

OPTIMASI NAIVE BAYES MENGGUNAKAN OPTIMIZE WEIGHTS DAN STRATIFIED PADA DATA KREDIT KOPERASI

Ade Suryanto¹, Ibnu Alfarobi², Taransa Agasya Tutupoly³, Risma Fauziahti⁴

¹Teknik Industri

²Sistem Informasi Akuntansi

³Teknik Informatika

⁴Sistem Informasi

¹Universitas Bina Sarana Informatika, Jl. Kamal Raya No. 18, Cengkareng Jakarta,

²Universitas Bina Sarana Informatika, Jl. Kamal Raya No. 18, Cengkareng Jakarta,

³STMIKNusaMandiri, Jl. Kramat Raya No. 18 Jakarta,

⁴STMIKNusaMandiri, Jl. Kramat Raya No. 18 Jakarta

Email: surya.ade@gmail.com , ibnu.iba@bsi.ac.id, taransa.tutupoly@gmail.com , rismafzhti@gmail.com

Abstrak

Kredit koperasi adalah penyedia dana untuk transaksi pinjam meminjam atas persetujuan dan kesepakatan antara pihak koperasi dengan nasabahnya, serta mewajibkan peminjam untuk membayar hutang dalam jangka waktu yang telah ditentukan. Pembatasan kredit di koperasi belum menemukan cara yang paling sesuai, karena koperasi belum mempunyai analisis kredit yang handal seperti perbankan dan selama ini koperasi hanya melakukan pendekatan secara personal dan survei lapangan. Klasifikasi data mining dengan model Naive Bayes, Optimize Weights dan Stratified dilakukan dengan pengujian-pengujian yang terukur melalui uji AUC dan ROC dengan bantuan Rapidminer. Hasilnya setelah dilakukan pengujian dengan model Naive Bayes ternyata menghasilkan $accuracy= 66.27\%$, $precision= 66.45\%$, $recall= 94.39\%$ dan hasil pengujian yang dilakukan dengan model Naive Bayes, Optimize Weights dan Stratified menghasilkan $accuracy= 86.67\%$, $precision= 89.47\%$, $recall= 89.47\%$. artinya $accuracy$ pengujian dengan menggunakan model Naive Bayes, Optimize Weights dan Stratified masih baik dan dapat dijadikan salah satu pedoman untuk deteksi pemberian kredit pada koperasi. Hasil pengujian menggunakan model Naive Bayes bukan satu-satunya untuk deteksi pemberian kredit koperasi, melainkan masih banyak model klasifikasi data mining yang kemungkinan hasilnya akan berbeda.

Kata Kunci: Curve AUC, ROC, Naive Bayes, Optimize Weights, Stratified

1) Pendahuluan

Perkembangan koperasi dari tahun ke tahun telah menuju kearah yang lebih positif. Berdasarkan data pemerintah dalam empat tahun terakhir, dengan angka pertumbuhan koperasi aktif rata-rata sebesar 2,5% periode 2012 hingga 2016. Pada 5 Juli 2017, Indonesia memiliki 26,8 juta anggota koperasi dan 152.282 unit koperasi. Dan kini koperasi simpan pinjam 19.509 unit koperasi di seluruh Indonesia (Bappenas, 2017).

Koperasi merupakan salah satu lembaga keuangan yang melakukan transaksi kredit dalam bentuk pinjaman anggota. Setiap anggota koperasi yang memiliki pinjaman memiliki kewajiban untuk membayar angsuran pinjaman sesuai dengan jangka waktu yang telah dilakukan. Ketidakkampaun anggota koperasi membayar angsuran pinjaman dalam jangka waktu yang telah ditentukan dapat menyebabkan pinjaman macet. Kondisi ini apabila dibiarkan terus menerus dapat berpengaruh langsung terhadap likuidasi koperasi untuk meminimalisir resiko pinjaman macet koperasi harus

menerapkan prinsip 5C (Character, Capital, Capacity, Collateral, Condition of economy) dalam mempertimbangkan keputusan pemberian pinjaman kepada anggota koperasi.

Koperasi memiliki permasalahan seperti menghilangnya nasabah, ketidakpastian pembayaran dana pinjaman bahkan ketidakmampuan nasabah dalam mengembalikan pinjaman dana kredit. Untuk penyediaan dana transaksi pinjam meminjam atas persetujuan dan kesepakatan pihak koperasi atau instansi keuangan dengan nasabahnya, serta mewajibkan peminjam untuk membayar utang dalam jangka waktu tertentu. Pemberian kredit dilakukan dengan mengidentifikasi dan menilai faktor yang mempengaruhi resiko kredit. Hilangnya pendapatan dan ancaman profitabilitas merupakan hal yang di waspadai dalam pemberian kredit. Dalam kredit semua jenis pinjaman yang harus dibayar kembali bersama sesuai dengan perjanjian yang telah disepakati (Hasibun, 2014:87). Untuk melindungi dana kredit, digunakan jaminan dapat berupa : jaminan benda berwujud (tanah,

bangunan, kendaraan bermotor, kebun, perhiasan dan lain-lain), jaminan tidak berwujud (sertifikat tanah, sertifikat saham, sertifikat obligasi dan lain-lain) dan jaminan orang (jaminan yang diberikan oleh seorang yang menyatakan kesanggupan untuk menanggung segala resiko apabila kredit tersebut macet).

