

Analisis Kelayakan Kredit Menggunakan *Classification Tree* dengan Teknik *Random Oversampling*

Lo Mei Ly Vebriyanti, Shantika Martha, Wirda Andani, dan Setyo Wira Rizki



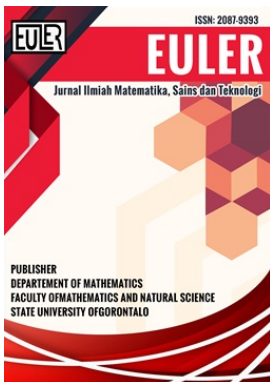
Volume 12, Issue 1, Pages 1–8, June 2024

Diterima 17 Januari 2024, Direvisi 7 Mei 2024, Disetujui 10 Mei 2024, Diterbitkan 16 Mei 2024

To Cite this Article : L. M. L. Vebriyanti, dkk., "Analisis Kelayakan Kredit Menggunakan *Classification Tree* dengan Teknik *Random Oversampling*", *Euler J. Ilm. Mat. Sains dan Teknol.*, vol. 12, no. 1, pp. 1–8, 2024, <https://doi.org/10.37905/euler.v12i1.24182>

© 2024 by author(s)

JOURNAL INFO • EULER : JURNAL ILMIAH MATEMATIKA, SAINS DAN TEKNOLOGI

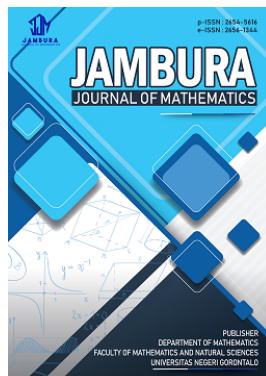


	Homepage	:	http://ejurnal.ung.ac.id/index.php/euler/index
	Journal Abbreviation	:	Euler J. Ilm. Mat. Sains dan Teknol.
	Frequency	:	Biannual (June and December)
	Publication Language	:	English (preferable), Indonesia
	DOI	:	https://doi.org/10.37905/euler
	Online ISSN	:	2776-3706
	Editor-in-Chief	:	Resmawan
	Publisher	:	Department of Mathematics, Universitas Negeri Gorontalo
	Country	:	Indonesia
	OAI Address	:	http://ejurnal.ung.ac.id/index.php/euler/oai
	Google Scholar ID	:	QF_r_gAAAAJ
	Email	:	euler@ung.ac.id

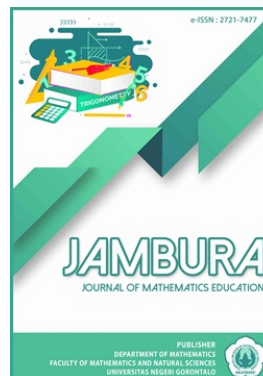
JAMBURA JOURNAL • FIND OUR OTHER JOURNALS



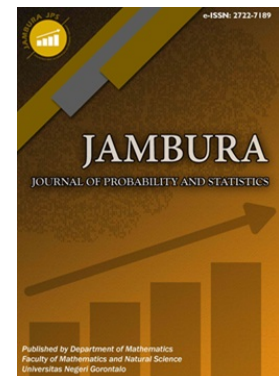
Jambura Journal of Biomathematics



Jambura Journal of Mathematics



Jambura Journal of Mathematics Education



Jambura Journal of Probability and Statistics

Analisis Kelayakan Kredit Menggunakan *Classification Tree* dengan Teknik *Random Oversampling*

Lo Mei Ly Vebriyanti^{1,*}, Shantika Martha¹, Wirda Andani¹, dan Setyo Wira Rizki¹

¹Jurusan Matematika, Universitas Tanjungpura, Indonesia

ARTICLE HISTORY

Diterima 17 Januari 2024

Direvisi 7 Mei 2024

Disetujui 10 Mei 2024

Diterbitkan 16 Mei 2024

KATA KUNCI

Kelayakan Kredit
Kredit Macet
Debitur
Classification Tree
Imbalanced Data

KEYWORDS

Creditworthiness
Bad Debt
Debtors
Classification Tree
Imbalanced Data

ABSTRAK. Kredit adalah kegiatan pemberian uang atau tagihan berdasarkan suatu perjanjian antara bank dan pihak lainnya. Pemberian kredit tidak terlepas dari risiko kredit macet sehingga analisis kredit wajib dilakukan terhadap calon debitur sebelum menyetujui pinjaman yang diajukan. Penelitian bertujuan untuk menganalisis kelayakan kredit menggunakan *Classification Tree* sebagai metode klasifikasi dengan teknik *Random Oversampling* untuk mengatasi imbalanced data. Penelitian ini menggunakan data sekunder, yaitu data status debitur salah satu bank di Kalimantan Barat. Data penelitian berjumlah 800 sampel data terdiri dari variabel kolektibilitas sebagai variabel target dan 10 variabel independen yaitu limit, rate, tenor, total angsuran, usia, gaji, premi dan admin, instansi, jenis kredit dan jenis kebutuhan. Metode yang digunakan adalah metode *Classification Tree* dengan teknik *Random Oversampling* untuk mengatasi imbalanced data. Pengklasifikasian diawali dengan *preprocessing data*, lalu data dibagi menjadi data latih dan uji dengan proporsi 70:30, 80:20 dan 90:10 untuk masing-masing perlakuan tanpa *Random Oversampling* dan dengan *Random Oversampling*. Selanjutnya, model klasifikasi dibentuk menggunakan data latih dan validasi model klasifikasi menggunakan data uji. Setelah itu, dilakukan evaluasi keseluruhan model untuk mengetahui model terbaik yang digunakan dalam proses klasifikasi. Berdasarkan hasil penelitian, model terbaik adalah model dari *Classification Tree* dengan teknik *Random Oversampling* proporsi 70:30, dengan nilai akurasi sebesar 89,17%, nilai *specificity* sebesar 75,00% dan nilai *recall* sebesar 89,66%. Dapat diartikan, model dapat digunakan untuk mengklasifikasikan data lancar dan non lancar debitur dengan variabel paling berpengaruh dalam pengklasifikasian status debitur adalah variabel total angsuran.

ABSTRACT. Credit is providing money or bills based on the agreement between a bank and another party. Lending is inseparable from bad credit risk, so credit analysis must be conducted on prospective debtors before approving a proposed loan. This research aims to analyze creditworthiness using a *Classification Tree* as a classification method with *Random Oversampling* to overcome imbalanced data. This study uses secondary data on the status of debtors from a bank in West Kalimantan. Research data amounted to 800 data samples consisting of collectability variables as target variables and 10 independent variables, namely limit, rate, tenor, total installments, age, salary, premium and admin, agency, type credit, and type need. The *Classification Tree* method with *Random Oversampling* is used to overcome imbalanced data. Classification begins with data preprocessing, then the data is divided into training and test data with proportions of 70:30, 80:20 and 90:10 for each treatment without *Random Oversampling* and with *Random Oversampling*. Next, a classification model is formed using training data, and the classification model is validated using test data. After that, an overall evaluation of the model is carried out to determine the best model used in the classification process. Based on the research results, the best model is the model *Classification Tree* with *Random Oversampling* in proportion 70:30, with an accuracy value of 89.17%, specificity of 75.00%, and recall of 89.66%. The model can be used to classify current and non-current debtor data. The most influential variable in classifying debtor status is the total installment variable.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License. *Editorial of EULER*: Department of Mathematics, Universitas Negeri Gorontalo, Jln. Prof. Dr. Ing. B. J. Habibie, Bone Bolango 96554, Indonesia.

