

Klasifikasi Kelayakan Kredit Pada Calon Debitur CS Finace Dengan Metode Naive Bayes

Hengki Tamando Sihotang¹, Rina Dewi²,
STMIK Pelita Nusantara, Program Studi Rekayasa Perangkat Lunak
E-mail : ¹hengkitamando@yahoo.com, rinadewi@gmail.com²

Abstrak

Kredit merupakan suatu sistem pinjam meminjam atas persetujuan yang telah disepakati diawal dengan phak bank atau institusi lembaga keuangan sejenis dengan nasabahnya. Dengan demikian analisa kelayakan pemberian kredit ini diharapkan dapat menganalisa solusi kesesuaian tingkat akurasi dalam sejumlah data yang sangat besar (big data) dengan melibatkan sebuah dataset yang nantinya akan dilakukan penambangan data dengan menggunakan tools rapid miner. Metode algoritma yang akan digunakan yakni metode data mining berupa Algoritma Naive Bayes Classifier. Klasifikasi data mining dapat membantu para analis kredit dalam hal menentukan signifikansi dan kelayakan pemberian kredit pada nasabah. Sehingga dari proses klasifikasi ini didapatkan berupa atribut penentu berupa kategori kelayakan dalam hal penentuan pemberian kredit kepada salah satu nasabah yang bersangkutan. Pengujian yang dilakukan yakni dengan menggunakan model confusion matrix yang melibatkan data training yang berbeda dalam hal jumlah atribut, dimana ekperimen pertama pengujian dilakukan terhadap 16 atribut data training kemudian ekperimen kedua pengujian dilakukan dengan melibatkan 9 atribut data training. Maka dari kedua eksperimen yang dilakukan akan diperoleh komparasi nilai akurasi, pengujian 16 atribut menghasilkan signifikansi akurasi 59,00% dan pengujian 9 atribut menghasilkan nilai akurasi sebesar 56,00 % sehingga dari kedua pengujian yang dilakukan dapat dipilih alternatif akurasi yang paling baik yakni pengujian dengan 16 atribut. Sehingga hasil dan output yang dihasilkan akan menghasilkan ketepatan pengujian atas suatu rekomendasi keputusan dalam membantu para profesional atau analis kredit di institusi keuangan dalam hal menentukan kelayakan pemberian kredit pada nasabahnya dengan bantuan algoritma yang digunakan untuk menyelesaikan suatu persoalan sehingga didapatkan pengidentifikasi terkait manakah yang layak atau tidak dikatakan sebagai penentuan pemberian kredit dengan mengacu pada penentuan kelas lancar atau bermasalah.

Kata Kunci—3-5 Data mining, klasifikasi kelayakan kredit, Naive Bayes Classifier

Abstract

Credit is a lending and borrowing system based on agreed agreements at the beginning with the phak bank or similar financial institution with its customers. Thus the feasibility analysis of credit granting is expected to be able to analyze the solution of the suitability of the accuracy level in a number of very large data (big data) by involving a dataset that will be mining data using rapid miner tools. The algorithm method that will be used is the data mining method in the form of the Naive Bayes Classifier Algorithm. Data mining classification can help credit analysts in determining the significance and feasibility of lending to customers. So that

the classification process is obtained in the form of determining attributes in the form of eligibility categories in terms of determining the provision of credit to one of the customers concerned. Tests carried out namely by using a confusion matrix model that involves training data that is different in terms of the number of attributes, where the first experiment testing is performed on 16 attributes of training data then the second experiment testing is done by involving 9 attributes of training data. So from the two experiments conducted will be obtained a comparison of the accuracy value, 16 attributes test produces a significance of 59.00% accuracy and 9 attributes test produces an accuracy value of 56.00% so that of the two tests conducted can be chosen the best alternative accuracy that is testing with 16 attributes. So the results and outputs produced will result in the accuracy of testing of a decision recommendation in helping credit professionals or analysts in financial institutions in terms of determining the eligibility of crediting to their customers with the help of an algorithm used to solve a problem so that an identifier related to which is feasible or not said as determining the granting of credit with reference to determining the current class or problematic.

Keywords—3-5 Data mining, Credit worthiness classification, Naive Bayes Classifier

1. PENDAHULUAN

Kepercayaan nasabah pada suatu instansi keuangan menjadi salah satu kemajuan badan usaha keuangan tersebut, karena nasabah memiliki peran penting dalam menunjang kegiatan operasional instansi keuangan dalam hal memilih alternatif penghimpunan dana dalam bentuk simpanan yang difavoritkan. dari penghimpunan dana tersebut pihak instansi akan menyalurkannya kembali kepada masyarakat dalam bentuk kredit guna meningkatkan taraf dan kualitas hidup masyarakat untuk turut serta dalam pembangunan nasional melalui instansi keuangan berupa koperasi, salah satu contohnya adalah koperasi simpan pinjam simpanan pemeungkeut banda (SPB).

Koperasi simpan pinjam merupakan suatu badan usaha keuangan berasaskan kekeluargaan dengan tujuan untuk mensejahterakan anggotanya. Dimana koperasi tersebut menyediakan jasa simpan pinjam bagi nasabahnya demi menjamin pelayanan ekstra. Namun disisi lain selalu ada resiko tertinggi yang harus dihadapi oleh instansi koperasi tersebut diantaranya ketidaktepatan manajemen waktu yang telah disepakati diawal, juga kegagalan saat pembayaran berupa permasalahan macetnya pembayaran kredit yang sebelumnya disalurkan. Seiring era globalisasi yang meningkat secara signifikan, memungkinkan bagi pihak koperasi untuk memperbaiki dan mengevaluasi kredit dengan menggunakan set data dalam jumlah yang sangat besar dengan menggunakan proses penambangan data menggunakan teknik algoritma data mining klasifikasi. Data mining telah terbukti memegang peran penting dilinimasa industri perbankan keuangan dalam skala besar seperti bank ataupun skala kecil seperti koperasi misalnya. Banyak yang dapat membantu para manajer untuk menggunakan aplikasi data mining sebagai alat untuk mematenkan suatu keputusan terkait kelayakan kredit nasabah.