Kredit macet di koperasi simpan pinjam sudah menjadi hal yang sangat mengkhawatirkan bagi kesalahan koperasi itu sendiri, dimana semakin banyaknya kredit yang macet dari waktu ke waktu dan sulit untuk ditagihkan, sementara kesederhanaan administrasi koperasi belum mampu melindungi koperasi dari kegagalan menagih hutang ke nasabah atau menjual jaminan yang ada penanggulangan dan pembatasan kredit macet di koperasi belum menemukan cara yang paling sesuai karena koperasi belum mempunyai analis kredit yang handal seperti perbankan dan mahal biayanya, selama analisa kredit dilakukan dengan melakukan pendekatan personal dengan mengisi blangko pengajuan kredit seret survey lapangan (Sucipto,2015).

Klasifikasi Data Mining dapat digunakan untuk menemukan resiko kredit. Data Mining adalah kegiatan yang meliputi pengumpulan data untuk menemukan keteraturan, pola atau hubungan dalam set data yang berukuran besar (Santoso, 2017).

Keluaran yang dihasilkan Klasifikasi Data Mining dapat digunakan untuk memperbaiki pengambilan keputusan bagi analis kredit dalam pemberian kredit. Klasifikasi Data Mining dengan model Naïve Bayes dengan pengujian menggunakan parameter biodata nasabah dengan jumlah nasabah sebanyak 1312 ternyata menghasilkan akurasi recall = 78.00%, artinya akurasi pengujian dengan menggunakan naïve bayes masih baik dan dapat dijadikan salah satu pedoman untuk deteksi kredit macet sebelum ada keputusan member atau tidak calon nasabah baru (Sucipto,2015).

Penelitian ini dibuat dengan menggunakan teknik data mining, teknik ini dilakukan untuk meminimalisir masalah tersebut. Teknik Data Mining merupakan proses menentukan pola dan informasi dari data yang berjumlah besar (Han, et al, 2013:8). Proses di dalam Data Mining untuk membedakan kelas data atau konsep yang bertujuan agar dapat digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui dikenal sebagai klasifikasi. Pemilihan Naïve Bayes Klasifikasi Data Mining untuk menentukan resiko kredit yang terjadi pada transaksi peminjaman.

2) Teori

a. Data Mining

Data mining adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi didalam database. Data mining adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam database, data warehouse, atau penyimpanan informasi lainnya.

b. Data Mining Menurut Para Ahli

Cara pandang dan pengetahuan yang berbeda membuat para ahli memberikan definisi berbeda dengan data mining. Lalu Larose menyatakan bahwa data mining adalah bidang yang di gabungkan dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistic, database, dan visualisasi untuk pengenalan permasalahan pengambilan informasi dari database yang besar. Data mining merupakan gabungan sejumlah disiplin ilmu komputer, (Clifton, 2010), yang didefinisikan sebagai proses penemuan pola-pola baru dari kumpulan-kumpulan data sangat besar, meliputi metode-metode yang merupakan irisan dari artificial intelligence, machine learning, statistics dan database systems (ACM, 2006).

Data mining adalah teknik data yang relatif cepat dan mudah untuk menemukan pengetahuan, pola dan relasi antar data, secara otomatis. Dengan menggabungkan empat disiplin ilmu komputer seperti pada definisi diatas, pengetahuan bisa ditemukan dalam lima proses yang berurutan : data selection, data cleaning, data transformation, data mining dan data integration (Fayyad et al, 1996).

c. Tahapan-Tahapan Data Mining

Istilah data mining atau knowledge discovery in database yang disingkat KDD sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD data mining sebagai berikut:

1) Data Cleaning

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, memperbaiki kesalahan pada data, seperti kesalahan cetak. juga dilakukan proses memperkaya data yang sudah ada dengan data atau informasi yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

2) Data Integration

Pola informasi yang dihasilkan dari proses data mining, perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut data integration.. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

3) Data Selection

Seleksi data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan suatu berkas, terpisah dari basis data operasional.