1. Pendahuluan

Kredit menurut undang-undang perbankan nomor 10 tahun 1998 merupakan suatu peminjaman uang atau hutang yang mewajibkan debitur untuk membayar utangnya sesuai jangka waktu yang telah disepakati dengan pemberian bunga sebagai imbalan balas jasa, berdasarkan perjanjian dan kontrak pinjaman antara bank dengan debitur [1]. Kredit yang diajukan oleh debitur

mempunyai risiko. Hal ini dikarenakan adanya kemungkinan debitur yang mendaftar namun mengalami kesulitan dalam pembayarannya yang menyebabkan terjadinya kredit macet. Sebelum menyepakati pinjaman yang diminta oleh debitur, perbankan dapat menganalisis kredit yang diajukan oleh debitur untuk menentukan apakah permohonan kredit akan dapat disetujui. Analisis kredit merupakan analisis yang berkaitan dengan beberapa faktor yang berpengaruh terhadap kelancaran atau tidaknya pelu-

*Penulis Korespondensi.

nasan kredit [2]. Pengambilan keputusan disetujui atau tidaknya seorang debitur dapat menggunakan data historis debitur yang disetujui oleh bank. Akan tetapi, perlu diketahui bahwa tidak semua penyetujuan kredit debitur adalah pembayar kredit yang baik. Terdapat beberapa debitur yang sudah disetujui tetapi terlambat membayar setelah beberapa bulan sehingga menyebabkan risiko terjadinya kredit macet [3]. Upaya untuk menangani kredit macet adalah dengan dilakukan klasifikasi untuk membedakan kelompok antara debitur lancar dan non lancar.

Klasifikasi bertujuan untuk mempelajari fungsi yang berbeda. Setiap fungsi memetakan setiap data yang dipilih ke salah satu kelompok kelas yang sebelumnya telah ditentukan [3]. Klasifikasi termasuk dalam *supervised learning*. Metode *supervised learning* bergantung pada data latih, yang jarang memiliki jumlah sebaran data yang sama untuk setiap kelas. Jika kelas tertentu memiliki jumlah sebaran data yang sangat besar dibandingkan kelas lainnya, itu dianggap sebagai data yang tidak seimbang (*imbalance data*) [4].

Salah satu faktor yang sangat berperan dalam kemampuan prediksi klasifikasi adalah *imbalance data*. Metode klasifikasi seperti *Classification Tree* cenderung membuat model *learning* yang tidak akurat untuk kelas minoritas dibandingkan dengan kelas mayoritas. Hal ini dikarenakan *Classification Tree* dirancang untuk asumsi distribusi kelas yang seimbang [5]. Keadaan data di mana ada ketidakseimbangan antara kelas data pertama dan kelas data lainnya disebut *imbalance data*. Salah satu masalah dalam klasifikasi adalah ketidakseimbangan data karena model klasifikasi cenderung membuat prediksi tentang kelas data yang besar (mayoritas) lebih akurat daripada kelas data yang kecil (minoritas). Akibatnya, prediksi kelas mayoritas lebih akurat daripada prediksi kelas minoritas pada data latih [6]. Salah satu cara untuk menangani *imbalance data* adalah dengan melakukan teknik *Random Oversampling*. Teknik *Random Oversampling* dapat menangani *imbalance data* dengan cara menambah data pada kelas minoritas tanpa mengurangi jumlah data pada kelas lainnya.

Decision Tree disebut sebagai *Classification Tree* apabila variabel targetnya berupa kategorik, sedangkan apabila variabel targetnya berupa numerik maka disebut dengan metode *Regression Tree* [7]. Metode *Classification Tree* merupakan metode non parametrik sehingga tidak perlu memperhatikan asumsi normalitas data. Struktur data dapat digunakan untuk mempermudah eksplorasi data dan pengambilan keputusan berdasarkan model yang diperoleh yang dapat dilihat secara visual. *Classification Tree* bertujuan untuk menghasilkan klasifikasi yang akurat dan menginterpretasikan prediksi data baru untuk setiap kategori yang muncul sebagai tanggapan dari penelitian [8].

Handayani, dkk. [9] melakukan penelitian terkait memprediksi tingkat risiko kredit dengan data mining menggunakan algoritma *Decision Tree C.45*. Penelitian ini bertujuan untuk memprediksi tingginya risiko kredit dengan menggunakan algoritma pohon keputusan C.45. Prediksi tingginya risiko menggunakan dua kelas, yaitu kelas lancar dan kelas non lancar. Hasil penelitian berhasil membuktikan bahwa metode hasil evaluasi menggunakan *confusion matrix* didapatkan akurasi sebesar 79%.

Berdasarkan hasil evaluasi penelitian [9] menggunakan *confusion matrix*, algoritma *Decision Tree C.45* dianggap mampu untuk memprediksi kelayakan kredit calon debitur sebesar 79%. Akan tetapi, masih terdapat pengembangan yang diperlukan untuk

meningkatkan analisis kelayakan kredit seperti menggunakan variasi proporsi data latih dan data uji untuk meningkatkan nilai akurasi. Oleh karena itu, penelitian ini mencoba melanjutkan dengan menggunakan berbagai variasi nilai latih dan nilai uji dalam upaya peningkatan nilai akurasi terbaik dengan metode *Classification Tree* untuk analisis kelayakan kredit dan teknik *Random Oversampling* untuk menangani adanya *imbalance data*. Penelitian ini dilakukan untuk menangani ketidakseimbangan kelas pada data kredit debitur salah satu bank di Kalimantan Barat dengan teknik *Random Oversampling* menggunakan metode *Classification Tree*. Hasil penelitian ini diharapkan dapat digunakan untuk keperluan dasar pengambilan keputusan penilaian kelayakan debitur di suatu instansi.

2. Metode

Data yang digunakan dalam penelitian ini berasal dari salah satu bank di Kalimantan Barat. Jumlah variabel yang digunakan dalam penelitian ini sebanyak 11 variabel, dimana berupa 10 variabel independen yang mencakup *limit*, *rate*, tenor, total angsuran, jenis kebutuhan, gaji, premi dan biaya admin, instansi, jenis kredit dan usia. Sedangkan untuk variabel target berjumlah sebanyak 1 variabel, yakni kolektibilitas. Jumlah sampel yang digunakan adalah sebanyak 800 data.

