

Penerapan SMOTE untuk Meningkatkan Kinerja Klasifikasi Penilaian Kredit

Muhammad Ibnu Choldun Rachmatullah*

Sekolah Vokasi, D3 Manajemen Informatika, Universitas Logistik dan Bisnis Internasional, Bandung, Indonesia

Email: muhammadibnucholdun@ulbi.ac.id

Email Penulis Korespondensi: muhammadibnucholdun@ulbi.ac.id

Submitted 30-01-2023; Accepted 27-02-2023; Published 27-02-2023

Abstrak

Teknik pembelajaran mesin/machine learning banyak digunakan di berbagai bidang dan data diperlukan untuk melatih model. Namun, distribusi kelas di sebagian besar kumpulan data dunia nyata ternyata tidak selalu seimbang, bahkan bisa sangat tidak seimbang. Jika data tidak seimbang, kinerja pengklasifikasi sangat tergantung pada kelas mayoritas sehingga menimbulkan permasalahan dalam penentuan kinerja. Salah teknik yang dapat diterapkan untuk menyeimbangkan data adalah Synthetic Minority Oversampling Technique (SMOTE). Pada penelitian ini, SMOTE diterapkan pada penilaian kredit (credit scoring) menggunakan dataset German Credit Data (GCD), dan kemudian dilakukan klasifikasi menggunakan empat metode klasifikasi, yaitu: random forest, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), dan Multilayer Perceptron (MLP). Ukuran kinerja penerapan SMOTE pada masing-masing metode pengklasifikasi diukur dengan menggunakan: recall, presisi, F1-Score, dan AUC. Nilai akurasi juga diukur untuk melihat apakah akurasi cocok untuk menghitung kinerja pada dataset yang tidak seimbang. Berdasarkan ukuran kinerja: recall, presisi, F1-Score, dan AUC, maka penerapan SMOTE pada dataset dan kemudian diklasifikasi menggunakan empat metode menunjukkan peningkatan kinerja. Ukuran kinerja tertinggi: recall = 82,00% dengan metode random forest, presisi = 75,35 dengan metode MLP, F1-Score = 76,93% dengan metode MLP, dan AUC = 0,832 dengan metode random forest. Nilai akurasi setelah SMOTE sedikit turun pada metode random forest, KNN, dan SVM, sedangkan dengan MLP nilai akurasinya sedikit meningkat. Kontribusi dari penelitian ini adalah untuk menunjukkan perlunya penanganan data tidak seimbang untuk meningkatkan kinerja algoritma pengklasifikasi khususnya untuk dataset penilaian kredit.

Kata Kunci: Credit scoring; SMOTE; Random Forest; KNN; SVM; MLP

Abstract

Machine learning techniques are widely used in various fields and data is needed to train models. However, the distribution of classes in most real-world datasets turns out to be not always balanced, and can be very imbalanced. If the data is imbalanced, the performance of the classifier is highly dependent on the majority class, causing problems in determining performance. One technique that can be applied to balance the data is the Synthetic Minority Oversampling Technique (SMOTE). SMOTE is applied to credit scoring using the German Credit Data (GCD) dataset, and then classified using four classification methods, namely: random forest, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Multilayer Perceptron (MLP). The performance measure of implementing SMOTE in each classifier method is measured using: recall, precision, F1-Score, and AUC. Accuracy values are also measured to see if the accuracy is suitable for calculating performance on imbalanced datasets. Based on performance measures: recall, precision, F1-Score, and AUC, then applying SMOTE to the dataset and then classifying it using four methods shows an increase in performance. The highest performance measure: recall = 82.00% with the random forest method, precision = 75.35 with the MLP method, F1-Score = 76.93% with the MLP method, and AUC = 0.832 with the random forest method. The accuracy value after SMOTE slightly decreased in the random forest, KNN, and SVM methods, while with MLP the accuracy value increased slightly. The contribution of this research is to show the need for imbalanced data handling to improve the performance of classifier algorithms, especially for credit rating datasets.

Keywords: Credit scoring; SMOTE; Random Forest; KNN; SVM; MLP

1. PENDAHULUAN

Teknik pembelajaran mesin (*machine learning*) banyak digunakan dalam berbagai aplikasi seperti perbankan, bioinformatika, keuangan, epidemiologi, pemasaran, diagnosis medis, analisis data meteorologi, dan di berbagai bidang lainnya [1]. Dalam domain ini, data diperlukan untuk melatih model. Namun, distribusi kelas di sebagian besar kumpulan data dunia nyata ternyata tidak selalu seimbang, bahkan bisa sangat tidak seimbang. Permasalahan data tidak seimbang ini menimbulkan tantangan besar bagi algoritme *machine learning* standar. Tantangan dataset yang tidak seimbang dalam masalah klasifikasi muncul ketika jumlah sampel dalam satu kelas jauh lebih banyak daripada kelas lainnya. Dalam keadaan seperti itu, kinerja pengklasifikasi biasanya tergantung kelas mayoritas dalam hal prediksi dan kurang memperhatikan kelas minoritas.