4) Data Transformation

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

5) Data Mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

d. Klasifikasi

Klasifikasi adalah satu bentuk analisis data yang menghasilkan model untuk mendeskripsikan kelas data yang penting. Klasifikasi merupakan proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau class data.dengan bertujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui (Suryana,2015)

1) Naive Bayes

Naive Bayes adalah merupakan salah satu algoritma data mining dengan metode klasifikasi. Naive Bayes adalah pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksikan peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai teorema bayes. Teorema bayes digunakan untuk menghitung probabilitas terjadinya suatu peristiwa berdasarkan pengaruh yang didapat dari hasil observasi. Metode ini penting karena beberapa alasan, termasuk berikut. Hal ini sangat mudah untuk membangun, tidak perlu ada yang rumit parameter estimasi skema berulang. Ini berarti dapat segera diterapkan

untuk besar data. Sangat mudah untuk menafsirkan, sehingga pengguna tidak terampil dalam teknologi dan dapat di pahami menagapa itu adalah membuat klasifikasi. Dan sangat penting, hal itu sering sangat baik dalam setiap diberikan aplikasi, tetapi biasanya dapat diandalkan untuk menjadi kuat dan melakukan dengan sangat baik (Wu, 2013).

Klasifikasi naive bayes didasarkan pada teorema bayes, diambil dari nama seorang ahli matematika yang juga menteri Prebysterin Inggris, Thomas Bayes (1702-1761) yaitu:

$$\frac{P(x|y) = P(y|x)P(x)}{P(y)}$$

Keterangan:

y = data dengan kelas yang belum diketahui

x = hipotesis data y merupakan suatu kelas spesifik

$P(x|y)$ = probabilitas hipotesis x berdasarkan kondisi y (posteriori probability)

$P(x)$ = probabilitas hipotesis x (prior probality)

$P(y|x)$ = probabilitas y berdasarkan kondisi pada hipotesis x

$P(y)$ = probabilitas dari y

Keuntungan metode Naïve Bayes:

- Menangani kuantitatif dan data diskrit
- Kokoh untuk titik noise yang diisolasi, misalkan titik yang dirata-ratakan ketika mengestimasi peluang bersyarat.
- Hanya memerlukan sejumlah kecil data pelatihan untuk mengestimasi parameter (rata-rata dan variansi dari variabel) yang dibutuhkan untuk klasifikasi.
- Menangani nilai yang hilang dengan mengabaikan instansi selama perhitungan estimasi peluang
- Cepat dan efisiensi ruang
- Kokoh terhadap atribut yang tidak relevan

Kekurangan Naïve Bayes:

- Tidak berlaku jika probabilitas kondisionalnya adalah nol, apabila nol maka probabilitas prediksi akan bernilai nol juga
- Mengasumsikan variabel bebas. Teorema tersebut dikombinasikan dengan Naïve dimana diasumsikan kondisi antara atribut saling bebas.

Persamaan dari teorema bayes adalah :

$$\frac{P(C_i|X) = P(X|C_i)P(C_i)}{P(X)}$$

2) Rapidminer

Rapidminer adalah salah satu perangkat lunak untuk pengolahan data mining. Pekerjaan yang dilakukan oleh rapidminer adalah berkisar dengan analisis teks, mengekstrak pola-pola dari

set yang besar dan mengkombinasikannya dengan metode kecerdasan buatan, dan database. Tujuannya adalah untuk mendapatkan informasi bermutu tertinggi dari teks yang diolah (Suryana, 2015). Rapidminer menyediakan prosedur data mining di dalamnya termasuk : ETL (Etraction, transformation, loading), data preprocessing, visualisasi, modelling dan evaluasi.

3. Metodologi Penelitian

a. Metode Pengumpulan Data

Penentuan jenis dan sumber data untuk memperoleh data yang benar-benar akurat merupakan hal yang sangat penting. Sumber data dalam penelitian ini adalah data pemberian kredit yang diambil dari Pusat Koperasi Kredit Jakarta pada tahun sebelumnya sebagai acuan untuk menemukan pola-pola tertentu yang bisa dijadikan atribut penentu. Berdasarkan sumber datanya, maka pengumpulan data dapat menggunakan data primer dan data sekunder. Data primer adalah sumber data yang langsung memberikan data kepada pengumpul data dan data sekunder merupakan sumber yang tidak langsung memberikan data kepada pengumpul data (sugiyono, 2017). Dalam pengumpulan data penelitian ini di peroleh dari Pusat Koperasi Kredit Jakarta. Data ini berupa usia, jumlah pinjaman, jangka waktu pinjaman, angsuran, jenis pekerjaan, masa kerja/usaha, penghasilan, status rumah, lama tinggal, jumlah tanggungan, nilai agunan, tujuan pinjaman, simpanan, prestasi pengambilan kredit yang lalu, partisipasi pengembangan kredit, status kredit.

b. Pengolahan Data Awal

Proses pengolahan awal data diperlukan untuk menyiapkan data yang benar-benar valid sebelum diproses. Pengolahan dilakukan dengan membersihkan data yang ganda, menyamakan batasan data, pengelompokan data, melakukan seleksi fitur dan preprocessing data (Gorunescu, 2013).

1) Integrasi Data

Data yang dapat digunakan dalam proses penentuan kredit adalah data piutang lancar, data agunan dan data peminjaman. Integrasi data adalah cara menggabungkan beberapa data dari tabel yang berbeda dengan melihat keamanan data berdasarkan atribut kunci (primary key), atribut tamu (foreign key) hingga melihat ketergantungan fungsionalnya. Integrasi data diperlukan karena perlu dilakukan seleksi fitur untuk mendapatkan pola yang merujuk pada hasil pemberian kredit.