Pada penelitian yang dilakukan oleh susanto tahun 2015 bahwasannya Hasil model plot view analisa data transaksi penjualan untuk melihat perilaku pola pembeli dengan metode naive bayes dapat diprediksi dengan density tertinggi pada class siang mencapai 16,8% pada hari minggu[1]. Ada beberapa penelitian yang relevan mengenai studi kasus diatas diantaranya penelitian mengenai *Implementasi Data Mining Untuk Penentuan Kelayakan Pemberian Kredit Nasabah Menggunakan Algoritma C4.5*, secara garis besar penelitian ini adalah penelitian yang dilakukan di Koperia (Koperasi Warga Komplek Gandaria) yang membahas tentang persoalan yang dihadapi oleh pengurus koperasi salah satunya sulitnya dalam menentukan pemberian kelayakan kredit, seringkali ada banyak resiko permasalahan yang dihadapi salah satunya pembayaran angsuran sering kali mengalami kemacetan dikarenakan memang sering menunggak. Hal ini tentu akan berdampak bagi sistem keuangan perusahaan yang bersangkutan[2].

Penelitian yang sama dilakukan oleh Heri Hidayanto dan Achmad Wahid Kurniawan yakni penelitian tentang *Klasifikasi Kelayakan Kredit Calon Debitur Bank Menggunakan Algoritma Decision Tree C4.5*, fokus pembahasannya berupa pengawasan yang dilakukan oleh bank terkait untuk manajemen resiko tertinggi yang akan dihadapi dengan cara membedakan kreditur yang memiliki potensi terjadi macet dalam hal angsuran kreditnya menggunakan model algoritma klasifikasi C4.5 yang dinilai dalam segi kelebihannya dapat divisualisasikan dalam bentuk pohon keputusan serta memiliki tingkat akurasi yang lebih tinggi dari metode klasifikasi lain [3].

Penelitian serupa dilakukan oleh Ika Menariyanti yakni tentang *Klasifikasi Data Mining Dalam Menentukan Pemberian Kredit Bagi Nasabah Koperasi*. Dimana, dalam penelitian ini menggunakan metode *Cross Validation, Confusion Matrix, ROC Curve, dan T-Test*, untuk menentukan output hasil yang paling akurat dalam penentuan pemberian kredit bagi nasabah koperasi. Karena sebagaimana kita ketahui bahwa pihak instansi menyadari akan resiko yang akan timbul salah satunya ancaman profitabilitas dan kelangsungan yang menjadi hal yang harus diwaspadai dalam pemberian kredit. Tujuan dari adanya mining data ini guna menyeleksi, mengidentifikasi, dan memilih kelayakan nasabah yang sekiranya dapat dijadikan faktor penentu kelayakan kredit nasabah di koperasi [4].

Penelitian yang sama dilakukan oleh Dina Rusdiana Sari yang membahas tentang *Penerapan Data Mining Klasifikasi Dengan Menggunakan Algoritma C4.5 Pada Data Nasabah Kredit Di Bank Kredit Desa (BKD) Kabupaten Rembang*, salah satu fokusnya adalah menciptakan suatu sistem informasi yang akurat dalam menentukan kategori macet dan lancar dalam hal kredit nasabah dengan menggunakan algoritma C4.5 [5]. Maulida Azkiya dkk, juga melakukan penelitian tentang *Klasifikasi Nasabah Bank "X" di Provinsi Lampung Menggunakan Analisis Diskriminasi Kernel*, dimana metode yang digunakan agak berbeda dengan penelitian-penelitian yang dilakukan sebelumnya yakni dengan menggunakan metode analisis diskriminasi yang digunakan sebagai metode utama pengklasifikasian. [6]

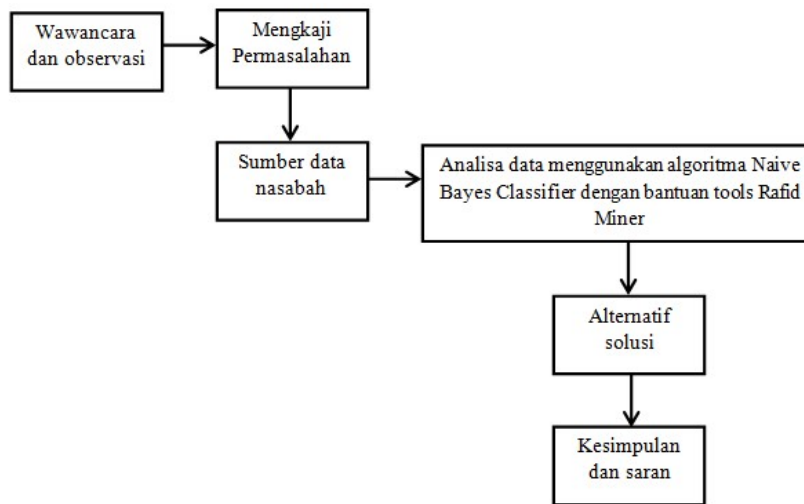
Berdasarkan analisis yang dilakukan dari beberapa penelitian-penelitian yang relevan didapatkan signifikansi studi kasus yang sama berupa kelayakan pemberian kredit pada suatu nasabah namun ada beberapa hal yang membedakan yakni dalam penentuan model dan metode algoritmanya salah satunya disini kami menggunakan algoritma Naive Bayes Classifier(NBC). Dimana dengan metode ini diharapkan dapat menghasilkan informasi yang akurat terkait penentuan kelayakan pemberian kredit. Karena algoritma klasifikasi Naive Bayes Classifier dinilai sangat efektif dalam menangani data kuantitatif atau kualitatif, perhitungannya cepat dan efisien.

2. METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah metode penilaian deskriptif kualitatif, Dimana penelitian ini menggambarkan gejala sosial dengan kata lain penelitian ini bertujuan untuk menggambarkan sifat sesuatu yang tengah berlangsung pada saat studi. Juga untuk memahami (*to understand*) fenomena atau gejala sosial dengan lebih menitikberatkan pada gambaran lengkap tentang fenomena yang dikaji. Harapannya ialah diperoleh pemahaman yang mendalam tentang fenomena untuk selanjutnya dihasilkan dari sebuah teori [7]. Selain itu kami menekankan aspek penelitian analitis yang mengkaji dan memperoleh sumber data berdasarkan analisis dokumen yang ada. Sumber data tersebut berasal dari responden berupa dokumen data penduduk yang akan dijadikan landasan riset sebagai dataset untuk menentukan klasifikasi kelayakan pemberian kredit pada badan usaha koperasi.

Penelitian dilakukan dengan melakukan wawancara terlebih dahulu dengan narasumber di objek penelitian yang kami tinjau salah satunya ada pimpinan koperasi, juga ada staf atau

pengurus koperasi . Pasca proses wawancara dilakukan, kami meminta sumber data nasabah disana sebagai dataset awal. Adapun proses penelitian dapat dilihat dalam bagan berikut ini :



Gambar 1. Langkah-langkah penelitian

Sumber data yang dihasilkan berupa dataset dengan jumlah lebih dari 700 record data yang berisi data nasabah. Kemudian data tersebut akan dijadikan landasan awal proses mining data sehingga didapatkan suatu output signifikansi dan analisa hasil yang menyatakan layak tidaknya dalam penentuan pemberian kredit. Set data dibagi kedalam dua bagian yakni ada data testing yang merupakan data yang digunakan untuk mengukur sejauh mana classifier berhasil melakukan klasifikasi dengan benar. Data training (training set) merupakan data yang digunakan oleh algoritma klasifikasi (decision tree, bayesian, neural network, SVM) untuk membentuk sebuah model classifier. Model ini merupakan representasi pengetahuan yang akan digunakan untuk prediksi kelas data baru yang belum pernah ada.

Data diperoleh dari salah satu Koperasi Simpanan Pameungkeut Banda (SPB) di Tasikmalaya tahun 2017. Data tersebut sebanyak 767 record yang memiliki *atribu Nama, Usia, Jumlah Pinjaman, Jangka waktu, Jumlah angsuran perbulan, Kategori pinjaman, Jenis, BSE, col, BGD, BGP, Saldo Nominatif, Tunggakan pokok, Tunggakan bunga, Status Kredit*. Atribut yang ada ditabel dataset merupakan keseluruhan data diskrit. Berikut sample datasetnya :

Tabel 1. Dataset keseluruhan

No	Nama	Usia	Jumlah Pinjaman	Jangka Waktu	Jml angsuran Per bulan	Kategori pinjaman	Jenis	BSE	Col	BGD	BGP	Saldo Nominatif	Tunggakan pokok	Tunggakan Bunga
1	Andi	40	345000	1	345000	100	301	6000	1	874	875	345000	345000	0
2	Angga	31	350000	7	55716	100	301	6000	1	874	875	390000	111428	0
3	Anhar	37	649926	6	108321	100	301	6000	1	874	875	649926	216642	0
4	Aminulloh	2	459168	12	38264	100	301	6000	1	874	875	459168	382640	0
5	Baban	34	3055499	8	381937,41	100	301	6000	1	874	875	3055499	1527749,48	0
6	Bambang	49	2000000	16	0	100	301	6000	1	874	875	-85000	0	0
.	Bani	31	8333334	10	833333,4	100	301	6000	1	874	875	8333334	333333,6	0
.	Delis	27	4435001	8	671098	100	301	6000	1	874	875	4435001	0	0
.	Desi	42	560000	8	95221	100	301	6000	1	874	875	660800	100800	0
767	Zaenal	36	2000000	4	550000	100	301	6000	2	874	000	1000000	1000000	100000

Didalam tabel menunjukkan data transaksi kredit baik yang bermasalah maupun tidak bermasalah (lancar). Dari keseluruhan atribut, ada beberapa atribut yang akan dijadikan sebagai landasan penentuan kredit nasabah yakni mengacu pada atribut *Usia,*

Jumlah Pinjaman, Jangka_waktu, Jumlah_angsuran_perbulan, Kategori_pinjaman, Jenis, BSE, col, BGD, BGP, Saldo_Nominatif, Tunggalan_pokok, Tunggalan_bunga, Status_Kredit.

