

Optimalisasi Data Tidak Seimbang Pada Data Nasabah Koperasi dalam Pemberian Pinjaman Menggunakan *Random Oversampling*

¹Richky Faizal Amir,²Andreyestha, ³Imam Nawawi, ⁴Rino Ramadhan

Fakultas Teknik & Informatika, Universitas Bin Sarana Informatika^{1,2,3}

Jl. Kramat Raya No.98, RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat, DKI Jakarta 10450^{1,2,3}

Alamat email richky.cfj@bsi.ac.id¹ · andreyes2505@bsi.ac.id² · imam.imw@bsi.ac.id³ · rino.rim@bsi.ac.id⁴

Abstract - The conditions set by a Cooperative for members who will make loans are one of the considerations for credit cooperatives to analyze the prospects for proposals whether they are feasible or not. To deal with this, it is necessary to hold a customer classification. Data mining has several classification algorithms that can be used for credit analysis [5], including the Random Forest and Algorithm C4.5. Prospective customer data received from cooperatives as a condition for applying for credit is processed using Random Forest data mining and C4.5 Algorithm to support credit analysis in order to obtain accurate information whether prospects applying for credit are feasible or not, this research was conducted to classify loans to prospective customers cooperatives using the Random Forest method and the C4.5 Algorithm which is optimized with Random Oversampling because the dataset is in an unbalanced condition. In testing the C4.5 Algorithm which was optimized with Random Oversampling, it obtained an accuracy of 78.03%, where the accuracy increased by 7.89% from the previous 70.14%. While the Random Forest with Random Oversampling has an accuracy value of 87.12%, an increase of 23.69% from the previous Random Forest test of 63.43%.

Keyword: Imbalanced Class, Random Oversampling, C4.5, Random Forest, Cooperative.

Abstract – Syarat-syarat yang ditetapkan oleh sebuah Koperasi bagi anggota yang akan melakukan peminjaman menjadi salah satu pertimbangan bagi koperasi kredit untuk menganalisa prospekajuan apakah layak atau tidak. Untuk menangani hal tersebut, maka butuh diadakan klasifikasi nasabah. Data mining memiliki beberapa algoritma klasifikasi yang dapat digunakan untuk analisis kredit[5], diantaranya terdapat Random Forest dan Algoritma C4.5. Data calon nasabah yang diterima dari koperasi sebagai syarat pengajuan kredit diolah menggunakan data mining Random Forest dan C4.5 Algorithm untuk mendukung analisis kredit guna mendapatkan informasi yang akurat apakah prospek yang mengajukan kredit layak dilakukan atau tidak, penelitian ini dilakukan untuk mengklasifikasikan pinjaman kepada calon nasabah koperasi dengan menggunakan metode Random Forest dan C4.5 Algorithm yang dioptimalkan dengan Random Oversampling karena dataset dalam kondisi tidak seimbang. Pada pengujian Algoritma C4.5 yang dioptimalkan dengan Random Oversampling mendapat akurasi 78.03%, dimana akurasi meningkat sebesar 7.89% dari sebelumnya 70.14%. Sedangkan Random Forest dengan Random Oversampling memiliki nilai akurasi sebesar 87.12% meningkat 23.69% dari pengujian Random Forest sebelumnya sebesar 63.43 %.

Keyword: Imbalanced Class, Random Oversampling, C4.5, Random Forest, Koperasi.

I. PENDAHULUAN

Badan usaha adalah entitas yang terkait dengan bidang usaha tertentu dan dicakup oleh istilah hukum, khusus, dan keuangan. Agen ekonomi juga dapat dicirikan sebagai entitas hukum dan moneter yang diorganisir untuk mengawasi unsur-unsur penciptaan untuk menghasilkan tenaga kerja dan produk dengan tujuan menghasilkan keuntungan (utilitas). Untuk itu penulis mengkaji lebih dalam setiap bagian dari perseroan terbatas (PT) khususnya yang berkaitan dengan saham dan koperasi, khususnya modal sertifikat koperasi [1].

Koperasi telah berkembang dari waktu ke waktu. Pengertian awal secara umum adalah bahwa koperasi merupakan wadah bagi kelompok-kelompok ekonomi yang rentan, dan bahwa koperasi tidak mementingkan diri sendiri, yang memungkinkan masing-masing untuk memenuhi kewajibannya sebagai anggota dan menerima remunerasi yang proporsional. sebuah organisasi yang bertujuan untuk selalu berusaha dengan semangat Penggunaan organisasi. Selain penggambaran unsur “untuk 4.444 golongan ekonomi lemah”, ada unsur kooperatif, apatis dan demokratis yang bersumber dari pernyataan bahwa anggota akan diberi imbalan atas jasanya sesuai dengan jasa dan partisipasinya [2].

