

PENERAPAN MWMOTE UNTUK MENGATASI KETIDAKSEIMBANGAN KELAS PADA KLASIFIKASI RISIKO KREDIT

Maria Ulfah¹, Triando Hamonangan Saragih², Dwi Kartini³, Muhammad Itqan Mazdadi⁴, Friska Abadi⁵

^{1,2} Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Jalan Jenderal Ahmad Yani KM 36, Banjarbaru, Kalimantan Selatan 70714

¹mariaulfah574@gmail.com, ²triando.saragih@ulm.ac.id, ³dwikartini@ulm.ac.id, ⁴mazdadi@ulm.ac.id, ⁵friska.abadi@ulm.ac.id

Corresponding author: Triando.saragih@ulm.ac.id

Abstrak

Salah satu bentuk usaha yang dijalankan oleh perbankan adalah pemberian kredit terhadap nasabah. Bank akan selalu berusaha mengoptimalkan penyaluran kredit terhadap nasabah, akan tetapi tidak menutup kemungkinan bahwa kredit yang diberikan tersebut memiliki risiko. Guna menekan dan meminimalisir risiko kredit pihak bank perlu melakukan analisis terhadap data yang dimiliki nasabah agar dapat mengambil keputusan apakah nasabah atau calon debitur layak diberikan pinjaman dalam bentuk kredit. Salah satu cara untuk menyelesaikan masalah analisa risiko kredit adalah dengan melakukan klasifikasi dengan menggunakan *machine learning*. Pada penelitian ini dilakukan klasifikasi dengan menggunakan algoritma *Support Vector Machine* (SVM) serta oversampling data dengan menggunakan MWMOTE dan *Improve* MWMOTE. Data yang digunakan pada penelitian ini adalah data *german credit risk* yang memiliki Kelas *bad credit* yang terdiri atas 300 data dan kelas *good credit* terdiri atas 700 data. Penelitian dilakukan dengan membandingkan klasifikasi SVM dengan dan tanpa oversampling. Hasilnya didapatkan bahwa nilai akurasi dari klasifikasi *Improve* MWMOTE SVM memiliki nilai tertinggi jika dibandingkan dengan SVM MWMOTE, dan SVM yaitu sebesar 77,95%.

Kata kunci : Risiko Kredit, SVM, MWMOTE, *Improve* MWMOTE, Klasifikasi

1. Pendahuluan

Kredit merupakan salah satu bentuk usaha yang dijalankan oleh perbankan. Berdasarkan UU Nomor 10 Tahun 1998 kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan atau kesepakatan pinjam-meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam untuk melunasi utangnya setelah jangka waktu tertentu dengan pemberian bunga. Kredit yang diberikan oleh pihak bank kepada nasabah tidak menutup kemungkinan mengandung risiko.

Risiko yang dihadapi dalam kegiatan penyaluran kredit adalah terjadinya kredit bermasalah. Kredit berkembang menjadi bermasalah dapat disebabkan oleh berbagai hal yang berasal dari nasabah. Kondisi eksternal dan pemberi kredit. Dalam penyaluran kredit belum tentu semua kredit atau pembiayaan yang di berikan selalu lancar dan bebas dari risiko kredit. Oleh karena itu, kredit yang di salurkan oleh suatu bank harus dikelola dengan baik untuk meminimalisir terjadi kerugian (Desda & Yurasti, 2019).

Pada penelitiannya, Lusiyanti (2018) melakukan klasifikasi untuk pengambilan keputusan menerima atau menolak permohonan kredit. Dalam

studi ini, dukungan *Support Vector Machine* (SVM) diimplementasikan untuk memprediksi risiko kredit. Data yang digunakan adalah data yang diperoleh dari salah satu lembaga pembiayaan. Dengan menggunakan fungsi aktivasi yang berbeda, akurasi 80.9524% diperoleh atau ada 51 data yang dapat diprediksi akurat dari data yang ada.

Agar hasil klasifikasi tidak bias, klasifikasi memerlukan data yang seimbang. Sedangkan untuk kasus risiko kredit buruk memiliki jumlah yang bisa dibilang sedikit, hal ini menyebabkan ketidakseimbangan data. Sehingga data tidak seimbang itu sendiri perlu diatasi. Salah satu cara untuk mengatasi ketidakseimbangan data adalah dengan metode *oversampling* atau pendistribusian data agar seimbang dengan menambahkan data sintetik pada kelas minoritas secara acak (Untoro & Buliali, 2018).

Pada penelitian Untoro & Buliali (2018) menggunakan MWMOTE untuk menangani permasalahan ketidakseimbangan data pada laboratorium kesehatan sehingga diperoleh hasil pengklasifikasian penyakit dengan tingkat akurasi lebih tinggi. Hasil pada penelitian ini menunjukkan bahwa MWMOTE dapat meningkatkan akurasi untuk permasalahan ketidakseimbangan data sebesar 3,13%.

Dikarenakan MWMOTE melakukan *generate* data sintetik berdasarkan anggota klaster kelas minoritas yang berdekatan dengan kelas mayoritas, data yang dihasilkan masih terpaku pada area tepi klaster sehingga memungkinkan hasil data sintesis masuk kedalam kelas mayoritas. Pada penelitian Saputra *et al.*, (2021) melakukan improvisasi proses *generate* data sintesis pada metode MWMOTE agar data yang dihasilkan terbesar luas dalam kelas minoritas. hasilnya menunjukkan metode yang diusulkan mampu meningkatkan performa klasifikasi untuk klasifikasi KNN dan C4.5 *Decision Tree* sebesar -.45% dan 0.96% dibandingkan dengan metode MWMOTE.