Teknik analisis data penelitian ini yaitu klasifikasi menggunakan metode *Classification Tree* dengan teknik *Random Oversampling* untuk mengatasi *imbalanced data*, Teknik *Random Oversampling* digunakan karena adanya ketidakseimbangan pada kelas data. Proses analisis dalam penelitian ini menggunakan *software R Studio*. Pada penelitian ini, akan dilakukan proses analisis *data mining* menggunakan metode *Classification Tree* untuk data debitur lancar dan non lancar pada data kredit yang diperoleh dari salah satu bank di Kalimantan Barat.

2.1. Confusion Matrix

Confusion matrix merupakan suatu metode yang menyediakan informasi klasifikasi yang diprediksi dengan benar oleh sistem klasifikasi. Metode ini umumnya digunakan untuk menghitung tingkat kinerja model dalam *data mining* [10]. *Confusion matrix* dilakukan untuk memprediksikan objek yang benar dan salah [11]. Contoh tabel *confusion matrix* [12] ditunjukkan pada Tabel 1.

Tabel 1. Confusion Matrix

		Nilai Prediksi	
		Negative (0)	Positive (1)
Nilai Aktual	Negative (0)	TN	FP
	Positive (1)	FN	TP

Terdapat lima parameter pada *confusion matrix* yang perlu dihitung, yaitu akurasi, presisi, *recall/sensitivity*, *specificity* dan *F1-Score*. Akurasi digunakan untuk mengukur total keseluruhan dari model klasifikasi yang bernilai *positive*, presisi digunakan untuk mengukur prediksi yang bernilai benar jika model memprediksi *positive*, *recall/sensitivity* digunakan untuk menghitung kebenaran prediksi dari data seluruh data *positive*, *specificity* digunakan untuk mengukur kebenaran prediksi dari data seluruh data *negative* dan *F1-Score* digunakan untuk menghitung rata-rata dari presisi dan *recall/sensitivity*. Nilai akurasi (1), presisi (2), *recall/sensitivity* (3), *specificity* (4) dan *F1-Score* (5) dapat diketahui menggunakan rumus

dari *confusion matrix* [13]:

$$\text{Akurasi} = \frac{TP + TN}{\text{Total}} \times 100\%, \quad (1)$$

$$\text{Presisi} = \frac{TP}{TP + FP} \times 100\%, \quad (2)$$

$$\text{Recall/sensitivity} = \frac{TP}{TP + FN} \times 100\%, \quad (3)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \times 100\%, \quad (4)$$

$$F1 - \text{Score} = 2 \times \frac{\text{Recall} \times \text{Presisi}}{\text{Recall} + \text{Presisi}} \times 100\%, \quad (5)$$

dengan

- TP (*True Positive*) = data *positive* yang terdeteksi dengan benar
- TN (*True Negative*) = data *negative* yang terdeteksi dengan benar
- FP (*False Positive*) = data *negative* yang terdeteksi sebagai data *positive*
- FN (*False Negative*) = data *positive* yang terdeteksi sebagai data *negative*

2.2. Classification Tree

Konsep dasar *Classification Tree* adalah mengubah data menjadi *rules* dalam bentuk struktur pohon yang mewakili keputusan. Ciri-ciri *Classification Tree* yang terbentuk dari beberapa elemen mengacu pada [14], yaitu:

1. *Root node* (node akar), yaitu node yang tidak memiliki lengan *input* dan tidak ada memiliki lengan *output*;
2. *Node internal*, yaitu setiap node non-terminal (bukan daun) yang memiliki satu lengan *input* dan beberapa lengan *output*. Node ini yang mewakili pengujian berdasarkan nilai fitur;
3. *Lengan*, yaitu setiap cabang yang menunjukkan nilai pengujian pada node non-terminal;
4. *Node daun* (terminal), yaitu node yang menentukan keputusan. Node ini memiliki satu lengan *input* dan tidak memiliki lengan *output*.

Beberapa langkah dalam pembentukan *splitting node tree* untuk kasus klasifikasi mengacu pada [15], yaitu:

1. Menentukan *class*;
2. Mengitung nilai *entropy* dari kelas variabel;
3. Menghitung *information gain* untuk setiap variabel;
4. Melakukan proses split. Variabel dengan *information gain* terbesar akan di split terlebih dahulu;
5. Mengulangi proses perhitungan *entropy* untuk variabel lainnya dan lakukan *split* hingga selesai;
6. Membentuk kelas untuk setiap lengan (cabang).

Pembuatan *Classification Tree* diperlukan perhitungan dua jenis *entropy*, yakni *entropy* dari variabel target dan *entropy* antara variabel prediktor dan variabel target, sehingga *entropy* dari variabel target dapat diketahui dengan menggunakan Persamaan (6) [15], yaitu

$$E(S) = \sum_{i=1}^n -p_i \log_2 p_i \quad (6)$$

dengan

$E(S)$ = *entropy* variabel target

n = jumlah total sampel

p_i = probabilitas dari sampel ke- i .

Entropy dari variabel prediktor dan variabel target dapat diketahui menggunakan Persamaan (7) [15], yaitu

$$E(T, X) = \sum_{c \in X} P(c) E(S) \quad (7)$$

dengan

$E(T, X)$ = *entropy* variabel target dan variabel prediktor

$c \in X$ = sampel untuk setiap variabel prediktor

$P(c)$ = probabilitas sampel

$E(S)$ = *entropy* variabel target.

Menghitung nilai *information gain* setelah seluruh nilai setiap *entropy* dihitung. Nilai *information gain* dapat diketahui dengan menggunakan Persamaan (8) [15], yaitu

$$\text{Information Gain}(T, X) = E(S) - E(T, X) \quad (8)$$

dengan

T = variabel target

X = variabel prediktor

$E(S)$ = *entropy* variabel target

$E(T, X)$ = *entropy* variabel target dan variabel prediktor.

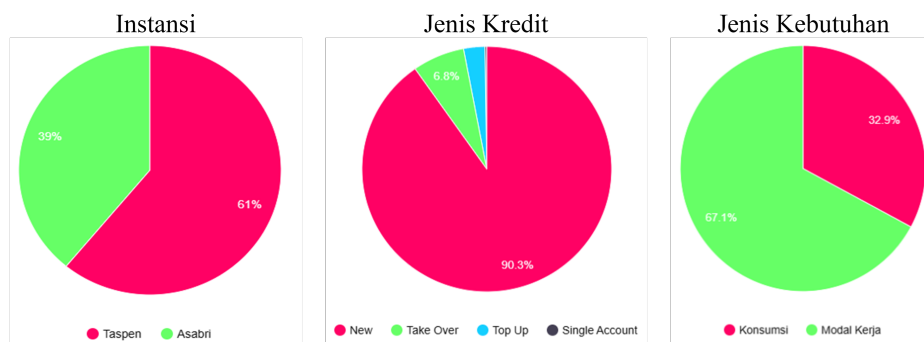
2.3. Random Oversampling (ROS)

Ketidakseimbangan data dapat memberikan pengaruh negatif terhadap hasil klasifikasi ketika kelas minoritas sering salah diklasifikasikan sebagai kelas mayoritas [16]. Adanya *imbalanced data* untuk proses klasifikasi mempengaruhi hasil klasifikasi pada data minoritas menjadi tidak sesuai atau tertutup oleh prediksi data mayoritas, sehingga diperlukan penanganan untuk permasalahan tersebut [6].