Dalam memberikan pelayanan, koperasi kredit tentunya memiliki syarat-syarat tertentu sebagai calon nasabah untuk menerima pinjaman. Koperasi perlu memeriksa apakah pihak yang berkepentingan akan menerima pinjaman. Pinjaman kepada nasabah merupakan sumber pendapatan utama bagi koperasi. Selain itu, kemungkinan kredit macet sangat berbahaya dan dapat mengurangi pendapatan koperasi yang mengeluarkan pinjaman. Masalah lainnya adalah tidak adanya prosedur yang sistematis untuk menentukan proyeksi risiko kredit dari prospek tersebut dan ketidakmampuan koperasi untuk menentukan apakah tepat untuk memberikan kredit kepada prospek Sulit untuk memprediksi risiko kredit yang

muncul melalui kepada pelanggan potensial. Koperasi perlu mengantisipasi potensi kredit macet dan selektif dalam memilih calon nasabah untuk menghadapi risiko kemacetan dan berkurangnya pendapatan. Dalam masyarakat saat ini, banyak usaha kecil memulai dan mendanai bisnis mereka secara mandiri. Namun, ketika mengembangkan usaha, mereka sering terkendala dengan pembiayaan (modal) [3].

Berbeda dengan bank, pengajuan kredit atau pinjaman ke koperasi biasanya mudah dan tetap memenuhi syarat pokok pinjaman. Pemeriksaan berbagai persyaratan seperti jumlah pinjaman, jangka waktu pembayaran, industri/jenis pekerjaan, jumlah tanggungan, status perkawinan, jumlah anggota keluarga, tanggungan, pendapatan, pajak, pembayaran tagihan listrik, telepon, dll. Formulir aplikasi ini biasanya digunakan untuk berkas permohonan pinjaman diajukan oleh calon pembeli yang mengisi formulir yang disediakan, selebihnya berdasarkan hasil analisa tim yang ditentukan oleh koperasi. Dalam beberapa kasus, prospek dengan persyaratan yang berbeda dapat menetapkan batas pendanaan pinjaman yang berbeda untuk mengurangi terjadinya kredit macet. Oleh karena itu, kita membutuhkan cara untuk memilih prospek [4], Data mining adalah analisis pemeriksaan kumpulan data untuk menemukan hubungan yang tidak terduga dan meringkas data dengan cara yang sebelumnya tidak diketahui, sehingga dapat dimengerti dan berguna bagi pemilik data. Data mining adalah bidang interdisipliner yang menggabungkan pembelajaran mesin, pengenalan pola, statistik, database, dan teknik visualisasi untuk memecahkan masalah pengambilan informasi dari database besar [5].

Dalam data mining, ada beberapa algoritma klasifikasi yang dapat digunakan untuk analisis kredit[5], diantaranya terdapat Random Forest dan Algoritma C4.5. Data calon nasabah yang diterima dari koperasi sebagai syarat pengajuan kredit diolah menggunakan data mining Random Forest dan C4.5 Algorithm untuk mendukung analisis kredit guna mendapatkan informasi yang akurat apakah prospek yang mengajukan kredit layak dilakukan. atau tidak, penelitian ini dilakukan untuk mengklasifikasikan pinjaman kepada calon nasabah koperasi dengan menggunakan metode Random Forest dan C4.5 Algorithm.

II. LANDASAN TEORI

Pada landasan teori akan dijelaskan mengenai materi yang berhubungan dengan penelitian yang dilakukan, serta metode-metode yang digunakan dalam penelitian.

1. Data Mining

Data mining banyak digunakan dalam bidang teknik dan sains, sebaliknya digunakan dalam aplikasi keuangan dan pemasaran untuk menyelesaikan tantangan di bidang masing-masing. Sistem pendukung keputusan berbasis data mining meningkatkan kinerja organisasi dengan menganalisis kenyataan di lapangan [6].

Data mining merupakan proses iteratif dan iteratif untuk menemukan pola atau model baru yang valid (sempurna), berguna dan dapat dimengerti dalam database yang sangat besar (large databases). Penambangan data melibatkan pencarian basis data besar untuk tren atau pola minat untuk membuat keputusan di masa depan. Pola-pola ini dikenali oleh alat khusus yang dapat memberikan analisis data yang berguna dan berwawasan luas, dan kemudian dapat dieksplorasi lebih menyeluruh, mungkin menggunakan alat pendukung keputusan lainnya [7].

