat dilihat pada tabel 4 :

Tabel 4. Data Seleksi

| Nama | Jenis Kelamin | Umur | Pekerjaan | Gaji | Jumlah Pinjaman | Jangka Waktu | Klasifikasi |
|------|---------------|------|-----------|------|-----------------|--------------|-------------|
|------|---------------|------|-----------|------|-----------------|--------------|-------------|

| | | | | | | | |
|------------------|-------------|-------------|------------|----------|------------|-------------|--------------|
| Suciati suparno | Perempuan | >60 Tahun | Pensiun | 0-1 Juta | 6-10 Juta | >40 Bulan | Tidak Lancar |
| Maedalena | Perempuan | 40-49 Tahun | Wiraswasta | 0-1 Juta | 6-10 Juta | 31-40 Bulan | Lancar |
| Endang Suharti | Perempuan | 40-49 Tahun | Karyawan | 2-3 Juta | >20 Juta | 11-20 Bulan | Lancar |
| Sugiri | Laki - Laki | 50-59 Tahun | Karyawan | 2-3 Juta | 6-10 Juta | 11-20 Bulan | Lancar |
| Usrek | Laki - Laki | 50-59 Tahun | Pensiun | 0-1 Juta | 6-10 Juta | >40 Bulan | Tidak Lancar |
| Rodi Ashari | Laki - Laki | 40-49 Tahun | Karyawan | 2-3 Juta | 16-20 Juta | 31-40 Bulan | Lancar |
| Sarimin | Laki - Laki | 50-59 Tahun | Karyawan | 0-1 Juta | 6-10 Juta | 11-20 Bulan | Lancar |
| M. Arsyad | Laki - Laki | 50-59 Tahun | Pensiun | 0-1 Juta | 16-20 Juta | 31-40 Bulan | Lancar |
| Suwanto | Laki - Laki | 30-39 Tahun | Karyawan | 0-1 Juta | 16-20 Juta | 31-40 Bulan | Lancar |
| Ahmad Handi | Laki - Laki | 30-39 Tahun | Wiraswasta | 2-3 Juta | >20 Juta | 11-20 Bulan | Lancar |
| Ny. Rohinah | Perempuan | >60 Tahun | Pensiun | 0-1 Juta | 0-5 Juta | 1-10 Bulan | Tidak Lancar |
| Indra Irawan | Laki - Laki | 30-39 Tahun | Wiraswasta | 0-1 Juta | 6-10 Juta | 21-30 Bulan | Lancar |
| Jatmoko Ardy | Laki - Laki | 50-59 Tahun | Karyawan | 2-3 Juta | 11-15 Juta | 31-40 Bulan | Lancar |
| Budiman | Laki - Laki | 40-49 Tahun | Karyawan | 2-3 Juta | 6-10 Juta | 21-30 Bulan | Lancar |
| Roulina Hasibuan | Laki - Laki | 50-59 Tahun | Pensiun | 0-1 Juta | 6-10 Juta | 21-30 Bulan | Lancar |
| Supeni M | Laki - Laki | 40-49 Tahun | Pensiun | 0-1 Juta | 0-5 Juta | 31-40 Bulan | Lancar |
| Yatno B | Laki - Laki | >60 Tahun | Pensiun | 0-1 Juta | 0-5 Juta | >40 Bulan | Tidak Lancar |
| Turini | Perempuan | 50-59 Tahun | Pensiun | 0-1 Juta | 0-5 Juta | >40 Bulan | Tidak Lancar |
| Mariatini Akbar | Perempuan | 50-59 Tahun | Pensiun | 0-1 Juta | 16-20 Juta | >40 Bulan | Tidak Lancar |

Transformasi Data (Data Transformation)

Pada tahap transformasi ini, sebelum melakukan proses pengolahan data menggunakan Weka, data hasil *preprocessing* dengan format XLSX di ubah terlebih dahulu atau di transformasikan (*save as*) ke dalam format CSV (*Comma Separated Values*) atau ARFF (*atribute Relation File Format*) agar file yang ada dapat diproses pada aplikasi weka. CSV merupakan suatu format data dalam basis data di mana setiap *record* dipisahkan dengan tanda koma atau titik koma. ARFF adalah format file yang digunakan dalam Weka.

Proses Klasifikasi

Proses klasifikasi merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data yang sudah di transformasi. Dalam penelitian ini metode yang digunakan dalam proses mining adalah metode *Naïve Bayes*.