Tabel set data ini merupakan suatu data murni yang belum diolah sedemikian rupa menjadi data training. Jika kita lihat secara seksama pada tabel data set ini maka akan didapati beberapa atribut yang tidak digunakan dan record yang duplikat. Maka dari itu diperlukan teknik preprocessing data yakni :

Dari keseluruhan dataset yang didapat itu dibagi menjadi data *training* dan *testing* dimana persentase sebanyak 80 % atau setara dengan 613 record data untuk pembagian data training yang nantinya akan digunakan sebagai pemodelan dalam metode Naive Bayes Classifier, model ini akan digunakan sebagai prediksi kelas data baru yang belum pernah ada. Berikut contoh sampel data training :

umur	jml_pinjaman	Jangka_waktu	jml_angsuran_per_bulan	kategori_pinjam	jenis	BSE	col	BGD	BGP	saldo_nominatif	tunggalan_pokok	tunggalan_bunga	status_kredit
45	3520000	12	293333,37	100	905	6000	1	874	875	2920000	0	0	LANCAR
44	4520000	12	376666,74	100	905	6000	1	874	875	4520000	553333,32	0	BERMASALAH
45	5000000	12	416666,74	100	905	6000	1	874	875	5000000	833333,32	0	BERMASALAH
45	10000000	10	1000000	100	905	6000	1	874	800	10000000	2000000	0	BERMASALAH
50	1275000	10	127500	100	905	6000	1	874	875	1275000	1147500	0	BERMASALAH
60	1200000	12	100000	100	905	6000	1	874	875	1184064,99	884064,99	0	BERMASALAH
46	1420000	72	19722,38	100	905	6000	1	874	875	1420000	236666,64	0	BERMASALAH
60	880000	8	110000	100	905	6000	1	874	875	880000	440000	0	BERMASALAH
35	2312000	50	46240	100	905	6000	1	874	875	2312000	462400	0	BERMASALAH
28	1000000	10	100000	100	905	6000	1	874	875	1000000	900000	0	BERMASALAH
29	3200000	10	320000	100	905	6000	1	874	000	2880000	320000	0	BERMASALAH
38	500000	5	100000	100	905	6000	1	874	875	400000	300000	0	BERMASALAH
52	950000	4	237500	100	905	6000	1	874	000	275000	0	0	LANCAR
54	2000000	14	200000	100	905	6000	1	874	000	971428,58	685714,26	0	BERMASALAH
43	700000	50	140000	100	905	6000	1	874	875	700000	140000	0	BERMASALAH

Gambar 2. Data training

Tidak hanya dibagi kedalam data training saja yang digunakan sebagai landasan pemodelan algoritma Naive Bayes, namun dari sejumlah besar dataset yang ada diperlukan pembagian data uji (data testing) yang digunakan untuk menguji sejauh mana classifier berhasil melakukan klasifikasi dengan benar. Berikut contoh sample data testing:

usia	jml_pinjaman	Jangka_waktu	jml_angsuran_per_bulan	kategori_pinjaman	jenis	BSE	col	BGD	bi_gol_penjamin	saldo_nominatif	tunggalan_pokok	tunggalan_bunga	status_kredit
36	439000,01	6	95300,01	100	901	6000	1	874	000	349250,01	845250,01	31250	BERMASALAH
42	783333,34	9	95370,37	100	901	6000	1	874	000	691866,34	691866,34	6664,03	BERMASALAH
34	5000000	24	245833,41	100	901	6000	1	874	000	3958333,35	395833,35	749999,99	BERMASALAH
32	1000000	12	90833,37	100	901	6000	1	874	000	60000,01	60000,01	7500	BERMASALAH
34	2000000	12	181666,74	100	901	6000	1	874	000	1100000	1100000	90000	BERMASALAH
30	1000000	3	343333,34	100	901	6000	1	874	000	1000000	1000000	30000	BERMASALAH
44	900000	3	327000	100	902	6000	1	874	000	600000	300000	4000	BERMASALAH
31	3000000	6	590000	100	902	6000	1	874	000	1500000	500000	90000	BERMASALAH
48	1000000	8	155000	100	902	6000	1	874	875	1500000	400000	0	BERMASALAH
45	400000	4	104000	100	902	6000	1	874	000	160000	160000	0	BERMASALAH
2	1762500	15	117500	100	903	6000	1	874	875	8246000	5497333,32	0	BERMASALAH
2	1500000	15	100000	100	903	6000	1	874	875	7500000	7500000	0	BERMASALAH
-43	275000	8	20000	100	903	6000	1	874	875	14598000	3831800	0	BERMASALAH
32	8246000	6	1374333,35	100	904	6000	1	874	875	1445000	722500	0	BERMASALAH
41	7500000	2	3750000	100	904	6000	1	874	875	10000000	4000000	0	BERMASALAH
							1	874	000	600000	600000	0	BERMASALAH
							1	874	000	100000	100000	0	BERMASALAH

Gambar 3. Data Testing

Rumus Algoritma Naive Bayes Classifier (NBC)

$$P(C|X) = \frac{P(x|c) P(c)}{P(x)} \dots\dots\dots(1)$$

Keterangan :

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

$$Posterior = \frac{prior \times likelihood}{evidence} \dots\dots\dots(2)$$

Keterangan :