Data mining adalah analisis otomatis dari sejumlah besar data yang digunakan untuk menemukan beberapa pola atau kecenderungan penting yang ketersediaannya biasanya diabaikan. Istilah penambangan data dan penemuan pengetahuan dalam basis data (KDD) biasanya digunakan secara bergantian untuk menggambarkan proses penggalian informasi tersembunyi dalam basis data besar. Padahal, kedua istilah tersebut memiliki konsep yang berbeda, namun terkait satu sama lain. Dan salah satu langkah dari keseluruhan proses KDD adalah data mining [8].

2. Klasifikasi

Proses menemukan model (atau fungsi) yang menggambarkan dan membedakan antara kelas data atau konsep yang digunakan untuk memprediksi kelas suatu objek yang penunjukan kelasnya tidak diketahui. Algoritme klasifikasi yang umum digunakan meliputi pohon keputusan/klasifikasi, pengklasifikasi Bayesian/Naive Bayes, jaringan saraf, analisis statistik, algoritme genetika, kumpulan kasar, k-tetangga terdekat, metode berbasis aturan, inferensi berbasis memori, dan Support Vector Machine (SVM) [9].

3. Random Forest

Dalam algoritma ini, sejumlah besar pohon keputusan dibangun saat mereka beroperasi bersama. Pohon keputusan bertindak sebagai pilar dalam algoritma ini. Random Forest didefinisikan sebagai kelompok pohon keputusan yang simpulnya ditentukan pada langkah pra-pemrosesan. Setelah membangun beberapa pohon, fitur terbaik dipilih dari subset fitur secara acak [10]. Random Forest umumnya menggunakan dua tingkat acak untuk memastikan keragaman pohon tunggal: (a) kumpulan data pelatihan yang berbeda dengan ukuran sampel yang sama, dan (b) kumpulan variabel penjelas yang berbeda untuk membagi setiap simpul [11] Rumus matematika untuk pengklasifikasi Random Forest adalah [10] :

$$n_{ij} = wC_j - w_{left(j)}C_{left(j)} - w_{right(j)}C_{right(j)}$$

$$n_{i \text{ sub}(j)} = \text{pentingnya simpul } j$$

$$w_{\text{ sub}(j)} = \text{jumlah tertimbang sampel yang mencapai simpul } j$$

$$C_{\text{ sub}(j)} = \text{nilai pengotor dari simpul } j$$

$$\text{left}(j) = \text{simpul anak dari pemisahan kiri pada simpul } j$$

$$\text{right}(j) = \text{simpul anak dari belahan kanan pada simpul } j.$$

4. Algoritma C4.5

Algoritma C4.5 adalah sekumpulan algoritma untuk masalah klasifikasi dalam pembelajaran mesin dan penambahan data. Algoritma C4.5 merupakan evolusi dari algoritma ID3. Algoritma C4.5 dan ID3 dikembangkan pada akhir 1970-an oleh seorang peneliti kecerdasan buatan bernama J. Rose Quinlan. Algoritma C4.5 membangun pohon keputusan top-down dimana atribut atas adalah root dan atribut bawah disebut daun [12].

Rumus algoritma C4.5 berupa pohon keputusan untuk mengetahui nilai entropi adalah sebagai berikut:

$$H(S) = - \sum p_i \log_2 p_i$$

Dengan :

S = Tetapkan Kasus

J = Fitur

N = Jumlah Partisi S

Pi = Proporsi Si Ke S

$$H(S, A) = H(S) - \sum_{i=1}^N p_i H(S_i)$$

Dengan : S = Tetapkan Kasus

J = Atribut

N = Jumlah Partisi Atribut A

[Si] = Jumlah Kasus Di Partisi Ke – I

|S| = Jumlah Kasus

5. Random Oversampling

Teknik pengambilan sampel memanipulasi data pelatihan untuk memperbaiki distribusi kelas miring, di antaranya SMOTE, borderline-SMOTE, ADASYN adalah metode yang paling representatif. Namun, kelemahan dari metode ini adalah pengaturan parameter dan model. Random Oversampling diperkenalkan, sebagai metode pengambilan sampel ulang untuk menangani kumpulan data yang tidak seimbang. Ada karena Random Oversampling menghasilkan sampel buatan yang cocok dari kedua kelas karena menggunakan pendekatan bootstrap yang dihaluskan. Selain itu, ROSE dapat melakukan estimasi model dan evaluasi akurasi dengan Receiver Operating Characteristic (ROC) dalam pembelajaran yang tidak seimbang.

6. Pinjaman

Pinjaman memiliki arti memberikan kepada orang lain suatu aset yang dapat diperoleh kembali atau diminta, yaitu meminjamkan tanpa mengharapkan imbalan apa pun [13], Pinjaman sering dipahami sebagai pengadaan barang secara angsuran atau angsuran di masa depan, atau peminjaman pinjaman uang untuk dilakukan dengan cara mencicil atau mencicil di lain waktu sesuai dengan suatu kontrak [13].