Kemajuan pesat dalam teknologi telah meningkatkan jumlah pengguna yang memunculkan kumpulan data yang lebih besar, misal data yang berkaitan dengan data penilaian kredit (*Credit Scoring*). Salah satu perhatian utama lembaga keuangan dan bank adalah pengklasifikasian pelamar yang layak diberikan kredit maupun pelamar yang tidak layak diberikan kredit. Prediksi yang tepat dari pelamar tersebut dapat menghasilkan penghematan pendapatan yang besar bagi lembaga keuangan maupun bank. Dalam beberapa tahun terakhir, penilaian kredit mendapat perhatian lebih besar dari komunitas data mining karena implikasinya yang sangat besar dalam menghasilkan pendapatan, pengurangan risiko keuangan, evaluasi risiko kredit, dan mempertahankan arus kas.

Secara tradisional, penilaian kredit dikategorikan menjadi dua jenis berdasarkan data yang digunakan dan tugas yang diberikan yaitu penilaian aplikasi dan penilaian perilaku. Penilaian aplikasi akan memperkirakan kemungkinan

pelamar gagal bayar untuk beberapa interval waktu tertentu. Biasanya, data yang digunakan dalam pelatihan model ini berisi informasi demografis dan keuangan pelamar bersama dengan status baik atau buruk mereka yang direkam untuk interval waktu lainnya. Penilaian aplikasi pemohon dilakukan sebelum pinjaman diberikan kepada pemohon. Penilaian kedua adalah penilaian perilaku dan dilakukan ketika pinjaman telah diberikan kepada pemohon. Penilaian perilaku juga memperkirakan kemungkinan pelamar gagal bayar pada interval waktu tertentu. Data yang digunakan untuk penilaian ini didasarkan pada kinerja pengembalian pinjaman oleh nasabah dengan status baik atau buruk. Dengan penilaian ini, lembaga keuangan dapat memantau perilaku pelanggan yang selanjutnya dapat mengarah pada pengambilan keputusan tentang status mereka. Untuk setiap lembaga keuangan untuk menghasilkan laba maksimum, diharapkan dapat memprediksi pelanggan secara akurat untuk interval waktu yang bervariasi (misalnya, bulan ke-2, bulan ke-4, bulan ke-6, dst). Prediksi pelanggan yang akurat ini, misalnya dapat menandai pelanggan dengan risiko tinggi dan memungkinkan lembaga keuangan untuk mengambil tindakan pencegahan yang diperlukan yang dapat menyelamatkannya dari kerugian besar. Klasifikasi dataset penilaian kredit adalah tugas yang sulit dalam membuat keputusan penting bagi pelanggan untuk diberikan atau ditolak pengajuan pinjamannya. Penilaian kredit adalah masalah mendasar yang dihadapi oleh komunitas data mining dan riset operasional [2]. Model penilaian kredit dikembangkan untuk klasifikasi pelanggan ke pelanggan yang baik atau pelanggan yang buruk. Namun, kumpulan data ini biasanya sangat tidak seimbang dengan lebih banyak pelamar yang baik dibandingkan dengan pelamar yang buruk. Secara konvensional, untuk klasifikasi yang tidak seimbang, algoritma dibiarkan ke kelas dengan jumlah pengamatan lebih banyak dengan memprediksi akurasi keseluruhan. Salah satu usaha untuk meningkatkan prediksi yang benar yaitu *True Positive Rate* (TPR) dari pengklasifikasi, diterapkan teknik resampling yang meliputi teknik *under-sampling* atau teknik *over-sampling* dataset. *Under-sampling* adalah proses mengubah probabilitas sebelumnya untuk kelas mayoritas, sedangkan *over-sampling* meningkatkan jumlah kelas minoritas [3] [4] [5]. Untuk teknik *under-sampling* dan teknik *over-sampling* dataset, banyak teknik telah dirancang yang bertujuan untuk meningkatkan tingkat prediksi untuk kelas. Secara umum, teknik *over-sampling* ternyata memiliki kinerja yang jauh lebih baik daripada teknik *under-sampling* karena informasi aslinya tidak hilang seperti halnya pada *under-sampling*. Penelitian sebelumnya berkaitan dengan pengklasifikasian penilaian kredit telah dilakukan banyak peneliti [6] [7] [8] [9] [10] [11] [12] [13]. Nguyen dan Huynh melakukan penelitian mengenai penilaian kredit perusahaan dengan menerapkan SMOTE dan pengklasifikasi yang dibandingkan adalah pengklasifikasi tunggal (*Logistic Regression*, *Decision Tree*, *Support Vector Machine*, dan *Multi-Layer Perceptron*) dan pengklasifikasi *ensemble* (baik homogen maupun heterogen) [10]. Lenka dkk. melakukan penelitian pengembangan model penilaian kredit dengan mengadakan seleksi fitur terlebih dahulu dan kemudian menerapkan SMOTE untuk menyeimbangkan kelas baru kemudian diproses dengan algoritma pengklasifikasi [11]. Li dkk. melakukan penelitian untuk menilai kredit dengan menerapkan transformasi fitur dan menggunakan model *ensemble*. Model tersebut dirancang dengan menggunakan pembobotan *factorization machine* (FM) dan *deep neural networks* (DNN) [12]. Khatir dan Bee melakukan penelitian untuk memprediksi kelayakan kredit pelanggan dengan melakukan seleksi fitur, penyeimbangan data, dan kemudian diklasifikasi dengan lima pengklasifikasi dan kemudian dibandingkan kinerjanya [13]. Pada penelitian sekarang ini bertujuan untuk menyeimbangkan dataset dengan menerapkan teknik Synthetic Minority Oversampling Technique (SMOTE) [14] [15] [16]. Setelah data seimbang dataset diklasifikasi dengan menggunakan empat metode pengklasifikasi yaitu: *Random Forest*, *K-Nearest Neighbor* (KNN), *Support Vector Machine* (SVM), dan *Multilayer Perceptron* (MLP) yang diterapkan pada dataset penilaian kredit tertentu. Jadi perbedaan penelitian ini dengan penelitian sebelumnya yang telah disebutkan adalah menggunakan algoritma Random Forest sebagai salah satu pengklasifikasinya.