- Posterior : Peluang masuknya sampel karakteristik tertentu dalam kelas C
- Prior : Peluang munculnya kelas C (sebelum masuknya sampel tersebut)
- Likelihood : Peluang kemunculan karakteristik sampel pada kelas C
- Evidence : Peluang kemunculan karakteristik sampel secara global

$$\begin{aligned}
 P(c|X_1, \dots, X_n) &= P(C)P(X_1, \dots, X_n|C) \\
 &= \frac{P(C)P(X_1|c)P(X_2|c, X_1) \dots P(X_n|c, X_1, X_2, \dots, X_{n-1})}{P(X_1|c)P(X_2|c, X_1) \dots P(X_n|c, X_1, X_2, \dots, X_{n-1})} \quad (3) \\
 &= P(C)P(X_1|c)P(X_2|c, X_1)P(X_3|c, X_1, X_2) \dots P(X_n|c, X_1, X_2, \dots, X_{n-1}) \\
 &= P(C)P(X_1|c)P(X_2|c, X_1)P(X_3|c, X_1, X_2) \dots P(X_n|c, X_1, X_2, \dots, X_{n-1})
 \end{aligned}$$

$$P(c|X_1, \dots, X_n) = P(c) \prod_{i=1}^n P(X_i|c) \quad (4)$$

$$P(c|X) = \frac{P(c) \prod_{i=1}^n P(X_i|c)}{\prod_{i=1}^n P(X_i|c)} \quad (5)$$

3. HASIL DAN PEMBAHASAN

Apabila diberikan input baru maka klasifikasi data calon nasabah dapat ditentukan melalui langkah berikut :

1. Data Inputan

Data inputan ini merupakan data testing dimana jika ada calon nasabah yang baru yang class label nya belum diketahui secara pasti dan masih kosong. Maka dari itu kita bisa mengklasifikasikan menggunakan perhitungan metode Naive Bayes Classifier.

Tabel 2. Data Testing

nama_nasabah	jenis_kelamin	usia	jml_pinjaman	jangka waktu	jml_angsuran_per_bulan	kategori_pinjaman	jenis_pinjaman
Ananda Ferdinan	L	36	439000,01	6	95300,01	100	301

BSE	col	BGD	BGP	saldo_nominatif	tunggakan_pokok	tunggakan_bunga	status_kredit
6000		2	874	000	345250,01	345250,01	31250?????

2. Menghitung jumlah class/label dan menentukan class yang sama

	BERMASALAH	LANCAR	100%
P(Usia ...)	27%	73%	100%
P(jml_pinjaman ...)	0,024038462	0,0009319	BERMASALAH
P(jangka waktu ...)	0	0	BERMASALAH
P(jml_angsuran_per_bulan ...)	0,052884615	0,064516129	BERMASALAH
P(kategori_pinjaman ...)	0	0	BERMASALAH
P(jenis_pinjaman ...)	0,942307692	0,9396681	BERMASALAH
P(BSE ...)	0	0,992811341	BERMASALAH
P(col ...)	0,009613385	0,24930681	BERMASALAH
P(BGD ...)	0,991192308	0,991192308	BERMASALAH
P(BGP ...)	0,947113385	0,377000932	BERMASALAH
P(saldo_nominatif ...)	0	0	BERMASALAH
P(tunggakan_pokok ...)	0	0	BERMASALAH
P(tunggakan_bunga ...)	0	0	BERMASALAH

3. Kalikan semua variabel lancar dan bermasalah

$$\begin{aligned}
 \text{LANCAR} &= P(\text{Usia} = 36 | Y = \text{Lancar}) * P(\text{jumlah pinjaman} = 439000.01 | Y = \text{Lancar}) * \\
 &P(\text{Jangka Waktu} = 6 | Y = \text{Lancar}) * P(\text{jml angsuran per bulan} = 95300,01 | Y = \text{Lancar}) * \\
 &P(\text{kategori pinjaman} = 100 | Y = \text{Lancar}) * P(\text{Jenis} = 301 | Y = \text{Lancar}) * \\
 &P(\text{BSE} = 6000 | Y = \text{Lancar}) * P(\text{col} = 2 | Y = \text{Lancar}) * P(\text{BGD} = 874 | Y = \text{Lancar}) * \\
 &P(\text{BGP} = 000 | Y = \text{Lancar}) * P(\text{saldo ominatif} = 345250,01 | Y = \text{Lancar}) * P(\text{Tunggakan pokok} = \\
 &345250,01 | Y = \text{Lancar}) * P(\text{Tunggakan bunga} = 31250 | Y = \text{Lancar})
 \end{aligned}$$

$$\text{LANCAR} = 0,024038462 * 0 * 0,052884615 * 0 * 1 * 0,942307692 * 1 * 0,009615385 * 0,995192308 * 0,947115385 * 0 * 0 * 0 = 0,00$$

$$\text{BERMASALAH} = P(\text{Usia} = 36 | Y= Bermasalah) * P(\text{jumlah pinjaman} = 439000.01 | Y= Bermasalah) * P(\text{Jangka Waktu} = 6 | Y= Bermasalah) * P(\text{jml_angsuran_per_bulan} = 95300,01 | Y= Bermasalah) * P(\text{kategori_pinjaman} = 100 | Y= Bermasalah) * P(\text{Jenis}=301|Y= Bermasalah) * P(\text{BSE}=6000|Y= Bermasalah) * P(\text{col}=2|Y=Lancar) * P(\text{BGD}=874|Y=Lancar)* P(\text{BGP}=000|Y=Lancar) * P(\text{saldo_ominatif}=345250,01|Y= Bermasalah) * P(\text{Tunggakan_pokok}= 345250,01|Y= Bermasalah) * P(\text{Tunggakan_bunga}= 31250|Y= Bermasalah)$$