III. METODOLOGI PENELITIAN

Kerangka penelitian dibuat untuk memfasilitasi pencapaian hasil penelitian. Kerangka penelitian yang menjadi acuan bagi peneliti saat melakukan penelitian.

1. Studi Literatur

Pada tahap ini, pencarian dari pemikiran berbagai sumber seperti buku, jurnal, dan referensi lainnya, meneliti baik konsep maupun teori untuk menemukan referensi yang cocok dan relevan.

2. Identifikasi Masalah

Pada tahap ini penulis mengidentifikasi hal apa saja yang menjadi permasalahan data mining mengenai pemberian pinjaman pada nasabah koperasi, sehingga penulis dapat menentukan rencana kerja serta menentukan data apa saja yang akan dibutuhkan dalam penelitian.

3. Pengumpulan Data

Pengumpulan data merupakan tahap dilakukannya pencarian data untuk memperoleh informasi tentang data-data milik nasabah koperasi. Atribut data yang digunakan adalah jaminan, pekerjaan, status, jumlah tanggungan, jenis kelamin, usia, tempat tinggal, pendapatan, dan hasil survei, di antara banyak atribut data pelanggan. Data yang digunakan adalah 150 data pelanggan, yang darinya dilakukan pembersihan data atau pemilihan data untuk menyaring data yang bising untuk mencari data yang dapat digunakan untuk data latih. Setelah proses pembersihan data atau pemilihan data, diperoleh 134 data bebas noise dan digunakan untuk data latih. Kemudian 134 data diuji dengan data tersebut. Data uji yang digunakan berasal dari data latih dimana data telah dibersihkan agar bebas dari noise. Data ini digunakan untuk menghitung algoritma Random Forest dan C4.5.

4. *Cleaning Data*

Cleaning data adalah pra-pemrosesan yang membersihkan data dalam fase ini atau membebaskannya dari data yang tidak konsisten atau bising. Pelanggan dengan atribut yang tidak lengkap dipisahkan. Selain itu, ada beberapa hal yang dilakukan dalam proses pembersihan data:

- Klasifikasi pinjaman nasabah yang berupa angka menjadi 4 kategori yaitu rendah (100 rb- 500 rb), sedang (600 rb – 1jt), tinggi (1,5 jt – 5 jt), dan sangat tinggi (6 jt – 10 jt).
- Klasifikasi jumlah tanggungan yang dimiliki nasabah menjadi 3 kategori yaitu ringan (1-2), sedang (3-5), banyak (6-<6).
- Klasifikasi usia nasabah menjadi 3 kategori yaitu muda (18-28 tahun), sedang (>28-40 tahun), tua (>40 tahun).
- Klasifikasi penghasilan yang dimiliki nasabah menjadi 3 kategori yaitu rendah (100 rb – 500 rb), sedang (1 jt – 5 jt), tinggi (<5 jt).
- Klasifikasi Pekerjaan menjadi 2 kategori yaitu BUMN dan Non BUMN.
- Kemudian hasil dari *cleaning data* ini disusun dalam bentuk format csv pada tools Notepad++

5. Transformasi Data

Transformasi data adalah proses mengubah dan mengintegrasikan data yang digunakan dalam proses penambangan. Pada tahap ini, format data diubah menjadi format yang sesuai dengan metode data mining yang digunakan. WEKA mendukung beberapa format file untuk input termasuk format CSV. CSV adalah File teks berformat khusus yang dapat menyimpan data dalam format tabel terstruktur.

6. *Preprocessing*

Pada fase ini, sebelum algoritma klasifikasi Random Forest dan C4.5 melakukan proses klasifikasi, diperlukan *preprocessing* untuk menyeimbangkan dataset menggunakan metode Random Oversampling (ROS) berdasarkan class Pinjaman yang berisi kriteria Rendah, Sedang, Tinggi dan Sangat Tinggi. Penyeimbangan data ini dilakukan agar hasil akurasi menjadi lebih baik.

7. Proses Klasifikasi

Pada fase ini, proses klasifikasi data menggunakan algoritma Random Forest dan algoritma C4.5 dilakukan menggunakan tools Weka yang diperoleh dari data training dan testing.

8. Pengujian Model

Model yang diperoleh dengan metode Random Forest dan C4.5 diuji menggunakan dataset yang digunakan, *10 K-Fold Cross Validation*. *Cross Validation* adalah bentuk sederhana dari teknik statistik. Jumlah fold standar untuk memprediksi tingkat error dari data adalah dengan menggunakan *10 K-Fold Cross Validation*.