2) Seleksi Fitur

Seleksi fitur dilakukan dengan mengambil sebagian variabel pada seluruh atribut yang ada untuk dijadikan atribut penentu dalam melakukan pemberian keputusan. Fitur yang diambil adalah atribut yang memiliki sifat ketergantungan fungsional dan merupakan bagian dari super key.

3) Data Cleansing

Proses cleansing merupakan tahapan yang penting, dimana data dibersihkan dari data yang tidak diperlukan (seperti: no.anggota, nama, alamat) dan menghapus data yang sama. Hal ini diusulkan untuk menjaga nilai ketergantungan fungsionalnya.

4) Transformasi Data

Proses transformasi data dikelompokkan berdasarkan kriteria yang mana untuk mempermudah pengolahan data selanjutnya.

c. Populasi dan sampel

Menurut Sugiyono (2017) populasi adalah wilayah generalisasi yang terdiri atas objek atau subjek yang mempunyai kualitas dan karakteristik tertentu yang ditetapkan oleh peneliti untuk dipelajari dan kemudian ditarik kesimpulannya. Sedangkan sampel adalah bagian dari jumlah dan karakteristik yang dimiliki oleh populasi tersebut. Populasi dalam penelitian ini adalah seluruh koperasi yang ada di wilayah DKI Jakarta. Adapun jumlah koperasi yang dapat diidentifikasi berdasarkan data dari Pusat Koperasi Kredit Jakarta.

Berdasarkan data set pada koperasi terdapat 567 data pemberian kredit dengan total data yang lancar sebanyak 361 data dan 206 data yang macet dan harus segera membayar kredit. Sedangkan sampel, peneliti menggunakan Rapidminer 7. Maka peneliti mengambil 10 data sebagai sampel dari populasi yang termasuk data training dan 1 data sebagai data testing yang ada di koperasi tersebut.

Tabel 1. Data Training

Umur	Jenis Pekerjaan	Jumlah Pinjam	Lama Pinjam	Status Kredit
44	Karyawan Ibu Rumah	50.000.000	60	LANCAR
41	Tangga	20.000.000	36	LANCAR
30	Karyawan	17.850.000	24	LANCAR
48	Dosen	60.000.000	48	LANCAR
41	Karyawan	185.000.000	60	MACET
49	Wiraswasta Ibu Rumah	15.000.000	15	MACET
42	Tangga	3.000.000	10	LANCAR
41	Guru	2.000.000	24	LANCAR
32	Mahasiswa	2.000.000	6	LANCAR

Umur	Jenis Pekerjaan	Jumlah Pinjam	Lama Pinjam	Status Kredit
44	Wiraswasta	4.000.000	16	LANCAR
35	Karyawan	25.000.000	30	?

Sumber: Koperasi Pusat Jakarta

Cara perhitungan Naive Bayes pada Tabel 1 untuk record paling akhir (umur 35 tahun) adalah sebagai berikut:

$$P=(Y=Lancar) = 8/10 = 0.8$$

$$P=(Y=Macet) = 2/10 = 0.2$$

Mencari nilai probabilitas pada tiap atribut

$$P=(umur= 35|Y=Lancar) = 2/8 = 0.25$$

$$P=(jenis pekerjaan= karyawan|Y=Lancar) = 3/8 = 0.375$$

$$P=(jumlah pinjam= 25.000.000|Y=Lancar) = 1/8 = 0.125$$

$$P=(lama pinjam= 30|Y=Lancar) = 3/8 = 0.375$$

$$P=(umur= 35|Y=Macet) = 0/2 = 0$$

$$P=(jenis pekerjaan= karyawan|Y= Macet) = 0/2 = 0$$

$$P=(jumlah pinjam= 25.000.000|Y=Macet) = 0/2 = 0$$

$$P=(lama pinjam= 30|Y= Macet) = 2/2 = 1$$

Menghitung hasil atribut Lancar dan Macet

$$P(status kredit=Lancar)*P(Y=Lancar)= 0.9$$

$$P(status kredit=Macet)*P(Y=Macet)=0.2$$

Jadi hasil yang diperoleh dan perhitungan menunjukkan bahwa pada tabel menghasilkan (P= Lancar) dengan probabilitas tertinggi.

d. Metode Analisis Data

Metode yang diusulkan dalam penelitian ini adalah klasifikasi data mining yang digunakan adalah Naive Bayes. Hal ini dilakukan untuk melihat sejauh mana perbedaan data setelah dan sebelum dilakukan processing data. Untuk menentukan klasifikasi yang digunakan pada suatu masalah diperlukan cara sistematis untuk mengevaluasi bagaimana metode yang bekerja dan membandingkan dengan yang lain. Evaluasi klasifikasi didasarkan pada objek yang benar dan salah (Gorunescu, 2013).