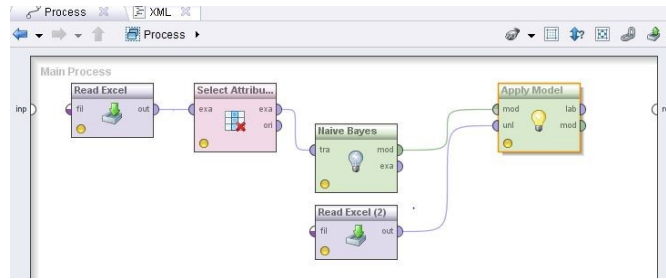
$$\text{BERMASALAH} = 0,0609319 * 0 * 0,064516129 * 0 * 1 * 0,9390681 * 0,992831541 * 0,29390681 * 0,995192308 * 0,577060932 * 0 * 0 * 0 = 0,00$$

DATA TESTING	jenis_pinjaman	BSE	col	BGD	BGP	saldo_nominatif	tunggakan_poko	tunggakan_bung
	301	6000	2	874	000	345250,01	345250,01	31250
	nama_nasabah	jenis_kelamin	usia	jml_pinjaman	Jangka waktu	jml_angsuran_per_bul	kategori_pinjaman	jenis_pinjaman
	Ananda Ferdinan	L	36	439000,01	6	95300,01	100	301
	Prediction	CLASS PREDICTION	LANCAR	BERMASALAH				
		BERMASALAH	0,0%	0,0%				

Maka hasil akhir diperoleh bahwa kelas dari sampel data testing ini memperoleh kategorikal kelas data “BERMASALAH” untuk calon nasabah bernama Ananda Ferdian.

3.2 Hasil Pengujian

Tahapan selanjutnya adalah melakukan eksperimen 16 atribut data training dengan menggunakan tools RapidMiner. Hal yang pertama dilakukan adalah mengimport dataset yang sudah disiapkan sebelumnya yang bersumber dari objek penelitian. Untuk menganalisa, dibutuhkan data training. Data training yang akan dimasukan kedalam rapid miner bisa dalam format .csv, .xls, .mdb dan lain-lain. Data yang kami gunakan adalah dalam format .xls yang dalam hal ini dataset yang diperoleh ini sudah diperoleh menjadi data training dengan menggunakan teknik preprocessing. Kemudian dilakukan pemodelan dengan menggunakan tools RapidMiner. Algoritma yang pertama diterapkan adalah algoritma *Naive Bayes Classifier*. Berikut hasil dari pengujian dengan menggunakan RapidMiner. Dalam hal ini mengkomparasikan atribut dengan jumlah atribut yang berbeda, yaitu eksperimen pertama kami menggunakan 16 atribut dan eksperimen kedua kami menggunakan 9 atribut atau menghilangkan salah satu atribut pada data training yang sama sehingga dapat diketahui tingkat akurasi yang lebih tinggi.



Gambar 4 . Implementasi Pemodelan Naive Bayes

Dari gambar 4 diatas, merupakan suatu bentuk implementasi pemodelan algoritma Naive Bayes Classifier dengan menggunakan tools rapid miner yang melibatkan sejumlah operator-operator yang berasal dari hirarki operator view. Dimana operator read excel digunakan untuk membaca file yang berekstensi .xls dalam hal ini data excel yang akan diimport yakni dengan melibatkan suatu data training (set training), Select Attribute berperan sebagai operator yang digunakan untuk mempermudah dalam hal pemilihan atribut, operator Naive Bayes adalah operator yang digunakan untuk metode algoritma yang akan diujung dalam pemodelan klasifikasi, operator Apply Model digunakan untuk upaya membaca data yang akan di estimasi berdasarkan data yang telah di pelajari.

Data View | Meta Data View | Plot View | Advanced Charts | Annotations

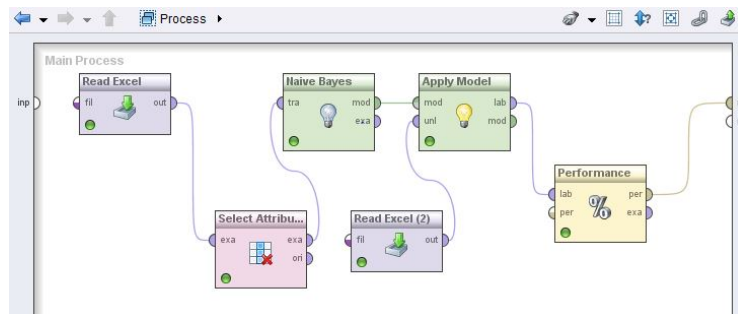
ExampleSet (102 examples, 4 special attributes, 15 regular attributes) View Filter (102 / 102): all

Row No.	status kredit	confidence(BER...	confidence(LAN...	prediction(status...	nama_nasabah	jenis_kel...	usia	jml_pinjam...	jangka waktu	jml_angsur...	kategori
1	BERMASALAH	1	0	BERMASALAH	Ananda Ferdinan	P	36	439000.010	6	95300.010	100
2	BERMASALAH	1	0	BERMASALAH	Asep Surya Somantri	P	42	783333.340	9	95370.370	100
3	BERMASALAH	1	0	BERMASALAH	Aryani Suherman	P	34	5000000	24	245833.410	100
4	BERMASALAH	1	0	BERMASALAH	Cici Sumiati	L	32	1000000	12	90833.370	100
5	BERMASALAH	1	0	BERMASALAH	Dudi Herdian	P	34	2000000	12	181666.740	100
6	BERMASALAH	1	0	BERMASALAH	Eti Safitri	P	30	1000000	3	343333.340	100
7	BERMASALAH	1	0	BERMASALAH	Elin Sukawati	P	44	900000	3	327000	100
8	BERMASALAH	1	0	BERMASALAH	Fera Husna	P	31	3000000	6	590000	100
9	BERMASALAH	1	0	BERMASALAH	Gunawan Liandar	L	48	1000000	8	155000	100
10	BERMASALAH	1	0	BERMASALAH	Hani Suryani	P	45	400000	4	104000	100
11	BERMASALAH	1	0	BERMASALAH	Ii Tari	L	2	1762500	15	117500	100
12	BERMASALAH	1	0	BERMASALAH	Jojo Suteja	L	2	1500000	15	100000	100
13	BERMASALAH	1	0	BERMASALAH	Jejen	L	-43	275000	8	20000	100
14	BERMASALAH	1	0	BERMASALAH	Syarifudin	L	32	8246000	6	1374333.350	100
15	BERMASALAH	1	0	BERMASALAH	Sofi Handayani	P	41	7500000	2	3750000	100
16	BERMASALAH	1	0	BERMASALAH	Sri Nurhayati	P	36	14598000	15	973200	100
17	BERMASALAH	1	0	BERMASALAH	Toni Rachmansyah	L	52	1445000	8	180625	100
18	BERMASALAH	1	0	BERMASALAH	Vina Nashruna Fauziah	L	28	10000000	10	1000000	100