9. Hasil Analisis

Hasil analisis menilai keakuratan data model yang diperoleh dengan komputasi Random Forest dan C4.5 menggunakan metode 10-fold cross-validation. Dalam teknik ini, kumpulan data secara acak dibagi menjadi K partisi. Kemudian dilakukan percobaan K, masing-masing percobaan menggunakan data partisi Kth sebagai data uji dan partisi sisanya sebagai data latih. Di mana hasil persentase maksimum diambil. Selain itu, format visualisasi setiap atribut dapat dilihat pada tool WEKA.

IV. HASIL DAN PEMBAHASAN

Analisis masalah bertujuan untuk mengetahui dan mengevaluasi sistem yang sedang berjalan berdasarkan kebutuhan koperasi guna mencari solusi atas permasalahan yang dihadapi koperasi. Penelitian ini menggunakan metode Nave Bayes. Analisis dilakukan untuk mencari informasi baru dari data yang sudah ada. Jika dihitung dengan Nave Bayes, data yang diperoleh dibagi menjadi beberapa kelas.

1. Profil Data

Data diperoleh berdasarkan observasi yang dilakukan, data pelanggan terdiri dari 150. Data yang digunakan adalah data pelanggan tahun 2019, dan memiliki atribut garansi BPKP sepeda motor, sertifikat properti, sertifikat bangunan, KK (pekerjaan, jumlah anak, ID (identitas, jenis kelamin, usia, tempat tinggal), pendapatan, hasil survei, dan pinjaman, data nasabah disajikan pada Tabel 1 dan data lengkap disajikan pada Lampiran 1.

Tabel 1. Data Nasabah

NO	Nama Karyawan	Jk	Usia	Status	Domisili	Pekerjaan	Penghasilan	Jaminan	Tanggungan	Hasil survei	Pinjaman
1	A. KADIR MUNIF	L	35	MENIKAH	TANGERANG	PEGAWAI	4.250.750	MOTOR	3	BAIK	SEDANG
2	A. RUSMAN	L		MENIKAH	TANGERANG	PEGAWAI			5		
3	ABDI PRANATA	L	41	MENIKAH	TANGERANG	PEGAWAI	5.325.000	MOTOR	4	BAIK	TINGGI
4	ABDUL RAZAK	L	49	MENIKAH	TANGERANG	PEGAWAI	5.150.500	MOTOR	6	CUKUP BAIK	SANGAT TINGGI
5	ACHMAD SAUKI	L	27	MENIKAH	TANGERANG	PEGAWAI	3.850.000	SERTIFIKAT	3	BAIK	RENDAH

2. Preprocessing

Preprocessing adalah salah satu tahapan kunci dari data dalam proses data mining. Langkah-langkah yang dilakukan adalah pembersihan data, integrasi data, pemilihan data, dan transformasi data. Untuk data yang tidak memiliki kelengkapan sebagai data atribut, hapus data dengan noise, bersihkan data, ganti atribut untuk mengintegrasikan data, hapus atribut yang tidak digunakan untuk analisis, pilih data, dan ubah format data Sesuai dengan teknik data mining , yang mengubah data menjadi format.

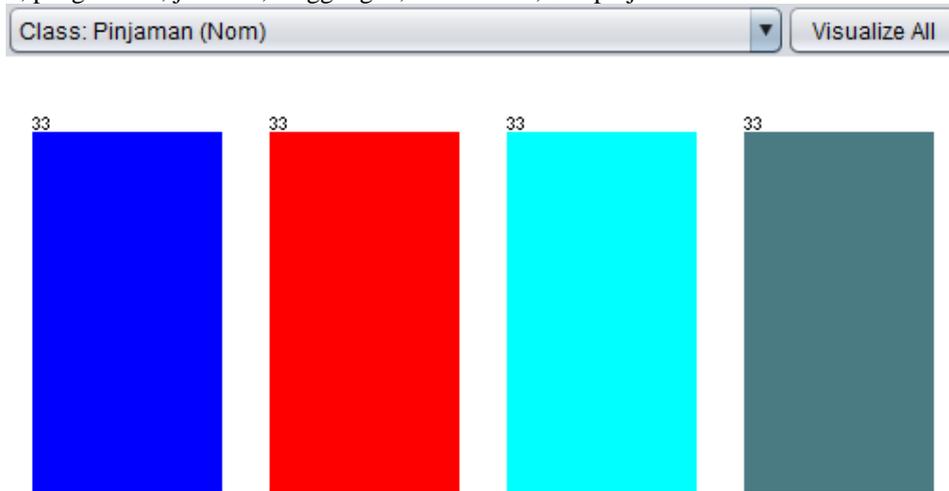
a. Cleaning Data

Data nasabah dilakukan cleaning data atau penyeleksian data untuk mengetahui data yang bisa digunakan untuk data training dengan menghilangkan data yang mengalami noise, misalnya redundansi data dan data inkonsisten.