Terdapat tiga pendekatan untuk mengatasi ketidakseimbangan data [5], yaitu pendekatan level data, pendekatan level algoritma dan pendekatan level *ensemble*. Pendekatan level data menggunakan berbagai teknik *resampling*, seperti *Random Oversampling* (ROS) dan *Random Undersampling* (RUS), untuk mengubah ketimpangan distribusi kelas. Pendekatan level algoritma mengubah langkah-langkah proses *learning* untuk menangani masalah distribusi kelas. Pendekatan level *ensemble* membagi kelas mayoritas menjadi beberapa subkelompok dengan ukuran yang sama dengan kelas minoritas. Kemudian, masing-masing subkelompok menerima instruksi yang berbeda, sehingga masing-masing menghasilkan model *learning* yang berbeda. Model *learning* terbaik dipilih oleh keputusan mayoritas atau *majority voting*. Contoh pendekatan level *ensemble* yaitu *Bagging* dan *Boosting*. Pendekatan level data, yaitu *Random Oversampling* (ROS) adalah metode yang secara acak meningkatkan dataset pada kelas minoritas untuk membuat distribusi kelas menjadi seimbang [16]. *Random Oversampling* dilakukan dengan tujuan rasio ketimpangan kelas dapat dikurangi.

3. Hasil dan Pembahasan

Pada bagian ini diberikan uraian hasil implementasi *Classification Tree* tanpa *Random Oversampling* dan *Classification Tree* dengan *Random Oversampling* menggunakan alat bantu *software R*



Gambar 1. Diagram Lingkaran Variabel Kategorik

Studio dengan membagi data menjadi beberapa proporsi. *Confusion Matrix* diterapkan untuk melihat nilai parameter terbaik yang akan digunakan untuk menentukan model terbaik.

3.1. Preprocessing Data

Analisis data penelitian diawali dengan dilakukannya *preprocessing data*. Proses *preprocessing data* termasuk pembersihan (*data cleaning*), integrasi (*data integration*), seleksi (*data selection*), dan transformasi (*data transformation*) [17]. Data sering mengandung rekaman dengan nilai variabel yang tidak lengkap, kosong, tidak konsisten, dan *noise*, sehingga memerlukan pembersihan data. Karena dapat berdampak negatif pada hasil penelitian, informasi ini harus dihilangkan daripada dimasukkan dalam proses *data mining*. Beberapa variabel yang diasumsikan berpengaruh terhadap penyebab kredit macet diantaranya adalah *limit*, *rate*, *tenor*, *total angsuran*, *usia*, *gaji*, *premi* dan *admin*, *instansi*, *jenis kredit* dan *jenis kebutuhan*. Tabel 2 menunjukkan beberapa variabel yang diambil dari data debitur lancar dan non lancar pada salah satu bank di Kalimantan Barat.

Tabel 2. Variabel Penelitian

Variabel	Keterangan
Limit	Jumlah kredit terbesar yang dapat diberikan bank kepada debitur
Rate	Biaya yang dibebankan oleh bank kepada debitur
Tenor	Lama waktu yang disepakati untuk melakukan pelunasan kredit
Total angsuran	Jumlah pembayaran kredit yang dilakukan setiap periode
Usia	Usia debitur
Gaji	Pendapatan debitur
Premi dan admin	Sejumlah uang yang wajib dibayarkan setiap bulannya
Instansi	Terdiri dari ASABRI dengan kategori "0" dan TASPEN dengan kategori "1"
Jenis kredit	Jenis kredit yang diajukan. Terdiri dari <i>new</i> dengan kategori "0", <i>take over</i> dengan kategori "1", <i>top up</i> dengan kategori "2" dan <i>top up single account</i> dengan kategori "3"
Jenis kebutuhan	Kebutuhan peminjaman, yang terdiri dari konsumsi dengan kategori "0" dan modal kerja dengan kategori "1"
Kolektibilitas	Klasifikasi status pembayaran kredit, terdiri dari non lancar dengan kategori "0" dan lancar dengan kategori "1"

3.2. Statistik Deskriptif

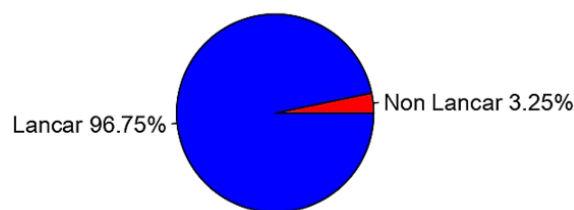
Statistik deskriptif bertujuan untuk memberikan informasi atau deskripsi pada objek data yang diteliti dalam penelitian. Analisis statistik deskriptif berkaitan dengan variabel kuantitatif seperti *limit*, *rate*, *tenor*, *total angsuran*, *usia*, *gaji*, *premi* dan *admin* yang ditunjukkan pada Tabel 3.

Diketahui bahwa pada nilai minimum pada variabel total angsuran sebesar Rp62.940,00 dan nilai maksimal pada variabel total angsuran sebesar Rp3.922.333,00. Tenor atau durasi pinjaman yang harus dibayarkan oleh debitur paling sebentar adalah 12 bulan dan paling lama adalah 187 bulan.

Hasil analisis statistik deskriptif untuk variabel instansi, jenis kredit dan jenis kebutuhan ditunjukkan pada Gambar 1.

Diketahui bahwa instansi yang paling banyak digunakan yaitu TASPEN dengan jumlah 488 debitur atau sebesar 61%, jenis kredit yang paling banyak digunakan yaitu *new* dengan jumlah 722 debitur atau sebesar 90,3%, dan jenis kebutuhan yang paling banyak digunakan yaitu kredit untuk modal kerja dengan jumlah 537 debitur atau sebesar 67,1%.

Kolektibilitas dikategorikan dalam bilangan biner, yaitu "0" untuk kategori debitur dengan pembayaran non lancar dan "1" untuk kategori debitur dengan pembayaran lancar. Proses dalam analisis statistika deskriptif menggunakan *software R Studio*. Diagram lingkaran diperoleh dari variabel target atau variabel kolektibilitas pada Gambar 2.