4. Hasil Dan Pembahasan

a. Hasil

Hasil dalam penelitian ini menggunakan sebanyak 567 data koperasi yang terdiri dari 16 atribut. Sebelum pembentukan metode Naive Bayes untuk klasifikasi kredit koperasi terlebih

dahulu dilakukan tahap preprocessing data yang terdiri dari integrasi data, seleksi fitur, dan transformasi data.

Integrasi data dengan penggantian atribut No.KTP dan tanggal peminjaman menjadi atribut Umur. Selain itu integrasi data juga dilakukan pada awal ketika pemilihan data-data dengan data peminjaman menggunakan primary key. Setelah integrasi data, dilanjutkan dengan tahap seleksi

Fitur dan tranformasi data. Seleksi fitur yang dilakukan adalah penghapusan atribut No.anggota dan penghapusan atribut-atribut yang tidak digunakan dalam analisis. Transformasi data dilakukan dengan mengubah tipe data numerik menjadi interval dan menginisialkan nilai atau karakter terlalu panjang pada beberapa atribut. Hasil transformasi data yang dilakukan dapat dilihat pada Tabel 2.

Tabel 2. Transformasi Data

Atribut	Kategori	Transformasi
Umur(th)	<30	U1
	30-39	U2
	40-49	U3
	50-59	U4
	>60	U5
Penghasilan (rupiah)	0-1000000	P1
	1000001-2000000	P2
	2000001-3000000	P3
	3000001-4000000	P4
	4000001-5000000	P5
	>5000000	P6
Pinjaman (rupiah)	0-2000000	PJ1
	2000001-4000000	PJ2
	4000001-6000000	PJ3
	6000001-8000000	PJ4
	>8000000	PJ5
Jangka Waktu (bulan)	1-24	J1
	25-48	J2
	>48	J3
Nilai Agunan	0-5000000	N1
	5000001-10000000	N2
	10000001-15000000	N3
	15000001-20000000	N4
	>20000000	N5
Jenis Pekerjaan	Mahasiswa	JP1
	Ibu Rumah Tangga	JP2
	Dosen	JP3
	Guru	JP4

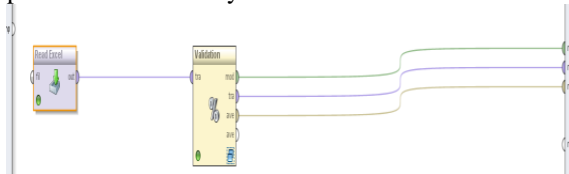
Atribut	Kategori	Transformasi
Status Rumah	Karyawan	JP5
	Wiraswasta	JP6
	PNS	JP7
	Pensiun	JP8
	Milik Sendiri	ST1
	Sewa/Kontrak	ST2
	Keluarga	ST3
	KPR	ST4
Tujuan	Modal Usaha	T1
	Renovasi Rumah	T2
	Pembelian Barang	T3
	Pembelian Kendaraan	T4
	Ibadah Keagamaan	T5
	Pendidikan	T6
	Pengobatan	T7
Status Kredit	Lancar	SK1
	Macet	SK2

Sumber: Koperasi Pusat Jakarta

Pada tabel diatas dapat di lihat bahwa hasil transformasi data seperti umur menjadi U1 agar karakter dapat membedakan atribut satu dengan atribut lain. Setelah transformasi data, selanjutnya mengetahui berapa banyak nilai accuracy, nilai precision, dan nilai recall dengan menggunakan Rapidminer.

1) Model Pengujian Pada Naive Bayes

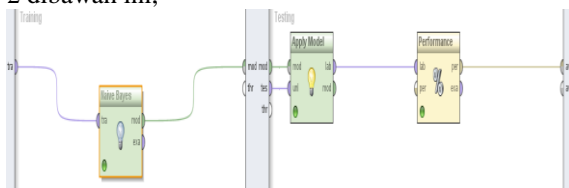
Dalam penelitian ini menggunakan dua pengujian yang pertama menggunakan naive bayes dan yang kedua menggunakan model optimize weights dan sample stratified. Berikut pada gambar dibawah ini ada pengujian pertama pada model naive bayes.



Sumber: Data Kredit Koperasi

Gambar 1. Model Split Validation pada Naive Bayes

Pada gambar 1 diatas data pemberian kredit dihubungkan dengan operator split validation yang di dalamnya terdapat proses seperti gambar 2 dibawah ini,



Sumber: Data Kredit Koperasi

Gambar 2. Pengujian *Split Validation* pada *Naive Bayes*

Pada gambar diatas adalah proses pengujian split validation yang didalamnya terdapat naive bayes yang terhubung dengan apply model dan performance classification, dimana training dan testing yang digunakan akan berpengaruh terhadap accuracy dan model yang dihasilkan.