2. METODOLOGI PENELITIAN

Metode penelitian terdiri dari tahapan-tahapan penelitian sebagai berikut, yang masing-masing tahapan diuraikan pada subbab berikutnya:

- Menyiapkan dataset
- Menyeimbangkan data dengan SMOTE
- Menentukan metode *split* data
- Menentukan metode pengklasifikasi
- Menghitung kinerja

2.1 Menyiapkan Dataset

Dataset yang digunakan dalam penelitian ini adalah data penilaian kredit yang diambil dari *UCI Machine Learning Repository*, yaitu dataset Statlog (*German Credit Data*) dengan menggunakan dataset yang sudah berbentuk numerik semua [17]. Dataset ini terdiri dari 1000 data, dengan 24 atribut input yang bersifat numerik semua, dan satu atribut output yang berisi apakah penilaian kredit “1=bagus” atau “2=buruk”. Pada dataset tersebut sebanyak 700 data tergolong “bagus” dan sebanyak 300 data tergolong “buruk”.

2.2 Menyeimbangkan Data dengan SMOTE

Tujuan dari penyeimbangan dataset adalah agar supaya setiap kelas mempunyai frekuensi/jumlah data yang sama. Pada penelitian ini untuk menyeimbangkan data menggunakan teknik yang disebut dengan *Synthetic Minority*

Oversampling Technique (SMOTE) yang pertama kali diusulkan oleh Chawla dkk [14] [15] [19]. Dengan proses SMOTE maka yang sebelumnya jumlah data antara kelas mayoritas dengan kelas minoritas tidak seimbang menjadi seimbang.

2.3 Menentukan Metode *Split Data*

Metode pemisahan data yang digunakan adalah *cross-validation*. *Cross-validation* merupakan salah satu metode yang berbasis statistik untuk menilai dan membandingkan algoritma pembelajaran dengan membagi data menjadi dua bagian: satu digunakan untuk mempelajari/*training* model dan yang sisanya difungsikan untuk memvalidasi atau menguji model [18] [19]. Pada *cross-validation*, data *training*/pelatihan dan *testing*/pengujian harus bertukar secara berurutan sehingga semua titik data mempunyai kemungkinan untuk diuji atau divalidasi. Formula paling umum dari *cross-validation* sering disebut *k-fold cross-validation*. Pada *k-fold cross-validation*, data dibagi-bagi menjadi *k* bagian berukuran sama. Pada data mining dan pembelajaran mesin, *10-fold cross-validation* adalah bentuk yang paling populer dan sering digunakan.

2.4 Menentukan metode pengklasifikasi

Dalam penelitian ini ada empat metoda pengklasifikasi yang digunakan, yaitu: *Random Forest* (RF), *K Nearest Neighbor* (KNN), *Support Vector Machine* (SVM), dan *Multilayer Perceptron* (MLP).

2.4.1 *Random Forest*

Decision Tree memberikan teknik mengklasifikasikan data dengan menghasilkan struktur seperti pohon. *Node* internal pohon mewakili pilihan (biner) untuk setiap atribut, sedangkan cabang pohon menandakan hasil dari pilihan yang diinginkan. Dalam beberapa tahun terakhir, banyak jenis pohon keputusan telah diperkenalkan oleh para peneliti, dan salah satu pengklasifikasi yang berbentuk pohon yang paling banyak digunakan adalah *Random Forest* (RF). RF mewakili kumpulan pohon yang diproduksi untuk menghindari risiko ketidakstabilan dan meminimalkan kemungkinan pelatihan sampel yang berlebihan. Pohon-pohon ini juga dibuat untuk mengurangi *overfitting* menggunakan teknik pemangkasan [20] [21]. *Random forest* ini secara progresif mengurangi node tanpa mengganggu kinerja *classifier* secara keseluruhan.

2.4.2 *K-Nearest Neighbor* (KNN)

K Nearest Neighbor adalah salah satu algoritma *machine learning* paling sederhana berdasarkan teknik *supervised learning*. Algoritma KNN mengasumsikan kesamaan antara data baru dan data yang ada dan memasukkan data baru ke dalam kategori yang paling mirip dengan kategori yang ada. Algoritma KNN menyimpan semua data yang ada dan mengklasifikasikan titik data baru berdasarkan kesamaan [22]. Artinya ketika data baru muncul maka dapat dengan mudah diklasifikasikan ke dalam kategori yang cocok dengan menggunakan algoritma KNN. KNN sering juga disebut algoritma *lazy learner* karena tidak langsung belajar dari set pelatihan melainkan menyimpan dataset dan pada saat klasifikasi, kemudian melakukan tindakan pada dataset. Algoritma KNN pada fase pelatihan hanya menyimpan dataset dan ketika mendapatkan data baru, kemudian mengklasifikasikan data tersebut ke dalam kategori yang sangat mirip dengan data baru.