Tabel 2. Data Seleksi

No	Nama Karyawan	Jk	Usia	Status	Domisili	Pekerjaan	Penghasilan	Jaminan	Tanggungan	Hasil survei	Pinjaman
1	A. KADIR MUNIF	L	35	MENIKAH	TANGERANG	PEGAWAI	4.250.750	MOTOR	3	BAIK	SEDANG
2	A. RUSMAN	L	-	MENIKAH	TANGERANG	PEGAWAI	-		5	-	
3	ABDI PRANATA	L	41	MENIKAH	TANGERANG	PEGAWAI	5.325.000	MOTOR	4	BAIK	TINGGI
4	ABDUL RAZAK	L	49	MENIKAH	TANGERANG	PEGAWAI	5.150.500	MOTOR	6	CUKUP BAIK	SANGAT TINGGI
5	ACHMAD SAUKI	L	27	MENIKAH	TANGERANG	PEGAWAI	3.850.000	SERTIFIKAT	3	BAIK	RENDAH

Pada tabel 2 terdapat data yang digaris merah, yaitu menunjukkan contoh data tidak lengkap atau mengandung noise sehingga data tersebut tidak bisa digunakan sebagai data training. Setelah dilakukan proses cleaning data atau penyeleksian data diperoleh 134 data yang tidak mengalami noise. Oleh karena itu pada penelitian ini penulis hanya menggunakan 134 data nasabah sebagai data training. Data nasabah dibagi menjadi 4 kelas yaitu 41 nasabah kelas RENDAH, 68 nasabah kelas SEDANG, 22 nasabah kelas TINGGI, dan 3 nasabah kelas SANGAT TINGGI, dataset yang tidak seimbang tersebut dilakukan preprocessing menggunakan metode Random Oversampling sehingga data menjadi seimbang sesuai dengan 4 kelas tersebut sesuai gambar 1. Atribut yang akan digunakan pada seluruh data nasabah yaitu jenis kelamin, usia, status, domisili, pekerjaan, penghasilan, jaminan, tanggungan, hasil survei, dan pinjaman.



Gambar 1. Dataset Setelah Preprocessing

b. Transformasi Data

Proses mentransformasikan data yang dilakukan untuk mengubah format data menjadi format yang nantinya sesuai dengan teknik data mining metode naïve bayes.

Tabel 3. Transformasi Data

Data Awal	Transformasi	Keterangan
Usia 18-28 Tahun	Muda	Usia 18 sampai 28 Tahun termasuk muda
Usia >28-40 Tahun	Sedang	Usia diatas 28 sampai 40 Tahun termasuk sedang
Usia >40 Tahun	Tua	Usia diatas 40 Tahun termasuk tua
Pekerjaan Pegawai	BUMN	Pekerjaan sebagai pegawai termasuk BUMN
Pekerjaan Cs Bandara, Airline, Outsourcing dll	Non BUMN	Pekerjaan sebagai Cs Bandara, Airline, Outsourcing dll termasuk Non BUMN
Penghasilan 100 rb-500 rb	Rendah	Penghasilan 100 rb sampai 500 rb termasuk Rendah
Penghasilan 1 jt- 5 jt	Sedang	Penghasilan 1 jt sampai 5 jt termasuk Sedang
Penghasilan > 5 jt	Tinggi	Penghasilan diatas 5 jt termasuk Tinggi
Tanggungan 1-2	Ringan	Tanggungan 1 sampai 2 termasuk Ringan
Tanggungan 3-5	Sedang	Tanggungan 3 sampai 5 termasuk Sedang
Tanggungan 6-<6	Tinggi	Tanggungan 6 sampai diatas 6 termasuk Tinggi
Pinjaman 100 rb-500 rb	Rendah	Penghasilan 100 rb sampai 500 rb termasuk Rendah
Pinjaman 600 rb - 1 jt	Sedang	Penghasilan 600 rb sampai 1 jt termasuk Sedang
Pinjaman 1.5 jt -5 jt	Tinggi	Penghasilan 1 jt sampai 5 jt termasuk Tinggi
Pinjaman 6 jt-10 jt	Sangat Tinggi	Penghasilan 1.5 jt sampai 10 jt termasuk Sangat Tinggi

Data latih dan data uji digunakan sebagai proses data mining. Data pelatihan digunakan sebagai pola data untuk membangun model data mining. Data yang digunakan terlebih dahulu harus diberi label, dan juga dilakukan proses preprocessing sedangkan data uji digunakan untuk pengujian dalam penerapan model klasifikasi. Data yang digunakan tidak perlu diberi tanda, tetapi harus melalui proses preprocessing. Di bawah ini adalah tabel data pelatihan yang digunakan untuk perhitungan Random Forest dan C4.5 Algorithm.