Seleksi Fitur

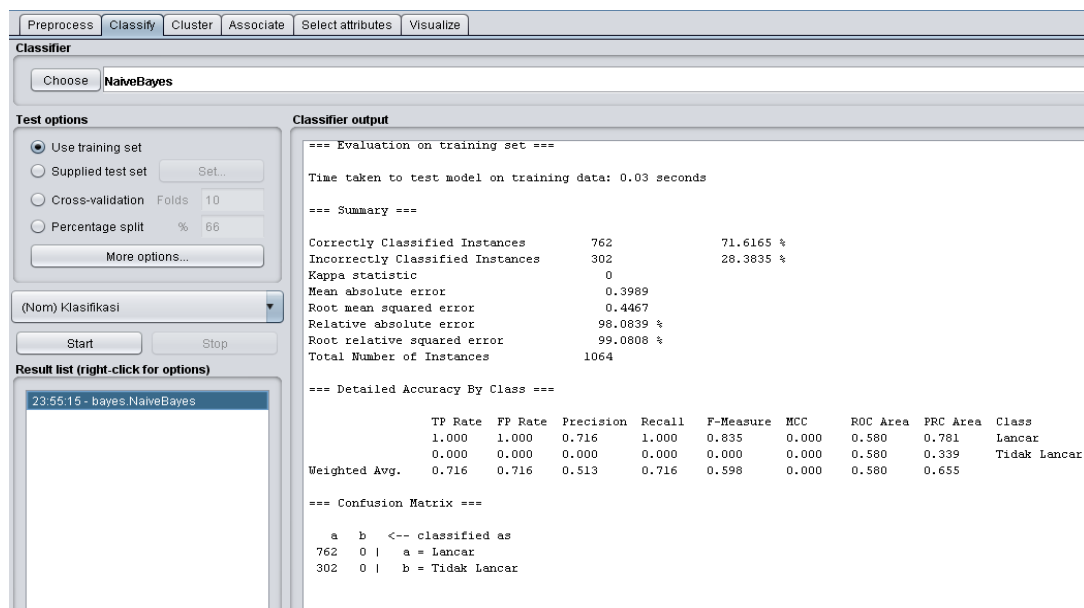
Proses seleksi fitur dalam penelitian ini menggunakan 300 data *testing* dengan jumlah 7 fitur yang ditentukan sebelumnya, akan tetapi karena 1 fitur tidak terlalu mempengaruhi dalam melakukan seleksi fitur/atribut, maka yang digunakan yaitu 6 fitur yang signifikan. Dari 6 fitur yang di seleksi atributnya/fitur menjadi 3 fitur yang digunakan dalam penelitian yaitu umur, gaji serta jangka waktu. Atribut seleksi fitur dapat dilihat pada tabel 5.

Tabel 5. Seleksi Fitur/atribut

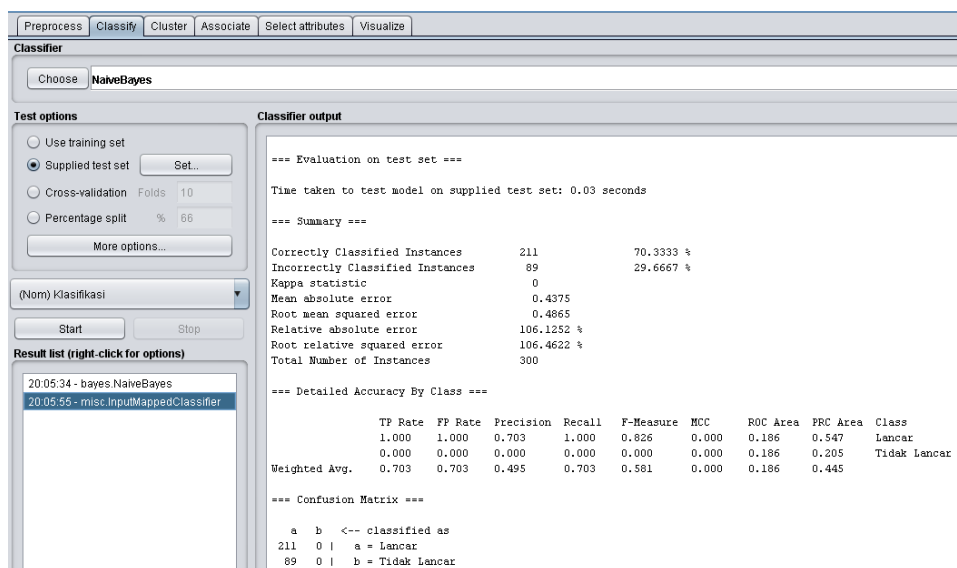
| No | Fitur | Format |
|----|-------|---------|
| 1 | Umur | Numerik |

| | | |
|---|--------------|---------|
| 2 | Gaji | Numerik |
| 3 | Jangka Waktu | Numerik |