2.4.3 *Support Vector Machine* (SVM)

Support Vector Machine atau SVM adalah salah satu teknik pembelajaran terawasi yang paling populer, yang dapat digunakan dalam permasalahan klasifikasi maupun permasalahan regresi. Tujuan dari teknik SVM adalah untuk menyediakan garis atau batas keputusan paling baik sehingga ruang *n*-dimensi dapat dipisahkan ke dalam kelas dan dapat dengan mudah menempatkan titik data baru ke dalam kategori yang tepat atau sesuai. Batas keputusan yang paling baik ini disebut *hyperplane*. SVM memilih titik/vektor ekstrim sehingga mempermudah dalam membuat *hyperplane*. Kasus yang ekstrim ini disebut sebagai support vector, dan karenanya tekniknya disebut sebagai *Support Vector Machine* [23].

2.4.4 *Multilayer Perceptron* (MLP)

Multilayer Perceptron (MLP) adalah salah satu bentuk jaringan saraf tiruan (*Artificial Neural Network*) yang terhubung penuh. MLP terdiri dari paling sedikit tiga lapisan yaitu : lapisan input, *hidden layer*, dan lapisan output. Kecuali simpul pada input, setiap simpul adalah *neuron* yang menggunakan fungsi aktivasi nonlinier. MLP menggunakan teknik pembelajaran terawasi yang disebut *backpropagation* untuk *training*. Jumlah *layer* yang minimal tiga dan mempunyai fungsi aktivasi yang non-linier ini membedakan MLP dengan perceptron linier [24].

2.5 Menghitung Kinerja

Untuk menghitung kinerja metode pengklasifikasi pada data yang seimbang kurang cocok jika menggunakan ukuran akurasi, karena nilai akurasi sangat tergantung pada kelas data mayoritas, padahal ternyata kelas minoritas sebenarnya yang seharusnya lebih mendapatkan perhatian. Walaupun bukan ukuran kinerja yang cocok untuk klasifikasi data tidak seimbang, nilai akurasi tetap akan dihitung untuk melihat perubahan nilai akurasi antara sebelum dan sesudah penerapan SMOTE. Untuk mengukur kinerja algoritma klasifikasi pada data tidak seimbang, empat metrik evaluasi digunakan. Metrik evaluasi ini didasarkan pada *confusion matrix* [25] [26] seperti yang ditunjukkan pada tabel 1. Matriks ini terdiri dari dua baris dan dua kolom yaitu: TP (*True Positive*), TN (*True Negative*), FP (*False Positive*), dan FN (*False Negative*).

Tabel 1. *Confusion matrix*

Kelas aktual	Kelas hasil prediksi	
	Positif	Negatif
Positif	TP	TN
Negatif	FP	FN

Rumus dari ukuran kinerja yang digunakan adalah sebagai berikut:

$$Recall = \frac{TP}{TP+FN} \quad (1)$$

$$Presisi = \frac{TP}{TP+FP} \quad (2)$$

$$F1 - Score = 2 * \frac{(Recall * Presisi)}{(Recall + Presisi)} \quad (3)$$

Area di bawah kurva ROC (*Receiver Operating Characteristic*) atau *Area Under Curve* (AUC) adalah metrik evaluasi lain yang digunakan untuk mengukur kinerja kumpulan data yang tidak seimbang [2]. ROC adalah kurva dua dimensi yang mewakili kompromi antara tingkat *True Positive* dan *False Positive*. Sedangkan area di bawah kurva ROC atau AUC digunakan untuk menilai keakuratan pengklasifikasi. Pengklasifikasi yang memberikan nilai AUC lebih tinggi artinya mempunyai kinerja yang lebih baik. Ukuran akurasi juga dihitung sebagai tambahan dengan rumus sebagai berikut:

$$Akurasi = \frac{TP+TN}{TP+FP+FN+TN} \quad (4)$$

3. HASIL DAN PEMBAHASAN

Dataset German Credit Data sebelum dilakukan penyeimbangan data, frekuensi yang dimiliki masing-masing kelas kurang seimbang, sedangkan setelah dilakukan proses SMOTE frekuensi untuk masing-masing kelas menjadi seimbang, seperti disajikan pada tabel 2.

Tabel 2. Frekuensi kelas sebelum dan sesudah SMOTE

Kelas	Sebelum SMOTE	Setelah SMOTE
1 ("Bagus")	700	700
2 ("Buruk")	300	700

Dari tabel 2 dapat dilihat bahwa frekuensi masing-masing kelas sebelum SMOTE adalah 700 dan 300, sedangkan setelah SMOTE frekuensi kedua kelas menjadi sama yaitu 700. Setelah dilakukan proses SMOTE maka diterapkan empat metode pengklasifikasi: KNN, *Random Forest*, SVM, dan MLP dan kemudian dihitung ukuran kinerjanya yaitu: *recall*, *presisi*, *F1-Score*, dan AUC. Dalam penelitian ini juga dihitung nilai akurasi untuk menunjukkan bahwa penggunaan ukuran akurasi kurang cocok diterapkan pada data tidak seimbang.