Pengujian pada metode naive bayes terdapat data pemberian kredit yaitu 178 dari jumlah 567. Berikut ini merupakan hasil dari klasifikasi menggunakan Rapidminer 7:

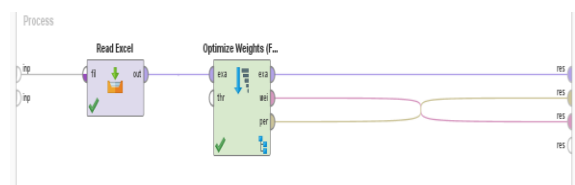
accuracy: 66.27%			
	true LANCAR	true MACET	class precision
pred LANCAR	101	51	66.45%
pred MACET	6	11	64.71%
class recall	94.39%	17.74%	

Sumber: Klasifikasi dengan Rapidminer

Gambar 3. Accuracy Naive Bayes

Pada gambar diatas dapat dilihat terdapat 178 data, 101 diklasifikasikan prediksi lancar ternyata lancar, 51 data diprediksikan lancar ternyata macet. Untuk 6 data diklasifikasikan prediksi macet ternyata lancar, serta 11 data diprediksikan sesuai yaitu macet. Dan nilai precision pada lancar sebanyak 66.45%, dan macet sebanyak 64.71%, nilai recall pada lancar sebanyak 94.39% dan macet sebanyak 17.74%, dan nilai accuracy pada metode naive bayes sebanyak 66.27%.

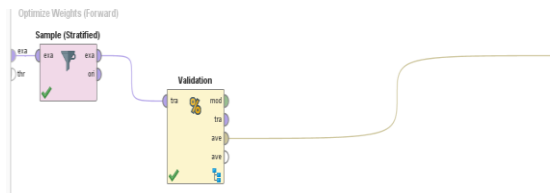
2) Model Pengujian Optimize Weights Dan Sample Stratified Pada Naive Bayes



Sumber: Data Kredit Koperasi

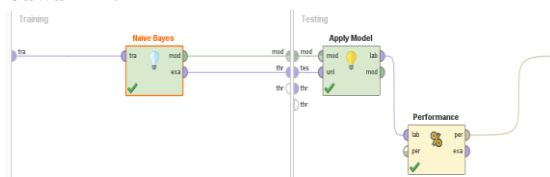
Gambar 4. Proses Optimize Weights

Pada gambar 4 diatas data pemberian kredit dihubungkan dengan operator optimize yang di dalamnya terdapat proses seperti gambar 5 dibawah ini,



Sumber: Data Kredit Koperasi
Gambar 5. Pengujian Optimize Weights pada Naive Bayes

Berdasarkan gambar 5 diatas adalah pengujian menggunakan optimize weights yang di dalamnya terdapat proses sample stratified dan split validation. Pada split validation didalamnya terdapat proses seperti gambar 6 dibawah ini.



Sumber: Data Kredit Koperasi
Gambar 6. Pengujian Split Validation pada Naive Bayes

Pada gambar diatas adalah proses pengujian split validation yang didalamnya terdapat naive bayes yang terhubung dengan apply model dan performance clasification, dimana training dan testing yang digunakan akan berpengaruh terhadap accuracy dan model yang dihasilkan.

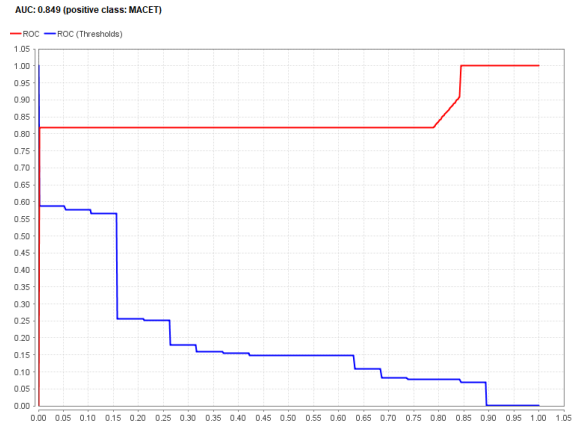
Pengujian pada metode naive bayes terdapat data pemberian kredit yaitu 22 dari jumlah 567. Berikut ini merupakan hasil dari klasifikasi menggunakan Rapidminer 7.

accuracy: 86.67%

	true LANCAR	true MACET	class precision
pred. LANCAR	17	2	89.47%
pred. MACET	2	9	81.82%
class recall	89.47%	81.82%	

Sumber: Klasifikasi dengan Rapidminer
Gambar 7. Accuracy Naive Bayes dengan Model Optimize Weights

Pada gambar diatas dapat dilihat terdapat 22 data, 17 diklasifikasikan prediksi lancar ternyata lancar, 2 data diprediksikan lancar ternyata macet. Untuk 2 data diklasifikasikan prediksi macet ternyata lancar, serta 9 data diprediksikan sesuai yaitu macet. Dan nilai precision pada lancar sebanyak 89.47%, dan macet sebanyak 81.82%, nilai recall pada lancar sebanyak 89.47% dan macet sebanyak 81.82%, dan nilai accuracy pada metode naive bayes sebanyak 86.67%.