Teori keputusan *bayes* adalah pendekatan statistik yang fundamental dalam pengenalan pola (*pattern recognition*), pendekatan ini didasarkan pada kuantifikasi *trade-off* antara berbagai keputusan klasifikasi dengan menggunakan probabilitas [10]. Penerapan algoritma pengklasifikasi *Naive Bayes* dapat mengurangi data *noise* pada *dataset* berukuran besar dan memiliki banyak kelas atau multi kelas sehingga akurasi klasifikasi algoritma *decision tree* dapat meningkat [11]. Untuk penerapan algoritma *Naive Bayes* data yang akan digunakan adalah data anggota pinjaman Koperasi Kredit Sejahtera. Penerapan algoritma dibagi menjadi dua, yaitu penerapan *Naive Bayes* dengan menggunakan aplikasi WEKA dan RAPIDMINER. Penerapan *naive bayes* dengan menggunakan weka digunakan untuk mengetahui klasifikasinya secara keseluruhan dan menampilkan hasil pengujian yang terbentuk dari data anggota pinjaman Koperasi Kredit Sejahtera. Dimana data training yang digunakan sebanyak 1064 *record* yang telah di filter serta data testingnya 300 *record*, variabel yang digunakan dalam proses pengujian algoritma *naive bayes* di weka adalah jenis kelamin, umur, pekerjaan, gaji, jumlah pinjaman serta jangka waktu.

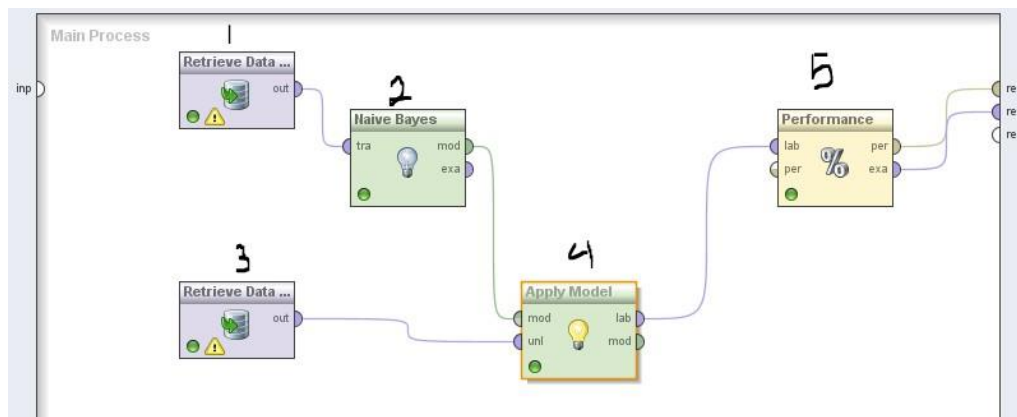


Gambar 2. Implementasi klasifikasi *naive bayes* data training dengan WEKA



Gambar 3. Implementasi klasifikasi *naive bayes* data testing dengan WEKA

Penerapan *naive bayes* dengan menggunakan rapidminer digunakan untuk menampilkan hasil pengujian yang terbentuk dari data anggota pinjaman Koperasi Kredit Sejahtera. Dimana data training yang digunakan sebanyak 1064 *record* yang telah di filter serta data testingnya 300 *record*, variabel yang digunakan dalam proses pengujian algoritma *naive bayes* di weka adalah jenis kelamin, umur, pekerjaan, gaji, jumlah pinjaman serta jangka waktu.



Gambar 4. Implementasi klasifikasi *naive bayes* data training dan data testing dengan RAPIDMINER.