Gambar 5. Signifikansi dan konsistensi output data view pemberian kredit nasabah

Dari gambar 5 diatas, merupakan output yang dihasilkan dari pengolahan data mining klasifikasi naive bayes. Tampak dalam tabel tersebut memang di peroleh signifikansi dan kekonsistenan hasil yang didapatkan berdasarkan perhitungan manualisasi. Dengan demikian nama nasabah dengan nama Ananda Ferdinan terbukti memiliki status kredit bermasalah dan ini relevan dengan perhitungan metode algoritma Naive Bayes

3.2.1 Integrasi antar operator untuk menghasilkan akurasi



Gambar 6. Implementasi akurasi dengan 16 atribut

Dari gambar 6 di atas merupakan bentuk implementasi untuk memperoleh nilai akurasi dalam artian mengukur dan mengevaluasi model klasifikasi, sejauh mana implementasi kalasifikasi dilakukan. Sehingga nantinya dengan adanya perolehan persentase akurasi akan dijadikan landasan untuk mengukur kedekatan antara nilai prediksi dengan nilai aktual. Operator *apply model* dan *performance* digunakan untuk membaca data yang diestimasi serta operator *performance* yang berasal dari repository *performance measurement* adalah operator yang berfungsi sebagai operator klasifikasi serta digunakan untuk evaluasi kinerja statistik tugas klasifikasi. Dengan adanya operator tersebut akan senantiasa memberikan daftar nilai kriteria kinerja tugas klasifikasi berupa nilai *accuracy* dan juga nilai *precision* yang digunakan untuk tingkat keterpatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem.

3.2.3 Akurasi

	true BERMASALAH	true LANCAR	class precision
accuracy: 59.00%			
pred. BERMASALAH	59	41	59.00%
pred. LANCAR	0	0	0.00%
class recall	100.00%	0.00%	

```

PerformanceVector

PerformanceVector:
accuracy: 59.00%
ConfusionMatrix:
True:  BERMASALAH  LANCAR
BERMASALAH:  59    41
LANCAR:  0      0
  
```

Gambar 7. Output nilai akurasi dengan 16 atribut

Dari gambar 7 di atas dapat dijelaskan bahwa output yang dihasilkan adalah berupa nilai akurasi sebesar **59.00%** dalam bentuk tabel view dalam lingkup eksperimen pertama, yakni melakukan pengujian dengan melibatkan 16 atribut data. Jadi nantinya hasil akurasi tersebut akan diperbandingkan dengan nilai akurasi dengan jumlah atribut yang berbeda melalui eksperimen kedua yakni sebanyak 9 atribut. Berikut hasil implementasi pengujian eksperimen kedua yang memperlihatkan nilai akurasi yang berbeda dengan eksperimen pertama.

3.2.4 Implementasi Pengujian 9 atribut

	true BERMASALAH	true LANCAR	class precision
accuracy: 56.00%			
pred. BERMASALAH	56	41	57.73%
pred. LANCAR	3	0	0.00%
class recall	94.92%	0.00%	

```

PerformanceVector

PerformanceVector:
accuracy: 56.00%
ConfusionMatrix:
True:  BERMASALAH  LANCAR
BERMASALAH:  56    41
LANCAR:  3      0
  
```

Gambar 8. Output nilai akurasi dengan 9 atribut

Dari gambar 8 diatas menunjukkan bahwa, signifikansi akurasi pada ekperimen kedua yang melibatkan 9 atribut menghasilkan nilai akurasi sebesar 56,00% lebih rendah dari klasifikasi akurasi pada ekperimen pertama. Untuk lebih jelasnya berikut komparasi nilai accuracy dan recall untuk algoritma Naive Bayes Classifier.

3.2.5 Evaluasi Model Confusion Matrix

Mengacu pada gambar 7 yang dijadikan sampel perhitungan matematis pemodelan evaluasi *confusion matrix* dimana, *TrueBERMASALAH* merupakan tupel positif didata set yang diklasifikasikan positif berjumlah 59, sedangkan *trueLANCAR* merupakan tuple negatif didata set yang diklasifikasikan negatif berjumlah 0. *FalseBERMASALAH* merupakan tupel positif didata set yang diklasifikasikan negatif berjumlah 0, sedangkan *FalseLANCAR* merupakan tupel negatif di data set yang diklasifikasikan positif berjumlah 41.