Sumber: Data Kredit Koperasi
Gambar 8. Grafik AUC Dan ROC Naive Bayes

Dan pada Gambar 8 adalah Kurva ROC dengan metode Naive Bayes dan mendapatkan hasil AUC sebesar 0.849 (positive class: Macet).

b. Pembahasan

1) Model Confusion Matrix Naive Bayes

Hasil dari pengujian model menggunakan metode naive bayes yang dilakukan adalah memprediksi kredit koperasi dengan untuk menentukan nilai accuracy, nilai precision, dan nilai recall dengan benar pada persamaan dibawah ini:

Tabel 3. Model Confusion Matrix Naive Bayes

accuracy: 66.27%

	true LANCAR	true MACET	class precision
pred LANCAR	101	51	66.45%
pred MACET	6	11	64.71%
class recall	94.39%	17.74%	

Sumber: Data Kredit Koperasi

$$\begin{aligned}
 TP(\text{True LANCAR}) &= 101 \\
 FP(\text{False LANCAR}) &= 6 \\
 TN(\text{True MACET}) &= 11 \\
 FN(\text{False MACET}) &= 51 \\
 \text{Accuracy} &= \\
 (TP+TN)/(TP+TN+FP+FN) &= \\
 (101+11)/(101+11+6+51) &= 112/169 \\
 &= 0.6627 (66.27\%) \\
 \text{Precision} &= TP/(FN+TP) \\
 &= 101/(51+101) \\
 &= 101/152
 \end{aligned}$$

$$= 0.66448 (66.45\%)$$

$$\begin{aligned} \text{Recall} &= \text{TP}/(\text{FP}+\text{TP}) \\ &= 101/(6+101) \\ &= 101/107 \\ &= 0.9439 (94.39\%) \end{aligned}$$

$$\begin{aligned} \text{Recall} &= \text{TP}/(\text{FP}+\text{TP}) \\ &= 17/(2+17) \\ &= 17/19 \\ &= 0.8947 (89.47\%) \end{aligned}$$

Hasil perhitungan terlihat pada tabel 4 dibawah ini.

Tabel 4. Nilai Accuracy, Precision dan Recall Naive Bayes

	Nilai (%)
Accuracy	66.27%
Precision	66.45%
Recall	94.39%

Sumber: Data Kredit Koperasi

2) Model Confusion Matrix Naive Bayes Optimize weight dan Stratified

Hasil dari pengujian model menggunakan metode naive bayes dengan optimize weights dan sample stratified yang dilakukan adalah memprediksi kredit koperasi dengan untuk menentukan nilai accuracy, nilai precision, dan nilai recall dengan benar pada persamaan dibawah ini:

Tabel 5. Model Confusion Matrix Naive Bayes Optimize weight Dan Stratified

accuracy: 86.67%			
	true LANCAR	true MACET	class precision
pred LANCAR	17	2	89.47%
pred MACET	2	9	81.82%
class recall	89.47%	81.82%	

Sumber: Data Kredit Koperasi

$$\begin{aligned} \text{TP(True LANCAR)} &= 17 \\ \text{FP(False LANCAR)} &= 2 \\ \text{TN(True MACET)} &= 2 \\ \text{FN(False MACET)} &= 9 \\ \text{Accuracy} &= \\ (\text{TP}+\text{TN})/(\text{TP}+\text{TN}+\text{FP}+\text{FN}) &= \\ &= (17+2)/(17+2+9+9) \\ &= 26/30 \\ &= 0.8667 (86.67\%) \\ \text{Precision} &= \text{TP}/(\text{FN}+\text{TP}) \\ &= 17/(2+17) \\ &= 17/19 \end{aligned}$$

Hasil perhitungan terlihat pada tabel 6 di bawah ini.

Tabel 6. Nilai Accuracy, Precision dan Recall Naive Bayes Optimize weight dan Stratified

	Nilai (%)
Accuracy	66.27%
Precision	66.45%
Recall	94.39%

Sumber: Data Kredit Koperasi

3) Hasil Perhitungan Pengujian

Dari hasil perhitungan pengujian diatas, baik menggunakan model confusion matrix naive bayes dan model confusion matrix naive bayes dengan optimize weights dan sample stratified. Terbukti bahwa hasil pengujian naive bayes dengan optimize weights dan sample stratified memiliki nilai accuracy yang lebih tinggi dibandingkan dengan naive bayes saja. Nilai accuracy untuk model confusion matrix naive bayes sebesar 66.27%, dan nilai accuracy untuk model confusion matrix naive bayes dengan optimize weights dan sample stratified sebesar 86.67% dengan selisih accuracy sebesar 20.40%, dapat dilihat pada tabel 7 di bawah ini:

Tabel 7. Hasil Perhitungan Pengujian

	Accuracy	AUC
NB	66.27%	
NB+OW+SS	86.67%	0.849

Sumber: Data Kredit Koperasi

Berdasarkan hasil pengujian yang dilakukan untuk memecahkan masalah hasil prediksi lancar dan macet, dapat disimpulkan bahwa hasil pengujian menggunakan optimize weights dan sample stratified pada metode Naive bayes saja mempunyai tingkat nilai accuracy sebesar 66.27% , dan hasil pengujian menggunakan optimize weights dan sample stratified pada metode naive bayes mempunyai tingkat nilai

accuracy sebesar 86.67% dan nilai AUC sebesar 0.849.