Keterangan proses dari rapidminer :

- 1) Data *training* merupakan data yang sudah dibuat sebelumnya berisi tentang data latihan sebanyak 1064 *record*.
- 2) *Naive bayes* yaitu metode atau model *naive bayes* yang digunakan pada rapidminer
- 3) Data *testing* digunakan untuk menguji model *naive bayes* yang sudah jadi, data *testing* yang digunakan sebanyak 300 *record*.
- 4) *Apply model* digunakan untuk mengecek data training dari model *naive bayes* yang digunakan nanti untuk dibandingkan dengan data testing
- 5) *Performance* yaitu untuk mengetahui sebaran atau mengecek datanya dan mengetahui akurasinya.

| accuracy: 70.33% | | | |
|--------------------|-------------------|-------------|-----------------|
| | true Tidak Lancar | true Lancar | class precision |
| pred. Tidak Lancar | 0 | 0 | 0.00% |
| pred. Lancar | 89 | 211 | 70.33% |
| class recall | 0.00% | 100.00% | |

Gambar 5. Hasil akurasi, presisi dan akurasi menggunakan RAPIDMINER

Pada tahapan pengujian ini, penulis menggunakan WEKA dan RAPIDMINER dimana metode yang dipakai adalah Naive Bayes, untuk menguji tingkat klasifikasi akurasinya dapat di lihat dari nilai *presisi*, *akurasi* dan *recall*. Berikut ini hasil dari pengujian yang telah dilakukan :

```

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      1.000    1.000    0.703     1.000   0.826     0.000    0.186    0.547    Lancar
      0.000    0.000    0.000     0.000   0.000     0.000    0.186    0.205    Tidak Lancar
Weighted Avg.   0.703    0.703    0.495     0.703   0.581     0.000    0.186    0.445

=== Confusion Matrix ===

  a  b  <-- classified as
211  0  |  a = Lancar
 89  0  |  b = Tidak Lancar
    
```

Gambar 6. Hasil Pengujian menggunakan Weka

| accuracy: 70.33% | | | |
|--------------------|-------------------|-------------|-----------------|
| | true Tidak Lancar | true Lancar | class precision |
| pred. Tidak Lancar | 0 | 0 | 0.00% |
| pred. Lancar | 89 | 211 | 70.33% |
| class recall | 0.00% | 100.00% | |

Gambar 7. Hasil Pengujian menggunakan Rapidminer

Berdasarkan informasi yang ditampilkan *confusion matrix*, untuk mendiskripsikan hasil evaluasi/proses klasifikasi yang menggunakan data *training* 1064 *record* data *testing* sebanyak 300 *record* dari KOPDIT Sejahtera. Bahwa hasil akurasi yang diperoleh dari pengujian kedua tools tersebut sama yaitu 70,33% artinya nilai akurasi dari klasifikasi cukup tinggi [12]. Akurasi merupakan hasil perhitungan semua nilai prediksi yang benar dibagi dengan keseluruhan data. Sedangkan hasil *recall* (*Sensitivity*) dari kedua pengujian tools tersebut juga menunjukkan nilai yang sama yaitu 70,33% artinya nilai prediksi dari kelas yang terklasifikasikan dengan benar adalah 70,33%. *Recall* atau *Sensitivity* dihitung dari jumlah prediksi positif yang benar dibagi dengan jumlah keseluruhan kelas yang

positif. Hasil *precision* (presisi) dari pengujian kedua tools memperoleh nilai yang sama juga yaitu 100% atau 1 artinya jumlah kelas yang terklasifikasi dengan benar sangat tinggi [12]. Presisi dihitung dari jumlah keseluruhan nilai prediksi positif yang benar dibagi dengan jumlah keseluruhan prediksi kelas yang benar.

Nilai akurasi 70,33% pada penelitian ini dipengaruhi atau tergantung dari data *training* dan data *testing* beserta *class* atau label klasifikasinya, semakin banyak data *training* dan data *testing* serta *class* yang bernilai benar maka akan mempengaruhi tingkat keakurasiannya. Berdasarkan *confusion matrix* biasanya banyak digunakan untuk mendeskripsikan evaluasi dari klasifikasi. Untuk mencapai nilai akurasi 100% atau 1 maka hasil prediksi sesuai dengan nilai yang sebenarnya yaitu sama-sama benar (TP / *True Positive*) dan hasil prediksi salah dan nilai yang sebenarnya juga salah (TN / *True Negative*) nilainya harus tinggi, jadi kalau nilai TP dan TN nya pada *confusion matrix* tinggi maka tingkat akurasi bisa mendekati 100% atau 1 [12]. Nilai akurasi akan dikatakan cukup tinggi jika mendekati nilai 1 dan nilai akan dikatakan cukup rendah atau terburuk jika nilainya hampir mendekati 0 [12].