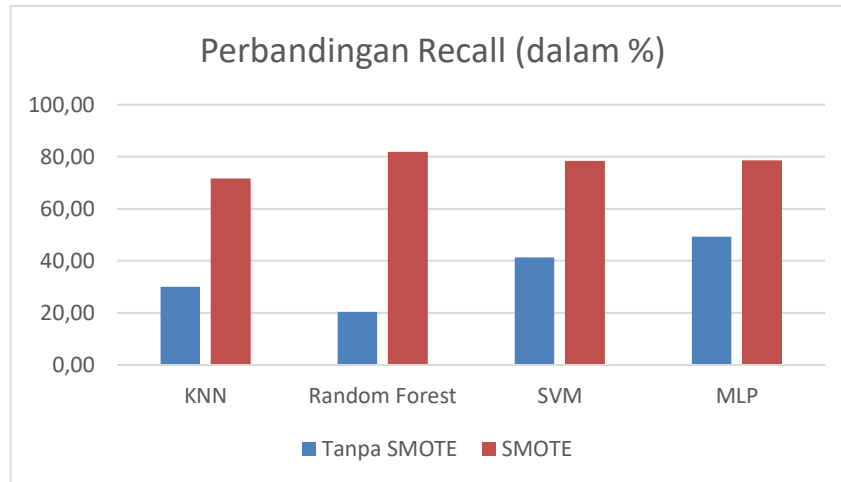
3.1 Recall

Perhitungan *recall* menggunakan rumus 1. Perhitungan ini diterapkan pada keempat metode pengklasifikasi yang digunakan dan hasilnya diperoleh seperti pada tabel 3. Pada tabel ini diperlihatkan hasil yang diperoleh jika dataset tanpa proses SMOTE dan jika pada dataset diterapkan SMOTE.

Tabel 3. Perbandingan recall

Metode	Tanpa SMOTE (%)	SMOTE (%)	Kenaikan (%)
KNN	30,00	71,57	41,57
Random Forest	20,33	82,00	61,67
SVM	41,33	78,43	37,10
MLP	49,33	78,57	29,24

Jika disajikan dengan menggunakan grafik seperti tersaji pada gambar 1, hasil *recall* untuk keempat metode pengklasifikasi yang digunakan adalah sebagai berikut:



Gambar 1. Perbandingan Recall

Dari tabel 3 dan gambar 1 dapat dilihat bahwa nilai *recall* untuk keempat metode pengklasifikasi yang digunakan jauh meningkat setelah dilakukan proses SMOTE pada dataset, bahkan dengan menggunakan metode *Random Forest* nilai *recall* naik dari 20,33% menjadi 82%, jadi ada kenaikan sekitar 59,67%.

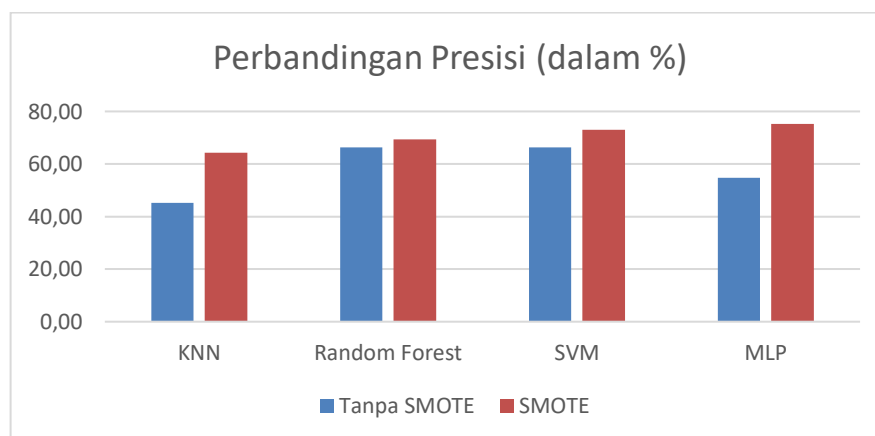
3.2 Presisi

Perhitungan presisi menggunakan rumus 2. Perhitungan ini diterapkan pada keempat metode pengklasifikasi yang digunakan dan hasilnya diperoleh seperti pada tabel 4. Pada tabel ini diperlihatkan hasil yang diperoleh jika dataset tanpa proses SMOTE dan jika pada dataset diterapkan SMOTE.

Tabel 4. Perbandingan presisi

Metode	Tanpa SMOTE (%)	SMOTE (%)	Kenaikan (%)
KNN	45,19	64,37	19,18
Random Forest	66,30	69,27	2,97
SVM	66,29	73,11	6,82
MLP	54,75	75,35	20,60

Jika disajikan dengan menggunakan grafik seperti tersaji pada gambar 2, hasil presisi untuk keempat metode pengklasifikasi yang digunakan adalah sebagai berikut:



Gambar 2. Perbandingan Presisi

Dari tabel 4 dan gambar 2 dapat dilihat bahwa nilai presisi untuk keempat metode pengklasifikasi yang digunakan meningkat setelah dilakukan proses SMOTE pada dataset. Dengan menggunakan MLP dan KNN kenaikan presisinya cukup tinggi, sedangkan dengan metode random forest dan SVM terjadi sedikit kenaikan. Nilai presisi paling tinggi setelah SMOTE dicapai oleh metode MLP yaitu 75,35%.