Gambar 2. Diagram Lingkaran Kolektibilitas

Diketahui bahwa dari 800 debitur terdapat sebanyak 96,75% atau setara 774 debitur yang dikategorikan lancar dan 3,25% atau setara 26 debitur yang dikategorikan non lancar. Data tidak seimbang (*imbalanced data*) terjadi ketika nilai variabel target memiliki jumlah kelas yang jauh lebih kecil daripada nilai kelas lainnya. Analisis data diawali dengan menangani permasalahan pada *imbalanced data* yang bertujuan agar keseimbangan kelas dapat terpenuhi. Penanganan *imbalanced data* diatasi dengan dilakukannya teknik *Random Oversampling* yaitu peningkatan jumlah sampel dari kelas minoritas. Hasil keseimbangan kelas

Tabel 3. Statistik Deskriptif Variabel Numerik

Variabel Numerik	Min	Max	Mean	Standar Deviasi
Limit	5.000.000,00	298.200.000,00	64.486.776,00	40.252.600,00
Rate	11,30	18,85	14,53	1,39
Tenor	12,00	187,00	144,00	46,48
Total angsuran	62.940,00	3.922.333,00	977.177,00	532.631,00
Usia	33,00	75,00	61,47	5,89
Gaji	730.000,00	4.841.200,00	2.388.434,00	1.164.850,00
Premi dan admin	257.428,00	23.004.400,00	4.160.484,00	3.169.030,00

kemudian dilakukan metode klasifikasi *Classification Tree*.

3.3. Pembagian Proporsi Data Latih dan Data Uji

Proses klasifikasi diawali dengan membagi data menjadi data latih dan data uji. Data latih adalah kumpulan data yang memiliki variabel kelas yang digunakan mesin untuk menentukan karakteristik dari kumpulan data untuk membuat model data. Data uji juga memiliki kategori yang digunakan untuk menguji ketepatan model dalam mengklasifikasikan data uji, tetapi variabel kelas ini disembunyikan selama proses klasifikasi. Digunakan proporsi 70:30, 80:20, dan 90:10 untuk membagi data latihan dan uji. Program R Studio digunakan untuk menyelesaikan tugas ini. Adapun proporsi data latih dan data ditunjukkan pada Tabel 4.

Tabel 4. Proporsi Data Latih dan Data Uji

Proporsi Data	Total Data Latih	Total Data Uji
70:30 tanpa ROS	560 data	240 data
70:30 dengan ROS	1.084 data	240 data
80:20 tanpa ROS	640 data	160 data
80:20 dengan ROS	1.238 data	160 data
90:10 tanpa ROS	720 data	80 data
90:10 dengan ROS	1.394 data	80 data

Proses tanpa *Random Oversampling* dengan menggunakan proporsi 70:30 diperoleh klasifikasi data yang masuk ke dalam data latih yaitu sebanyak 560 data dan data uji sebanyak 240 data, proses tanpa *Random Oversampling* dengan menggunakan proporsi 80:20 diperoleh klasifikasi data yang masuk ke dalam data latih yaitu sebanyak 640 data dan data uji sebanyak 160 data, dan proses tanpa *Random Oversampling* dengan menggunakan proporsi 90:10 diperoleh klasifikasi data yang masuk ke dalam data latih yaitu sebanyak 720 data dan data uji sebanyak 80 data. Proses dengan *Random Oversampling* dengan menggunakan proporsi 70:30 diperoleh data latih digandakan menjadi sebanyak 1.084 data, proses dengan *Random Oversampling* dengan menggunakan proporsi 80:20 diperoleh data latih digandakan menjadi sebanyak 1.238 data, dan proses dengan *Random Oversampling* dengan menggunakan proporsi 90:10 diperoleh data latih digandakan menjadi sebanyak 1.394 data.

Model pembelajaran klasifikasi yang dibangun adalah *Classification Tree*, sedangkan teknik yang digunakan untuk mengatasi *imbalanced data* yaitu *Random Oversampling*. Data dibagi menjadi tiga proporsi split data latih dan data uji, yaitu proporsi 70:30, proporsi 80:20 dan proporsi 90:10 dengan membandingkan hasil antara tanpa teknik *Random Oversampling* dan dengan teknik *Random Oversampling* pada data.

3.4. Classification Tree tanpa Random Oversampling

Pemodelan menggunakan metode *Classification Tree* tanpa *Random Oversampling* diawali dengan melakukan pemodelan data latih menggunakan *minsplit* 100. Pembentukan model *Classification Tree* tanpa *Random Oversampling* menggunakan proporsi 70% data latih dan 30% data uji dilakukan dengan bantuan *software R Studio*. Hasil kinerja model *Classification Tree* tanpa *Random Oversampling* untuk *minsplit* 100 pada data latih menggunakan *confusion matrix* disajikan pada Tabel 5.

Tabel 5. Klasifikasi Model *Classification Tree* tanpa *Random Oversampling* Menggunakan *Minsplit* 100 pada Data Latih

		Nilai Prediksi	
		Tidak Lancar	Lancar
Nilai Aktual	Tidak Lancar	0	18
	Lancar	0	542

Diperoleh hasil dari *confusion matrix Classification Tree* tanpa *Random Oversampling* dengan proporsi 70:30, sehingga dapat dihitung tingkat akurasi, presisi, *recall*, *specificity* dan *F1-Score* model pada data latih dengan menggunakan persamaan (1)-(5), diperoleh nilai akurasi sebesar 96,79%, presisi 96,79%, *recall/sensitivity* 100,00%, *specificity* 0,00% dan *F1-Score* 98,37%. Nilai akurasi menunjukkan kemampuan pada model *Classification Tree* tanpa *Random Oversampling* dapat mengklasifikasikan kelayakan pemberian kredit sebesar 96,79%. Model *Classification Tree* dapat dengan benar mengklasifikasi debitur yang memiliki potensi lancar pada data latih sebanyak 542 debitur atau berkisar 100,00% dan dengan benar mengklasifikasi debitur yang memiliki potensi non lancar pada data latih sebanyak 0 debitur atau sebesar 0,00%. Dapat disimpulkan bahwa model memperoleh nilai akurasi sebesar 96,79%, namun model *Classification Tree* tanpa *Random Oversampling* kurang baik dalam mengklasifikasikan data debitur non lancar.

Selanjutnya dilakukan validasi model dari hasil pembentukan model pada data latih menggunakan data uji. Kinerja model *Classification Tree* tanpa *Random Oversampling* menggunakan *minsplit* 100 guna memprediksi status kelancaran dan non lancar debitur pada data uji disajikan menggunakan *confusion matrix* untuk mendapatkan hasil yang disajikan pada Tabel 6.