Tabel 4. Data Training

NO	Jk	Usia	Status	Domisili	Pekerjaan	Penghasilan	Jaminan	Tanggungan	Hasil survei	Pinjaman
1	L	SEDANG	MENIKAH	TANGERANG	BUMN	SEDANG	MOTOR	SEDANG	BAIK	SEDANG
2	L	TUA	MENIKAH	TANGERANG	BUMN	TINGGI	MOTOR	SEDANG	BAIK	TINGGI
3	L	TUA	MENIKAH	TANGERANG	BUMN	TINGGI	MOTOR	BANYAK	CUKUP BAIK	SANGAT TINGGI
4	L	MUDA	MENIKAH	TANGERANG	BUMN	SEDANG	SERTIFIKAT	SEDANG	BAIK	RENDAH
5	L	SEDANG	MENIKAH	TANGERANG	BUMN	SEDANG	MOTOR	SEDANG	BAIK	SEDANG

3. Hasil Klasifikasi *Random Forest*

Klasifikasi menggunakan algoritma Random Forest dilakukan dengan menggunakan *10-cross validation*. Berikut merupakan hasil klasifikasi menggunakan tools WEKA:

Tabel 5. Hasil Klasifikasi Random Forest

Perhitungan	Persentase
Akurasi	63.43 %
Sensitivitas/Recall/Tprate	0,634
Fprate	0,260
Precision/PPV	0,639
F-Measure	0,632
Area under ROC/ AUC	0,784

hasil klasifikasi Random Forest dengan menggunakan *10-cross validation* menunjukkan hasil prediksi benar dengan akurasi sebesar 63.43 % dan prediksi salah dengan persentase 36.56 %.

4. Hasil Klasifikasi Algoritma C4.5

Klasifikasi menggunakan Algoritma C4.5 dilakukan dengan menggunakan *10-cross validation*. Berikut merupakan hasil klasifikasi menggunakan tools WEKA:

Tabel 6. Hasil Klasifikasi Algoritma C4.5

Perhitungan	Persentase
Akurasi	70.14 %
Sensitivitas/Recall/TPrate	0,701
Fprate	0,251
Precision/PPV	0,713
F-Measure	0,685
Area under ROC/ AUC	0,752

hasil klasifikasi Algoritma C4.5 pada Tools WEKA dengan menggunakan *10-cross validation* menunjukkan hasil yang lebih baik dibanding Random Forest, prediksi benar dengan akurasi sebesar 70.14 % dan prediksi salah dengan persentasi 29.85 %.

5. Hasil Klasifikasi Algoritma C4.5 dengan Random Oversampling

Klasifikasi menggunakan Algoritma C4.5 dengan optimalisasi Random Oversampling dilakukan dengan menggunakan *10-cross validation*. Berikut merupakan hasil klasifikasi menggunakan tools WEKA:

Tabel 7. Hasil Klasifikasi C4.5 dengan Random Oversampling

Perhitungan	Persentase
Akurasi	78.03 %
Sensitivitas/Recall/Tprate	0,780
Fprate	0,073
Precision/PPV	0,786
F-Measure	0,782
Area under ROC/ AUC	0,896

hasil klasifikasi Algoritma C4.5 dengan optimalisasi Random Oversampling dengan menggunakan *10-cross validation* menunjukkan peningkatan dari hasil menggunakan algoritma C4.5 sebelumnya, dimana akurasi meningkat sebesar 7.89%, prediksi benar dengan akurasi sebesar 78.03% dan prediksi salah dengan persentasi 21.96 %.

6. Hasil Klasifikasi Algoritma Random Forest dengan Random Oversampling

Klasifikasi menggunakan Algoritma Random Forest dengan optimalisasi Random Oversampling dilakukan dengan menggunakan *10-cross validation*. Berikut merupakan hasil klasifikasi menggunakan tools WEKA:

Tabel 8. Hasil Klasifikasi C4.5 dengan Random Oversampling

Perhitungan	Persentase
Akurasi	87.12 %
Sensitivitas/Recall/Tprate	0,871
Fprate	0,043
Precision/PPV	0,873
F-Measure	0,872
Area under ROC/ AUC	0,963

hasil klasifikasi Algoritma Random Forest dengan optimalisasi Random Oversampling dengan menggunakan *10-cross validation* menunjukkan peningkatan dari hasil menggunakan algoritma Random Forest sebelumnya, dimana akurasi meningkat sebesar 23.69%, prediksi benar dengan akurasi sebesar 87.12% dan prediksi salah dengan persentasi 12.87 %.