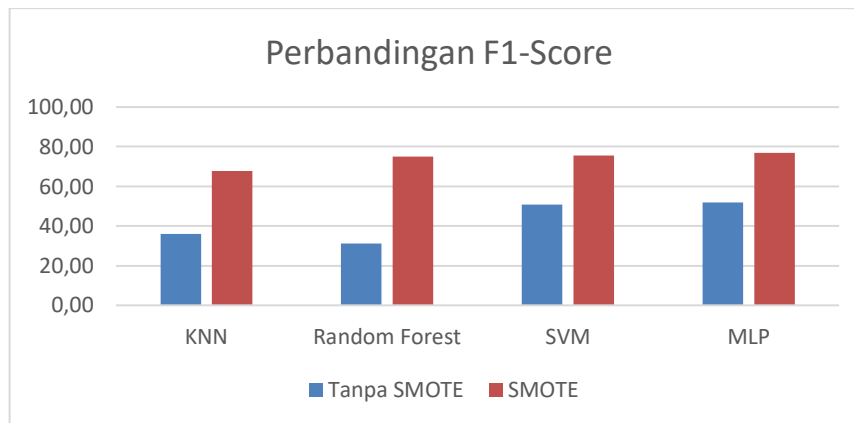
3.3 F1-Score

Perhitungan *F1-Score* menggunakan rumus 3. Perhitungan ini diterapkan pada keempat metode pengklasifikasi yang digunakan dan hasilnya diperoleh seperti pada tabel 5. Pada tabel ini diperlihatkan hasil yang diperoleh jika dataset tanpa proses SMOTE dan jika pada dataset diterapkan SMOTE.

Tabel 5. Perbandingan F1-Score

Metode	Tanpa SMOTE (%)	SMOTE (%)	Kenaikan (%)
KNN	36,06	67,78	31,72
Random Forest	31,12	75,10	43,98
SVM	50,92	75,68	24,76
MLP	51,90	76,93	25,03

Jika disajikan dengan menggunakan grafik seperti tersaji pada gambar 3, hasil *F1-Score* untuk keempat metode pengklasifikasi yang digunakan adalah sebagai berikut:



Gambar 3. Perbandingan F1-Score

Dari tabel 5 dan gambar 3 dapat dilihat bahwa nilai *F1-Score* untuk keempat metode pengklasifikasi yang digunakan jauh meningkat setelah dilakukan proses SMOTE pada dataset. Dengan menggunakan random forest diperoleh kenaikan *F-Score* yang paling tinggi dibanding tiga metode lainnya. Nilai *F1-Score* paling tinggi setelah SMOTE dicapai oleh metode MLP yaitu 76,93%.

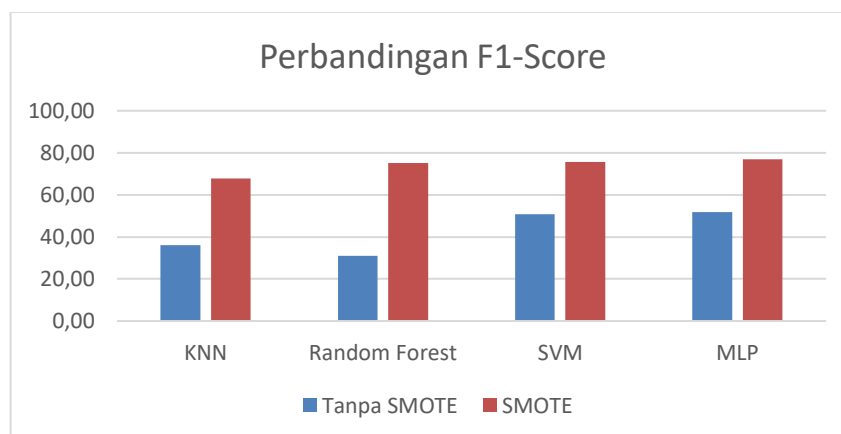
3.4 Area Under Curve (AUC)

Hasil AUC pada keempat metode pengklasifikasi yang digunakan didapatkan hasil seperti pada tabel 6. Pada tabel 6 ini diperlihatkan hasil yang diperoleh jika dataset tanpa proses SMOTE dan jika pada dataset diterapkan SMOTE.

Tabel 6. Perbandingan AUC

Metode	Tanpa SMOTE (%)	SMOTE (%)	Kenaikan (%)
KNN	0,669	0,720	0,051
Random Forest	0,779	0,832	0,053
SVM	0,786	0,809	0,023
MLP	0,730	0,827	0,097

Jika disajikan dengan menggunakan grafik seperti tersaji pada gambar 4, hasil AUC untuk keempat metode pengklasifikasi yang digunakan adalah sebagai berikut:



Gambar 4. Perbandingan F1-Score

Dari tabel 6 dan gambar 4 dapat dilihat bahwa nilai AUC untuk keempat metode pengklasifikasi yang digunakan meningkat setelah dilakukan proses SMOTE pada dataset. Dengan MLP diperoleh kenaikan AUC yang paling tinggi

(0,097) dibanding tiga metode lainnya. Nilai AUC paling tinggi setelah SMOTE dicapai oleh metode random forest yaitu 0,832.