Dari hasil pengujian menggunakan kedua tools diatas yang menggunakan data *testing* sebanyak 300 *record*, yang berhasil diklasifikasikan dengan benar yaitu sebanyak 211 data anggota KOPDIT Sejahtera yang melakukan pinjaman, dan sebanyak 89 data anggota KOPDIT Sejahtera yang melakukan pinjaman tidak berhasil diklasifikasikan dengan benar. Berdasarkan gambar 6 dan gambar 7 dapat dilihat bahwa perbandingan hasil akurasi dipengaruhi karena, semakin banyak data atau *record* yang digunakan maka akan lebih baik tingkat akurasi.

4. Kesimpulan Dan Saran

Berdasarkan penelitian yang telah dilakukan dapat diambil kesimpulan sebagai berikut:

- 1) Algoritma *Naïve Bayes* dapat digunakan sebagai salah satu pilihan untuk mengklasifikasikan data anggota kelalaian pinjaman Koperasi Kredit Sejahtera. Penelitian ini menggunakan 1064 *record* data training dan 300 *record* data testing dengan 6 fitur didalamnya untuk mengetahui hasil klasifikasi menggunakan WEKA dan Rapidminer.
- 2) Dari pengujian menggunakan WEKA dan Rapidminer yang menggunakan 1064 *record* data *training* serta 300 *record* data *testing* menghasilkan tingkat akurasi sebesar 70,33%, recall 70,33% dan presisi 100% atau 1. Jadi semakin banyak data testing yang digunakan maka akan semakin mempengaruhi nilai akurasi.

Berdasarkan simpulan yang telah diuraikan terdapat beberapa saran untuk penelitian selanjutnya, diantaranya:

- 1) Agar menambahkan kelayakan awal atau pemrosesan awal sebelum memilih atribut dari data tersebut guna untuk mengetahui tingkat kelayakan dari atribut tersebut.
- 2) Penelitian selanjutnya diharapkan dapat membuat atributnya fleksibel supaya atribut atau fiturnya bisa ditambah atau dikurangi.

5. Daftar Pustaka

- [1]. Kasmir, "Dasar-Dasar Perbankan. 1st ed," Jakarta, Indonesia, 2010.
- [2]. Costa G. et al, *Data Mining for Effective Risk Analysis in a Bank Intelligence Scenario.*, 2007.
- [3]. P3RI PTPN VII (persero) Kopdit Sejahtera, "Anggaran Dasar dan Anggaran Rumah Tangga," 2017.
- [4]. H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode *Naïve Bayes*," *ILKOM Jurnal Ilmiah*, vol. 10, no. 2, 2018.
- A. N. Putri, "Penerapan *Naive Bayesian* Untuk Perankingan Kegiatan Di Fakultas Tik Universitas Semarang," *Jurnal SIMETRIS*, vol. 8, no. 2, 2017.
- [5]. R. D. Gunawan, T. Oktavia, and R. I. Borman, "Perancangan Sistem Informasi Beasiswa Program Indonesia Pintar (PIP) Berbasis Online (Tudi Kasus : SMA N 1 Kota Bumi)," *Jurnal Mikrotik*, vol. 8, no. 1, 2018.
- A. E. Kumala, R. I. Borman, and P. Prasetyawan, "Sistem Informasi Monitoring Perkembangan Sapi Di Lokasi Uji Performance (Studi Kasus : Dinas Peternakan Dan Kesehatan Hewan Provinsi Lampung)," *Jurnal TEKNOKOMPAK*, vol. 12, no. 1, 2018.
- [6]. A. Saleh, "Implementasi Metode Klasifikasi *Naive Bayes* Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," vol. vol 2(3), pp. 207-217, 2015.
- [7]. M. Ridwan, H. Suyono, and M. Sarosa, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma *Naive Bayes Classifier*," *Jurnal EECCIS*, vol. 7, no. 1, 2013.
- [8]. W. Muslehatin, M. Ibnu, and M. Mustakim, "Penerapan *Naive Bayes Classification* untuk Klasifikasi Tingkat Kemungkinan Obesitas Mahasiswa Sistem Informasi UIN Suska Riau," in *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) 9*, Pekanbaru, 2017.
- [9]. F. Ariani, A. Amir, N. Alan, and K. Rizal, "Klasifikasi Penetapan Status Karyawan Dengan Menggunakan Metode *Naïve Bayes*," *Paradigma - Jurnal Komputer dan Informatika*, vol. XX, no. 2, 2018.
- [10]. S. Adinugroho and Y. Sari Arum, *Implementasi Data Mining Menggunakan WEKA*. Malang: UB Press, 2018.