V. KESIMPULAN

Penelitian ini mencoba mencari metode dengan akurasi terbaik pada algoritma Random Forest dan C4.5, untuk memberikan hasil akurasi tertinggi pada data nasabah koperasi dalam pemberian pinjaman. Data-data nasabah koperasi ini mengalami banyak noise hal ini diketahui melalui proses cleaning data, data nasabah yang dapat digunakan yaitu 134 data nasabah. Data nasabah koperasi juga dalam kondisi tidak seimbang sehingga mempengaruhi perhitungan akurasi pada metode yang diuji.

Dalam hasil pengujian yang dilakukan metode *Random Forest* menggunakan *10-cross validation* mendapat akurasi 63.43% dan algoritma C4.5 menggunakan *10-cross validation* mendapat akurasi 70.14%. Sedangkan pengujian dengan dua metode tersebut yang dioptimalkan dengan Random Oversampling, mengalami peningkatan akurasi dengan hasil *Random Forest* mendapat akurasi 87.12% dan C4.5 mendapat akurasi 78.03%. Akurasi tertinggi pada pengujian klasifikasi data nasabah koperasi didapat dari algoritma Random Forest yang dioptimalkan dengan Random Oversampling sebesar 87.12%.

REFERENSI

- [1] A. B. Nugraha, “Perbedaan Saham dan Sertifikat Modal Koperasi Ditinjau Dari Kajian Yuridis Menurut Hukum Koperasi Indonesia,” *J. Ilmu Sos. dan Pendidik.*, vol. 5, no. 4, pp. 2598–9944, 2021, doi: 10.36312/jisip.v5i4.2599/http.
- [2] I. G. T. Isa and G. P. Hartawan, “Perancangan Aplikasi Koperasi Simpan Pinjam Berbasis Web (Studi,” *J. Ilm. Ilmu Ekon.*, vol. 5, no. 10, pp. 139–151, 2017.
- [3] E. Supriyanto and N. Ismawati, “Sistem Informasi Fintech Pinjaman Online Berbasis,” *J. Sist. Informasi, Teknologi Inf. dan Komput.*, vol. 9, no. 2, pp. 100–107, 2019.
- [4] A. Sucipto, “Prediksi Kredit Macet Melalui Perilaku Nasabah Pada Koperasi Simpan Pinjam Dengan Menggunakan Metode Alogaritma Klasifikasi C4.5,” *J. DISPROTEK*, vol. 6, no. 1, pp. 75–87, 2015.
- [5] D. P. Utomo and M. Mesran, “Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung,” *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 437, 2020, doi: 10.30865/mib.v4i2.2080.
- [6] Dr. T. Senthil Kumar, “Data Mining Based Marketing Decision Support System Using Hybrid Machine Learning Algorithm,” *J. Artif. Intell. Capsul. Networks*, vol. 2, no. 3, pp. 185–193, 2020, doi: 10.36548/jaicn.2020.3.006.
- [7] S. Al Syahdan and A. Sindar, “Data Mining Penjualan Produk Dengan Metode Apriori Pada Indomaret Galang Kota,” *J. Nas. Komputasi dan Teknol. Inf.*, vol. 1, no. 2, 2018, doi: 10.32672/jnkti.v1i2.771.
- [8] D. Hartama, A. Perdana Windarto, and A. Wanto, “The Application of Data Mining in Determining Patterns of Interest of High School Graduates,” *J. Phys. Conf. Ser.*, vol. 1339, no. 1, 2019, doi: 10.1088/1742-6596/1339/1/012042.
- [9] H. Annur, “Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes,” *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018, doi: 10.33096/ilkom.v10i2.303.160-165.
- [10] K. Shah, H. Patel, D. Sanghvi, and M. Shah, “A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification,” *Augment. Hum. Res.*, vol. 5, no. 1, 2020, doi: 10.1007/s41133-020-00032-0.
- [11] L. Cheng, X. Chen, J. De Vos, X. Lai, and F. Witlox, “Applying a random forest method approach to model travel mode choice behavior,” *Travel Behav. Soc.*, vol. 14, no. August 2018, pp. 1–10, 2019, doi: 10.1016/j.tbs.2018.09.002.
- [12] V. Anestiviya, A. Ferico, and O. Pasariibu, “Analisis Pola Menggunakan Metode C4.5 Untuk Peminatan Jurusan Siswa Berdasarkan Kurikulum (Studi Kasus : Sman 1 Natar),” *J. Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 80–85, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>.
- [13] H. Sabijono, V. Ilat, and Y. N. Makaombohe, “Rasio Likuiditas Dan Jumlah Kredit Terhadap Profitabilitas Perbankan Di Bursa Efek Indonesia,” *J. Ris. Ekon. Manajemen, Bisnis dan Akunt.*, vol. 2, no. 1, pp. 617–626, 2014.