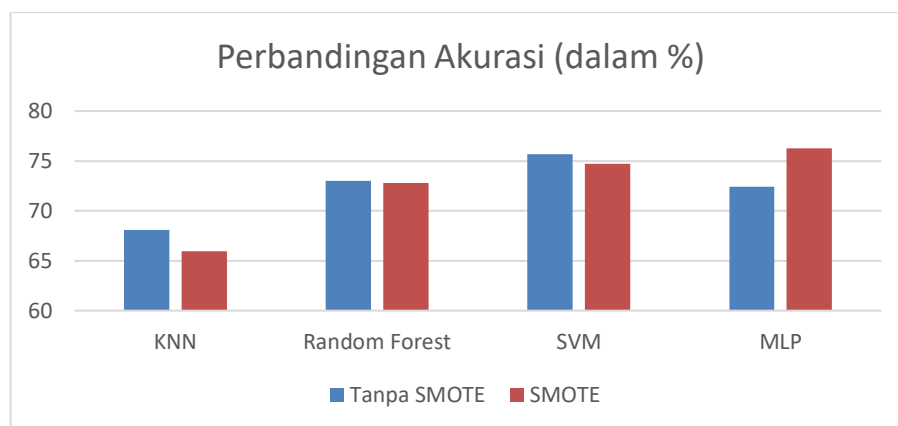
3.5 Akurasi

Perhitungan akurasi menggunakan rumus 4. Perhitungan ini diterapkan pada keempat metode pengklasifikasi yang digunakan dan hasilnya diperoleh seperti pada tabel 7. Pada tabel 7 ini diperlihatkan hasil yang diperoleh jika dataset tanpa proses SMOTE dan jika pada dataset diterapkan SMOTE.

Tabel 7. Perbandingan akurasi

Metode	Tanpa SMOTE (%)	SMOTE (%)	Kenaikan (%)
KNN	68,10	65,93	-2,17
Random Forest	73,00	72,79	-0,21
SVM	75,70	74,71	-0,99
MLP	72,40	76,29	3,89

Jika disajikan dengan menggunakan grafik seperti tersaji pada gambar 5, hasil akurasi untuk keempat metode pengklasifikasi yang digunakan adalah sebagai berikut:



Gambar 5. Perbandingan akurasi

Dari tabel 7 dan gambar 5 dapat dilihat bahwa nilai akurasi untuk keempat metode pengklasifikasi yang digunakan, pada KNN, *random forest*, dan SVM nilai akurasi sedikit menurun, sedangkan dengan MLP sedikit meningkat. Jadi dapat dikatakan bahwa nilai akurasi tidak berubah secara drastis setelah pada dataset diterapkan SMOTE untuk menyeimbangkan data.

4. KESIMPULAN

Tantangan dataset yang tidak seimbang dalam masalah klasifikasi muncul ketika jumlah sampel dalam satu kelas jauh lebih banyak daripada kelas lainnya. Dalam keadaan seperti itu, pengklasifikasi biasanya tergantung kelas mayoritas dalam hal prediksi dan kurang memperhatikan kelas minoritas. Agar supaya metode pengklasifikasi tidak hanya tergantung pada kelas mayoritas, maka dilakukan teknik penyeimbangan data pada dataset yang salah satunya adalah menggunakan teknik *Synthetic Minority Oversampling Technique* (SMOTE). Dengan menggunakan dataset German Credit Data yang bertujuan untuk menentukan pelanggan apakah layak diberikan kredit atau tidak dilakukan proses SMOTE dan kemudian diklasifikasi dengan metode: *Random Forest*, *K-Nearest Neighbor (KNN)*, *Support Vector Machine (SVM)*, dan *Multilayer Perceptron (MLP)* dan selanjutnya diukur kinerjanya. Dengan menggunakan ukuran kinerja *recall*, presisi, *F1-Score*, dan AUC ternyata setelah diterapkan SMOTE pada keempat metode pengklasifikasi, semua ukuran kinerjanya meningkat. Sedangkan untuk ukuran kinerja akurasi dengan menggunakan keempat metode pengklasifikasi, ternyata setelah dilakukan SMOTE kinerja *Random Forest*, KNN, dan SVM sedikit menurun, hanya MLP yang sedikit meningkat. Ini menunjukkan akurasi kurang sesuai untuk mengukur kinerja pada dataset tidak seimbang.

REFERENCES

- [1] E. Sudarmaji and S. Ambarwati, "Pembelajaran Mesin untuk Menilai Kelayakan Kredit Proyek Retrofit : Multinomial Logit," no. September, 2022, doi: 10.24123/jati.v15i2.4912.
- [2] M. Anis and M. Ali, "Investigating the Performance of Smote for Class Imbalanced Learning : A Case Study of Credit Scoring Datasets," *Eur. Sci. J.*, vol. 13, no. 33, pp. 340–353, 2017, doi: 10.19044/esj.2017.v13n33p340.
- [3] C. C. Tusell-rey, O. Camacho-nieto, and Y. Cornelio, "Customized Instance Random Undersampling to Increase Knowledge Management for Multiclass Imbalanced Data Classification," 2022.