Berikut rumus untuk menghitung model confusion matrix :

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \dots\dots\dots(6)$$

$$\text{Presisi} = \frac{TP}{FP+TP} * 100\% \dots\dots\dots(7)$$

$$\text{Recall} = \frac{TP}{FN+TP} * 100\% \dots\dots\dots(8)$$

Keterangan :

- TP = True Positif
- TN = True Negatif
- FP = False Positif
- FN = Flase Negatif

Pembuktian matematis nilai akurasi pada data training 16 atribut :

$\begin{aligned} \text{Akurasi} &= \frac{TP+TN}{TP+TN+FP+FN} * 100\% \\ &= \frac{59+0}{59+0+0+41} * 100\% \\ &= \frac{59}{100} * 100\% \\ &= 0,59 * 100\% \\ &= \mathbf{59,00} \end{aligned}$	$\begin{aligned} \text{Presisi} &= \frac{TP}{FP+TP} * 100\% \\ &= \frac{59}{0+59} * 100\% \\ &= \frac{59}{59} * 100\% \\ &= 1 * 100\% \\ &= \mathbf{100,00\%} \end{aligned}$	$\begin{aligned} \text{Recall} &= \frac{TP}{FN+TP} * 100\% \\ &= \frac{59}{41+59} * 100\% \\ &= \frac{59}{100} * 100\% \\ &= \frac{59}{100} * 100\% \\ &= \mathbf{59,00\%} \end{aligned}$
---	--	---

Dari perhitungan matematis diatas, dapat diperoleh signifikansi nilai akurasi=**59,00**, presisi=**100,00%** dan juga Recall = **59,00%**. Hal ini sinkron dengan hasil akhir yang ada di tabel output nilai perbandingan pengujian yang dilakukan menggunakan 16 atribut data dibawah ini :

Tabel 3. Komparasi Data precision, Accuracy, dan Recall untuk Algoritma Naive Bayes Classifier

Hasil pengujian menggunakan 16 atribut		Hasil pengujian menggunakan 9 atribut data	
<i>Accuracy</i>	59,00%	<i>Accuracy</i>	56,00%
<i>Precision</i>	100,00%	<i>Precision</i>	94,00%
<i>Recall</i>	59,00%	<i>Recall</i>	57,00%

Jadi, dari tabel hasil perbandingan diatas dapat dibuktikan bahwa penggunaan 16 atribut lebih tinggi dibandingkan dengan penggunaan 9 atribut.

4. KESIMPULAN

Dengan dihasilkannya klasifikasi kelayakan penentuan pemberian kredit nasabah dengan jumlah 16 atribut menghasilkan persentase akurasi sebesar 59,00 % merupakan tingkat akurasi yang baik dibandingkan dengan hasil penelitian klasifikasi kelayakan pemberian kredit nasabah yang menggunakan 9 atribut yang menghasilkan signifikansi persentase akurasi sebesar 56,00 %, sehingga kelancaran proses penilaian kelayakan kredit dapat tercipta, dan pembayaran terlambat (menunggak atau macet) sudah terprediksikan dari awal untuk dapat diwaspadai agar dapat meminimalisir meningkatnya kredit macet. Tingkat akurasi algoritma Naive Bayes eksperimen ini sudah mencapai tingkat baik, sehingga dapat meningkatkan ketelitian dalam proses klasifikasi dan prediksi dengan cara menambahkan beberapa atribut dari history pembayaran kredit nasabah yang ada sehingga dihasilkan pola klasifikasi yang akurat.

5. SARAN

Untuk peneltian selanjutnya, pengembangan analisis solusi dikembangkan dengan metode klasifikasi yang lain seperti algoritma C4.5 dan yang lainnya untuk melihat komparasi (perbandingan) dalam hal kelebihan metode-metode klasifikasi yang ada. Untuk meningkatkan nilai akurasi, sebaiknya menggunakan data dalam jumlah yang besar yang banyak bersumber dari beberapa perusahaan Bank.

DAFTAR PUSTAKA

- [1] Susanto, S., Mulyani, E. D. S., & Nurhasanah, I. R. (2015). Penerapan Data Mining Classification Untuk Prediksi Perilaku Pola Pembelian Terhadap Waktu Transaksi Menggunakan Metode Naive Bayes. *Proceedings Konferensi Nasional Sistem dan Informatika (KNS&I)*.
- [2] Fachry Husaini., 2018, Algoritma Klasifikasi Naive Bayes Untuk menilai Kelayakan Kredit di Bank Mandiri Kredit Mikro, Program Studi Teknik Informatika, Universitas Muhammadiyah Jember, Jember.
- [3] Heri Hidayanto., 2016, Achmad Wahid Kurniawan., Klasifikasi Kelayakan Kredit Calon Debitur Bank Menggunakan Algoritma Decision Tree C4.5, Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang.
- [4] Menarianti, I. (2015). Klasifikasi data mining dalam menentukan pemberian kredit bagi nasabah koperasi. *Jurnal Ilmiah Teknosains*, 1(1/November).

- [5] Dina Rusdiana Sari., Data Mining Klasifikasi Dengan Menggunakan Algoritma C4.5 Pada Data Nasabah Kredit di Bank Kredit Desa (BKD) Kabupaten Rembang, Universitas Dian Nuswatoro, Semarang.
- [6] Azkiya, M., Mukid, M. A., & Ispriyanti, D. (2015). Klasifikasi Nasabah Kredit Bank “X” Di Provinsi Lampung Menggunakan Analisis Diskriminan Kernel. *Jurnal Gaussian*, 4(4), 937-946.
- [7] Rani, L. N. (2015). Klasifikasi Nasabah Menggunakan Algoritma C4. 5 Sebagai Dasar Pemberian Kredit. *Komputer Teknologi Informasi*, 2(2).