Diperoleh hasil dari *confusion matrix Classification Tree* tanpa *Random Oversampling* dengan proporsi 70:30, sehingga dapat dihitung tingkat akurasi, presisi, *recall*, *specificity* dan *F1-Score* model pada data uji dengan menggunakan persamaan (1)-(5), diperoleh nilai akurasi sebesar 96,67%, presisi 96,67%, *recall/sensitivity* 100,00%, *specificity* 0,00% dan *F1-Score* 98,31%. Nilai akurasi me-

Tabel 6. Klasifikasi Model *Classification Tree* tanpa *Random Oversampling* Menggunakan *Minsplit* 100 pada Data Uji

Nilai Aktual	Tidak Lancar Lancar	Nilai Prediksi	
		Tidak Lancar	Lancar
	Tidak Lancar	0	8
	Lancar	0	232

nunjukkan kemampuan pada model *Classification Tree* tanpa *Random Oversampling* dapat mengklasifikasikan kelayakan pemberian kredit sebesar 96,79%. Model *Classification Tree* dapat dengan benar mengklasifikasi debitur yang memiliki potensi lancar pada data uji sebanyak 232 debitur atau berkisar 100,00% dan dengan benar mengklasifikasi debitur yang memiliki potensi non lancar pada data uji sebanyak 0 debitur atau sebesar 0,00%. Dapat disimpulkan bahwa model memperoleh nilai akurasi sebesar 96,67%, namun model *Classification Tree* tanpa *Random Oversampling* kurang baik dalam mengklasifikasikan data debitur non lancar.

Ketimpangan nilai *recall/sensitivity* dan *specificity* dalam mengklasifikasikan debitur lancar dan non lancar diasumsikan karena adanya *imbalanced data*. Klasifikasi pada *imbalanced data* merupakan masalah utama karena dapat menyebabkan kurang tepatnya pihak kreditur dalam mengambil kebijakan kelayakan kredit yang tepat.

3.4.1. Classification Tree dengan Random Oversampling

Pemodelan menggunakan metode *Classification Tree* dengan *Random Oversampling* diawali dengan melakukan pemodelan data latih menggunakan *minsplite* 100. Pembentukan model *Classification Tree* dengan *Random Oversampling* menggunakan proporsi 70% data latih dan 30% data uji dilakukan dengan bantuan *software* R Studio. Setelah dilakukan *Random Oversampling*, jumlah data yang sebelumnya 800 data menjadi sebanyak 1.084 data. Hasil kinerja model *Classification Tree* dengan *Random Oversampling* untuk *minsplite* 100 pada data latih menggunakan *confusion matrix* disajikan pada Tabel 7.

Tabel 7. Klasifikasi Model *Classification Tree* dengan *Random Oversampling* Menggunakan *Minsplit* 100 pada Data Latih

Nilai Aktual	Tidak Lancar Lancar	Nilai Prediksi	
		Tidak Lancar	Lancar
	Tidak Lancar	542	0
	Lancar	63	479

Diperoleh hasil dari *confusion matrix Classification Tree* dengan *Random Oversampling* menggunakan proporsi 70:30, sehingga dapat dihitung tingkat akurasi, presisi, *recall*, *specificity* dan *F1-Score* model pada data latih dengan menggunakan persamaan (1)-(5), diperoleh nilai akurasi sebesar 94,19%, presisi 100,00%, *recall/sensitivity* 88,38%, *specificity* 100,00% dan *F1-Score* 93,83%. Nilai akurasi menunjukkan kemampuan pada model *Classification Tree* tanpa *Random Oversampling* dapat mengklasifikasikan kelayakan pemberian kredit sebesar 94,19%. Model *Classification Tree* dapat dengan benar mengklasifikasi debitur yang memiliki potensi lancar pada data latih sebanyak 479 debitur atau sebesar 88,38% dan dengan benar mengklasifikasi debitur yang memiliki poten-

si non lancar pada data latih sebanyak 542 debitur atau berkisar 100,00%. Jika dicermati, model *Classification Tree* dengan *Random Oversampling* dengan proporsi 70:30 tidak memiliki perbedaan nilai *recall/sensitivity* dan *specificity* yang timpang. Dapat disimpulkan bahwa model memperoleh nilai akurasi sebesar 94,19%, serta model *Classification Tree* dengan *Random Oversampling* sudah baik dalam mengklasifikasikan data debitur lancar dan non lancar.

Selanjutnya dilakukan validasi model dari hasil pembentukan model pada data latih menggunakan data uji. Kinerja model *Classification Tree* dengan *Random Oversampling* menggunakan *minsplite* 100 digunakan untuk memprediksi status kelancaran dan non lancar debitur pada data uji diperoleh hasil menggunakan *confusion matrix* yang disajikan pada Tabel 8.

Tabel 8. Klasifikasi Model *Classification Tree* dengan *Random Oversampling* Menggunakan *Minsplit* 100 pada Data Uji

Nilai Aktual	Tidak Lancar Lancar	Nilai Prediksi	
		Tidak Lancar	Lancar
	Tidak Lancar	6	2
	Lancar	24	208

Diperoleh hasil dari *confusion matrix Classification Tree* dengan *Random Oversampling* dengan proporsi 70:30, sehingga dapat dihitung tingkat akurasi, presisi, *recall*, *specificity* dan *F1-Score* model pada data latih dengan menggunakan persamaan (1)-(5), diperoleh nilai akurasi sebesar 89,17%, presisi 99,05%, *recall/sensitivity* 89,66%, *specificity* 75,00% dan *F1-Score* 94,12%. Nilai akurasi menunjukkan kemampuan pada model *Classification Tree* tanpa *Random Oversampling* dapat mengklasifikasikan kelayakan pemberian kredit sebesar 89,17%. Model *Classification Tree* dapat dengan benar mengklasifikasi debitur yang memiliki potensi lancar pada data uji sebanyak 208 debitur atau sebesar 89,66% dan dengan benar mengklasifikasi debitur yang memiliki potensi non lancar pada data uji sebanyak 6 debitur atau berkisar 75,00%. Jika dicermati, model *Classification Tree* dengan *Random Oversampling* dengan proporsi 70:30 tidak memiliki perbedaan nilai *recall/sensitivity* dan *specificity* yang timpang. Dapat disimpulkan bahwa model memperoleh nilai akurasi sebesar 96,67%, model *Classification Tree* dengan *Random Oversampling* sudah dapat mengklasifikasikan data debitur lancar dan debitur non lancar dengan baik.

3.4.2. Evaluasi Model

Evaluasi model bertujuan untuk menilai seberapa baik model mampu menghasilkan prediksi yang akurat. Evaluasi model pada data latih bertujuan untuk melihat seberapa baik model bisa mempelajari pola yang ada dalam data latih, sementara evaluasi model pada data uji bertujuan untuk melihat seberapa baik model bisa memprediksi secara akurat terhadap data. Evaluasi model dilakukan dengan menggunakan *confusion matrix* yang bertujuan untuk mengukur kinerja model dalam memprediksi data pada data latih dan data uji. *Confusion matrix* menampilkan hasil perhitungan akurasi, presisi, *recall/sensitivity*, *specificity* dan *F1-score* pada model tersebut. Hasil model *Classification Tree* tanpa *Random Oversampling* dan *Classification Tree* dengan *Random Oversampling* pada data latih ditampilkan pada Tabel 9.