- [4] L. Qadrini, Hikmah, and Megasari, “Oversampling , Undersampling , Smote SVM dan Random Forest pada Klasifikasi Penerima Bidikmisi Sejava Timur Tahun 2017,” vol. 3, no. 4, pp. 386–391, 2022, doi: 10.47065/josyc.v3i4.2154.
- [5] W. Chaipanha and P. Kaewwichian, “SMOTE VS . Random Undersampling for Imbalanced Data- Car Ownership Demand Model,” no. March, 2022, doi: 10.26552/com.C.2022.3.D105-D115.
- [6] X. Dastile, T. Celik, H. Vandierendonck, and S. Member, “Model-Agnostic Counterfactual Explanations in Credit Scoring,” *IEEE Access*, vol. 10, no. April, pp. 69543–69554, 2022, doi: 10.1109/ACCESS.2022.3177783.
- [7] X. Sika and R. Amelia, “Klasifikasi Kelayakan Pemberian Kredit Pada Calon Debitur Menggunakan Naïve Bayes,” vol. 9, no. 6, pp. 1833–1839, 2022, doi: 10.30865/jurikom.v9i6.5131.
- [8] C. Wang and Z. Xiao, “applied sciences A Deep Learning Approach for Credit Scoring Using Feature Embedded Transformer,” 2022.
- [9] D. W. Triscowati and L. D. Jayanti, “Penilaian Kredit Pada Data Tak Seimbang Menggunakan Random Forest Credit Scoring In Imbalance Data Using Random Forest,” vol. 1, pp. 25–31, 2021.
- [10] H. B. Nguyen and V. Huynh, “On Sampling Techniques for Corporate Credit Scoring,” *J. Adv. Comput. Intell. Intell. Informatics*, vol. 24, no. 1, 2020.
- [11] S. R. Lenka, S. K. Bisoy, R. Priyadarshini, and M. Sain, “Review Article Empirical Analysis of Ensemble Learning for Imbalanced Credit Scoring Datasets : A Systematic Review,” *Wirel. Commun. Mob. Comput.*, vol. 2022, 2022.
- [12] H. Li *et al.*, “A novel method for credit scoring based on feature transformation and ensemble model,” *PeerJ Comput. Sci.*, pp. 1–18, 2021, doi: 10.7717/peerj-cs.579.
- [13] A. Almustfa, H. Adam, and M. Bee, “Machine Learning Models and Data-Balancing Techniques for Credit Scoring : What Is the Best Combination ?,” *Risk*, vol. 10, no. 169, 2022.
- [14] N. V Chawla, K. W. Bowyer, and L. O. Hall, “SMOTE : Synthetic Minority Over-sampling Technique,” *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.
- [15] A. Fernandez, S. Garcia, F. Herrera, and N. V Chawla, “SMOTE for Learning from Imbalanced Data : Progress and Challenges , Marking the 15-year Anniversary,” *J. Artif. Intell. Res.*, vol. 61, pp. 863–905, 2018.
- [16] M. Deng, Y. G. Id, C. Wang, and F. Wu, “An oversampling method for multi-class imbalanced data based on composite weights,” *PLoS One*, vol. 16, no. 11, pp. 1–15, 2021, doi: 10.1371/journal.pone.0259227.
- [17] D. Dua and C. Graff, “No Title,” *UCI Mach. Learn. Repos.*, 2019.
- [18] N. Darapureddy, N. Karatapu, and T. K. Battula, “Research of Machine Learning Algorithms using K-Fold Cross Validation,” *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6, pp. 215–218, 2019, doi: 10.35940/ijeat.F1043.0886S19.
- [19] I. K. Nti, “Performance of Machine Learning Algorithms with Different K Values in K-fold Cross- Validation,” *J. Inf. Technol. Comput. Sci.*, vol. 6, pp. 61–71, 2021, doi: 10.5815/ijitcs.2021.06.05.
- [20] U. Erdiansyah, A. I. Lubis, and K. Erwansyah, “Komparasi Metode K-Nearest Neighbor dan Random Forest Dalam Prediksi Akurasi Klasifikasi Pengobatan Penyakit Kutil,” *J. MEDIA Inform. BUDIDARMA*, vol. 6, pp. 208–214, 2022, doi: 10.30865/mib.v6i1.3373.
- [21] I. F. Yulianti and P. R. Sihombing, “Penerapan Metode Machine Learning dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah di Indonesia Implementation of Machine Learning Method in Risk Classification on Low Birth weight in Indonesia,” *Matrik J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 20, no. 2, pp. 417–426, 2021, doi: 10.30812/matrik.v20i2.1174.
- [22] M. L. Suliztia and A. Fauzan, “Comparing Naïve Bayes , K-Nearest Neighbor , And Neural Network Classification Methods Of Seat Load Factor In Lombok Outbound Flights,” *J. Mat. Stat. Komputasi*, vol. 16, no. 2, pp. 187–198, 2020, doi: 10.20956/jmsk.v.
- [23] N. H. Ovirianti, M. Zarlis, and H. Mawengkang, “Support Vector Machine Using A Classification Algorithm,” *Sink. J. dan Penelit. Tek. Inform.*, vol. 7, no. 3, pp. 2103–2107, 2022.
- [24] A. A. Almazloun, A.-R. Al-Hinnawi, and R. De Fazio, “Assessment of Multi-Layer Perceptron Neural Network for Pulmonary Function Test ’ s Diagnosis Using ATS and ERS Respiratory Standard Parameters,” *Computers*, vol. 11, 2022.
- [25] R. Indransyah *et al.*, “Klasifikasi Sentimen Pergelaran MotoGP di Indonesia Menggunakan Algoritma Correlated Naïve Bayes Clasifier,” pp. 60–66, 2022.
- [26] A. Nabillah, S. Alam, and M. G. Resmi, “Twitter User Sentiment Analysis Of TIX ID Applications Using Support Vector Machine Algorithm,” vol. 3, no. 1, pp. 14–27, 2022.