5. Kesimpulan

Hasil penelitian yang dilakukan dengan menggunakan klasifikasi data mining menggunakan metode naive bayes dan naive bayes dengan optimize weights dan sample stratified dengan Rapidminer 7. Menghasilkan 14 atribut pada masing-masing metode. Dan untuk deteksi status kredit pada metode naive bayes menghasilkan nilai accuracy sebesar 86.67%, nilai recall dan nilai precision pada lancar dan macet sebesar 89.47% dan 81.82%. Dan untuk deteksi status kredit pada metode naive bayes dengan optimize weights dan sample stratified menghasilkan nilai accuracy sebesar 66.27%, nilai recall pada lancar 94.39% dan pada macet sebesar 17.74%, nilai precision pada lancar sebesar 66.45% dan pada macet sebesar 64.71%. Untuk pengujian atau mencari nilai probabilitas pada tiap atribut dengan metode naive bayes menghasilkan status kredit pada lancar sebesar 0.9 dan status kredit pada macet sebesar 0.2, jadi hasil yang diperoleh dalam pengujian ini menunjukkan bahwa status kredit pada lancar dengan probabilitas tertinggi.

6. Referensi

- [1] Aryanto, Kadek. 2017. Implementasi Metode C4.5 Dan Naive Bayes Berbasis Adaboost Untuk Memprediksi Kelayakan Pemberian Kredit.
- [2] International Journal of Natural Science and Engineering Ciptohartono, Claudia. 2014. Algoritma Klasifikasi Naïve Bayes Untuk Menilai Kelayakan Kredit
- [3] Gufroni, Acep. Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia. IJNS Vol. 03 No. 02 (2017) 299-305. ISSN: 2476-8812
- [4] Karyadiputra, Erfan. Analisis Algoritma Naive Bayes Untuk Klasifikasi Status Kesejahteraan Rumah Tangga Keluarga Binaan Sosial. "Technologia" Vol 7, No.4, Oktober – Desember 2016
- [5] Kriestanto, Danny. Penerapan Naïve Bayes Untuk Prediksi Kelayakan Kredit. IJNS – Vol. 1, No. 1, Februari 2016
- [6] Leidiyana, Henny. Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor. 1(1) : 65-76 (2013)
- [7] Masripah, Siti. Komparasi Algoritma Klasifikasi Data Mining Untuk Evaluasi Pemberian Kredit. Vol.3, No.1, Juni 2016, 187 – 193. ISSN: 2527-9777
- [7] Menarianti, Ika. Klasifikasi Data Mining Dalam Menentukan Pemberian Kredit Bagi Nasabah Koperasi. IJNS Vol. 1 No. 1 November 2015. ISSN: 2476-9436
- [8] Insani, Nur. 2015. Analisis Klasifikasi Pada Nasabah Kredit Koperasi X Menggunakan Decision Tree C4.5 Dan Naïve Bayes
- [9] Sucipto, Adi. Prediksi Kredit Macet Melalui Perilaku Nasabah Pada Koperasi Simpan Pinjam Dengan Menggunakan Metode Algoritma Klasifikasi C4.5. IJNS Volume 6 no. 1, Januari 2015
- [10] Suryana, Eko. Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu). IJNS Vol. 11 No. 2, September 2015

PROFIL PENULIS

Ade Suryanto, M.Kom. Menyelesaikan Pendidikan Strata Dua(S2) di STMIK Nusa Mandiri Jakarta, lulus pada tahun 2015. Penulis bergabung di Universitas Bina Sarana Informatika sejak tahun 2006 hingga sekarang sebagai Dosen Tetap dan sudah memiliki Jabatan Fungsional Akademik.

Ibnu Alfarobi, M.Kom. Lahir di Brebes, 1 Juli 1989, lulus Program Sarjana di STMIK Nusa Mandiri pada tahun 2015 dan Program Pasca Sarjana di STMIK Nusa Mandiri pada tahun 2017. Saat ini masih aktif sebagai Dosen tetap di Universitas Bina Sarana Informatika.

Taransa Agasya Tutupoly, M.Kom. Lahir di Bogor, 25 Agustus 1989. Lulusan S2 Ilmu Komputer STMIK Nusa Mandiri. Saat ini aktif mengajar sebagai Dosen di STMIK Nusa Mandiri Jakarta.

Risma Fauziahti, S.Kom. Lahir di Jakarta, 16 September 1996. Lulus Program Sarjana di STMIK Nusa Mandiri Program Studi Sistem Informasi pada tahun 2018. Saat ini masih terdaftar sebagai karyawan di Biznet.