Tabel 9. Perbandingan Hasil Model pada Data Latih

Model	Akurasi	Presisi	Recall	Specificity	F1-Score
Proporsi Split Data Latih 70:30					
CT	96,79%	96,79%	100,00%	0,00%	98,37%
CT + ROS	94,19%	100,00%	88,38%	100,00%	93,83%
Proporsi Split Data Latih 80:20					
CT	96,72%	96,72%	100,00%	0,00%	98,33%
CT + ROS	94,26%	100,00%	88,53%	100,00%	93,92%
Proporsi Split Data Latih 90:10					
CT	96,81%	96,81%	100,00%	0,00%	98,38%
CT + ROS	93,97%	96,09%	91,68%	96,27%	93,83%

Hasil model *Classification Tree* tanpa *Random Oversampling* dan *Classification Tree* dengan *Random Oversampling* pada data uji ditunjukkan pada Tabel 10.

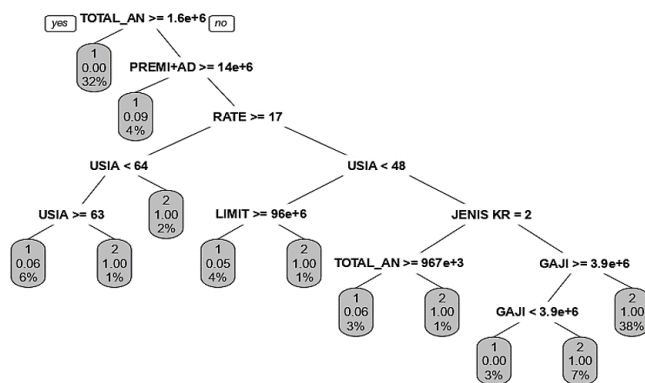
Tabel 10. Perbandingan Hasil Model pada Data Uji

Model	Akurasi	Presisi	Recall	Specificity	F1-Score
Proporsi Split Data Uji 70:30					
CT	96,67%	96,67%	100,00%	0,00%	98,31%
CT + ROS	89,17%	99,05%	89,66%	75,00%	94,12%
Proporsi Split Data Uji 80:20					
CT	96,87%	96,88%	100,00%	0,00%	98,42%
CT + ROS	88,75%	98,58%	89,68%	60,00%	93,92%
Proporsi Split Data Uji 90:10					
CT	96,25%	96,25%	100,00%	0,00%	98,09%
CT + ROS	91,25%	97,30%	93,51%	33,33%	95,37%

Diketahui perbandingan performa klasifikasi model terbaik dari model *Classification Tree* tanpa *Random Oversampling* dan *Classification Tree* dengan *Random Oversampling*. Setiap model *Classification Tree* menghasilkan nilai akurasi, presisi, *recall* dan *F1-score* yang sangat baik, namun tidak dengan nilai *specificity*. Nilai *specificity* pada model *Classification Tree* tanpa *Random Oversampling* sebesar 0,00%, hal ini berarti *Classification Tree* tanpa *Random Oversampling* tidak dapat mengklasifikasikan data debitur non lancar dengan tepat. Setelah dilakukan *Random Oversampling*, nilai *specificity* yang diperoleh dari model *Classification Tree* sudah sangat baik pada data latih dan telah mengalami sedikit peningkatan nilai *specificity* pada data uji. Sehingga dapat disimpulkan bahwa model *Classification Tree* tanpa *Random Oversampling* dapat mengklasifikasikan debitur lancar dengan baik, namun model tidak dapat mengklasifikasikan debitur non lancar dengan baik. Hal ini terjadi karena *imbalanced data* yang menyebabkan kelas non lancar pada kolektibilitas diabaikan oleh model, sehingga diperlukan teknik *Random Oversampling* untuk mengatasi permasalahan tersebut.

Pembentukan *Classification Tree* dari model terbaik yaitu *Classification Tree* dengan *Random Oversampling* dilakukan dengan menggunakan program R Studio. Gambar 3 menunjukkan hasil struktur model klasifikasi dari debitur lancar dan non lancar salah satu bank. Berdasarkan Gambar 3, aturan pengklasifikasian menggunakan *Classification Tree* disajikan pada Tabel 11.

Berdasarkan Gambar 3 dan Tabel 11 dapat disimpulkan bahwa dari 10 variabel independen yang digunakan dalam penelitian ini, variabel yang tidak terlalu mempengaruhi hasil prediksi dalam mengklasifikasikan debitur lancar dan non lancar pada bank tersebut adalah variabel tenor, jenis kebutuhan dan instansi. Se-



Gambar 3. Classification Tree dengan Random Oversampling Menggunakan Proporsi 70:30

Tabel 11. Aturan Pengklasifikasian Menggunakan Classification Tree

Status	Karakteristik
Lancar	Total angsuran < Rp1.600.000, premi < Rp14.000.000, rate ≥ 17 dan usia ≥ 64 tahun.
	Total angsuran < Rp1.600.000, premi < Rp14.000.000, rate ≥ 17 dan usia < 63 tahun.
	Total angsuran < Rp1.600.000, premi < Rp14.000.000, rate < 17, usia < 48 tahun dan limit < Rp96.000.000.
	Premi < Rp14.000.000, rate < 17, usia ≥ 48 tahun, jenis kredit top up dan total angsuran < Rp967.000.
Non Lancar	Total angsuran < Rp1.600.000, premi < Rp14.000.000, rate < 17, usia ≥ 48 tahun dan jenis kredit non top up.
	Total angsuran ≥ Rp1.600.000.
	Total angsuran < Rp1.600.000 dan premi ≥ Rp14.000.000.
	Total angsuran < Rp1.600.000, premi < Rp14.000.000, rate ≥ 17 dan usia 63 tahun.
Non Lancar	Total angsuran < Rp1.600.000, premi < Rp14.000.000, rate < 17, usia < 48 tahun dan limit ≥ Rp96.000.000.
	Premi < Rp14.000.000, rate < 17, usia ≥ 48 tahun, jenis kredit top up dan total angsuran berkisar antara Rp967.000-Rp1.600.000.

dangkan, variabel total angsuran menjadi *variable importance* yang dapat diartikan sebagai variabel yang memiliki pengaruh terbesar, dikarenakan variabel yang menjadi *root node* (node akar) adalah variabel total angsuran.

4. Kesimpulan

Penerapan *Classification Tree* menggunakan *Random Oversampling* lebih baik jika dibandingkan dengan *Classification Tree* tanpa *Random Oversampling*. Hal ini dikarenakan model *Classification Tree* tanpa *Random Oversampling* memperoleh nilai *specificity* sebesar 0,00%, sehingga tidak dapat mengklasifikasikan data debitur non lancar dengan tepat. *Classification Tree* menggunakan *Random Oversampling* dengan proporsi split data 70:30 dapat mengklasifikasikan debitur non lancar sebesar 75,00%. *Classification Tree* dengan *Random Oversampling* menggunakan proporsi 70:30 memiliki nilai akurasi sebesar 89,17%. Berdasarkan pohon

dari model terbaik, dapat diartikan bahwa variabel yang memiliki pengaruh paling besar dalam pengklasifikasian status debitur adalah variabel total angsuran, sedangkan variabel yang tidak berpengaruh adalah variabel tenor, jenis kebutuhan dan instansi. Dengan demikian, analisis kelayakan kredit menggunakan metode *Classification Tree* dengan *Random Oversampling* dapat memprediksi dengan baik sebesar 89,17%, sehingga hasil tersebut dapat menjadi acuan dalam pengambilan keputusan untuk seleksi calon debitur.

Kontribusi Penulis. Lo Mei Ly Vebriyanti: analisis, visualisasi, dan penulisan naskah. Shantika Martha: Konseptualisasi, tinjauan penulisan, dan administrasi Project. Wirda Andani: Konseptualisasi, metodologi, tinjauan penulisan, dan supervisi. Setyo Wira Rizki: Konseptualisasi, metodologi, dan supervisi. Semua penulis telah membaca dan menyetujui versi manuskrip yang diterbitkan.

Ucapan Terima Kasih. Para penulis menyampaikan terima kasih kepada editor dan reviewer atas pembacaan yang cermat, kritik yang mendalam, dan rekomendasi yang praktis untuk meningkatkan kualitas tulisan ini.

Pembiayaan. Penelitian ini mendapatkan dukungan anggaran dari DIPA PNBPN Universitas Tanjungpura dengan kontrak SP DIPA-023.17.2.677517/2022 tahun anggaran 2022.

Konflik Kepentingan. Para penulis menyatakan tidak ada konflik kepentingan yang terkait dengan artikel ini.

Referensi

- [1] K. Fatmawati, A. P. Windarto, Solikhun, and M. R. Lubis, "Analisa SPK dengan Metode AHP dalam Menentukan Faktor Konsumen dalam Melakukan Kredit Barang," *KOMIK (Konferensi Nasional Teknologi Informatika dan Komputer)*, vol. 1, no. 1, pp. 314–321, Oct. 2017, doi: <http://dx.doi.org/10.30865/komik.v1i1.515>.
- [2] N. Iriadi and H. Leidiyana, "Prediksi Pinjaman Kredit dengan Support Vector Machine dan K-Nearest Neighbors pada Koperasi Serba Usaha," *Jurnal Penelitian Ilmu Komputer, System Embedded & Logic*, vol. 01, no. 02, pp. 115–124, 2013.
- [3] N. L. Hanun and A. U. Zailani, "Penerapan Algoritma Klasifikasi Random Forest untuk Penentuan Kelayakan Pemberian Kredit di Koperasi Mitra Sejahtera," *Infotech: Journal of Technology Information*, vol. 6, no. 1, pp. 7–14, Jun. 2020, doi: <http://dx.doi.org/10.37365/jti.v6i1.61>.
- [4] A. A. Arifiyanti and E. D. Wahyuni, "SMOTE: Metode menyeimbangkan Kelas pada Klasifikasi Data Mining," *Jurnal Sistem Informasi dan Telematika (Telekomunikasi, Multimedia dan Informatika)*, vol. 15, no. 1, pp. 34–39, Feb. 2020, doi: <http://dx.doi.org/10.36448/jsit.v13i1.2539>.
- [5] J. Prasetya, "Penerapan Klasifikasi Naive Bayes dengan Algoritma Random Oversampling dan Random Undersampling pada Data Tidak Seimbang Cervical Cancer Risk Factors," *Leibniz: Jurnal Matematika*, vol. 2, no. 2, pp. 11–22, Jul. 2022, doi: <http://dx.doi.org/10.59632/leibniz.v2i2.173>.
- [6] R. D. Fitriani, H. Yasin, and Tarno, "Penanganan Klasifikasi Kelas Data Tidak Seimbang dengan Random Oversampling Pada Naive Bayes (Studi Kasus: Status Peserta KB IUD di Kabupaten Kendal)," *Jurnal Gaussian*, vol. 10, no. 1, pp. 11–20, Feb. 2021, doi: <http://dx.doi.org/10.14710/j.gauss.v10i1.30243>.
- [7] Z. W. Mardika, M. A. Mukid, and H. Yasin, "Pembentukan Pohon Klasifikasi Biner dengan Algoritma Cart (Classification and Regression Trees) (Studi Kasus: Kredit Macet di PD. BPR-BKK Purwokerto Utara)," *Jurnal Gaussian*, vol. 5, no. 3, pp. 583–592, 2016, doi: <https://doi.org/10.14710/j.gauss.5.3.583-592>.
- [8] Z. K. Malta and Sutikno, "Analisis Karakteristik Tingkat Kesejahteraan di Kota Surabaya Menggunakan Metode Pohon Klasifikasi," *Jurnal Sains dan Seni ITS*, vol. 8, no. 2, pp. D424–D431, 2019, doi: <http://dx.doi.org/10.12962/j23373520.v8i2.46867>.
- [9] N. Handayani, H. Wahyono, J. Trianto, and D. S. Permana, "Prediksi Tingkat Risiko Kredit dengan Data Mining Menggunakan Algoritma Decision Tree C.45," *JURIKOM (Jurnal Riset Komputer)*, vol. 8, no. 6, p. 198–204, Dec. 2021, doi: <http://dx.doi.org/10.30865/jurikom.v8i6.3643>.
- [10] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 4, no. 2, p. 113, Dec. 2018, doi: <http://dx.doi.org/10.26418/jp.v4i2.27526>.
- [11] R. K. Dinata, Safwandi, N. Hasdina, and N. Azizah, "Analisis K-Means Clustering pada Data Sepeda Motor," *Informatika Journal*, vol. 5, no. 1, pp. 10–17, 2020, doi: <https://doi.org/10.19184/isj.v5i1.17071>.
- [12] D. Normawati and S. A. Prayogi, "Implementasi Naive Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, vol. 5, no. 2, pp. 697–711, Sep. 2021, doi: <http://dx.doi.org/10.30645/j-sakti.v5i2.369>.
- [13] M. D. Muafa and L. Iswari, "Pengembangan Aplikasi Berbasis Web dengan Rshiny untuk Data Klasifikasi Menggunakan Metode Naive Bayes," *AUTOMATA*, vol. 3, no. 1, 2022.
- [14] A. Maksun and D. Swanjaya, "Perbandingan Antara Metode Decision Tree Dan Support Vector Machine Pada Model Rekomendasi Mobil Bekas," in *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, 2021, pp. 167–173.
- [15] B. Purnama, *Pengantar Machine Learning*. Bandung: Informatika Bandung, 2019.
- [16] R. Amelia, Indahwati, Erfiani, A. Fitrianto, and A. Rizki, "Komparasi Teknik Undersampling dan Oversampling pada Regresi Logistik Biner dalam Menduga Faktor Determinan Berhenti Merokok Penduduk Lanjut Usia," *Jurnal TIMES*, vol. 10, no. 2, pp. 1–11, Dec. 2021.
- [17] A. N. Kholifah and N. Insani, "Analisis Klasifikasi pada Nasabah Kredit Koperasi X Menggunakan Decision Tree C4.5 dan Naive Bayes," *Jurnal Kajian dan Terapan Matematika*, vol. 5, no. 6, pp. 1–8, 2016.