Berdasarkan uraian diatas, maka penelitian yang diangkat yaitu melakukan perbandingan klasifikasi *Support Vector Machine* (SVM) dengan penerapan MWMOTE serta *Improve* MWMOTE untuk menyeimbangkan kelas pada dataset Risiko Kredit untuk mengetahui berapa tingkat akurasi yang didapatkan dari klasifikasi risiko kredit menggunakan *Support Vector Machine* (SVM).

2. Tinjauan Pustaka

2.1 Risiko Kredit

Bank merupakan jantung perekonomian Nasional, hal ini dikarenakan uang yang mengalir ke bank kemudian diedarkan kembali oleh bank ke sistem perekonomian untuk menjalankan kegiatan perekonomian. Semakin berkembangnya suatu bank akan diiringi oleh tantangan yang harus dihadapi oleh bank sebagai Lembaga keuangan yang berdasarkan kepercayaan (*agent of trust*) (Nuryanto *et al.*, 2020).

Menurut Silalahi *et al.* (2017) dalam kehidupan perekonomian suatu negara, bank memiliki peranan penting dalam perekonomian. Salah satu pelayanan dalam dunia perbankan adalah memberi pinjaman kredit kepada nasabah yang memenuhi syarat perbankan. Kredit merupakan fasilitas keuangan yang memungkinkan seseorang atau badan usaha meminjam uang untuk membeli produk dan membayarnya kembali dalam jangka waktu yang ditentukan.

Risiko kredit didefinisikan sebagai risiko kerugian sehubungan dengan pihak peminjam (*counterparty*) tidak dapat dan atau tidak mau memenuhi kewajiban untuk membayar kembali dana yang di pinjamnya secara penuh pada saat jatuh tempo atau sesudahnya. Risiko Kredit dapat didefinisikan dengan kerugian yang didapatkan karena pihak peminjam tidak mampu atau tidak mau memenuhi atau membayar kewajibannya saat jatuh tempo (Dana *et al.*, 2019).

Kasus kredit macet ini bisa disebabkan oleh beberapa faktor seperti faktor eksternal dan faktor internal dari koperasi itu sendiri. Faktor eksternal yang mengakibatkan terjadinya kredit macet adalah

kondisi ekonomi secara makro baik itu pertumbuhan ekonomi dan tingkat inflasi, itikad tidak baik dari nasabah, dan akibat adanya bencana alam. Sedangkan faktor internal yang mengakibatkan Kasus kredit macet ini bisa terjadinya kredit macet adalah prosedur pemberian kredit yang tidak jelas atau kurang selektif, itikad tidak baik dari petugas koperasi, dan lemahnya sistem pengawasan kredit (Brahmana *et al.*, 2020).

2.2 Data Mining

Data mining adalah suatu pencarian dan analisa pada suatu koleksi data (*database*) sehingga ditemukan suatu pola yang menarik dengan tujuan mengekstrak informasi dan pengetahuan yang akurat dan potensial,serta dapat dipahami dan berguna bagi pengambilan keputusan (Setiyani, 2020).

Di dalam data mining terdapat istilah lain yang mempunyai makna serupa dengan data mining yakni *Knowledge Discovery in Database* (KDD). Data mining serta KDD memiliki tujuan yang sama yakni menggunakan data yang tersedia pada basis data kemudian mengolah data untuk mendapatkan sebuah informasi baru yang bermanfaat. Selain itu, banyak istilah lain yang memiliki makna serupa dengan data mining misalnya, *Knowledge Discovery in Database*, ekstraksi pengetahuan, analisis pola / data, arkeologi data, dan pengerukan data. Banyak orang memperlakukan data mining sebagai sinonim untuk istilah lain yang populer digunakan, penemuan pengetahuan dari data, atau KDD, sementara yang lain melihat penambangan data hanya sebagai langkah penting dalam proses penemuan pengetahuan (Utomo, 2020).

2.3 SVM

SVM merupakan teknik data mining yang relatif baru untuk melakukan prediksi, baik untuk kasus klasifikasi maupun regresi. SVM pertama kali diperkenalkan oleh Vapnik, Boser, dan Guyon (1992). Prinsip dasar SVM adalah linier classifier, yaitu kasus klasifikasi dimana data dapat dipisahkan secara linier, kemudian dikembangkan sehingga dapat bekerja pada data yang tidak dapat dipisahkan secara linier dengan memasukan konsep trik kernel (Lusiyanti, 2018).

Beberapa jenis kernel yang umumnya digunakan dalam SVM adalah sebagai berikut :

a. Polynomial

$$k(X, Y) = (X, Y + c)^d \quad (1)$$

X, Y : merupakan vektor di ruang input

c : merupakan bias parameter pada fungsi kernel

d : merupakan polynomial *degree*

b. RBF (*Gaussian Radial Basic Function*)

$$k(X, Y) = \text{Exp}^{-\gamma |X, Y|^2}, \gamma \quad (2)$$

X, Y : merupakan vektor di ruang input

γ : merupakan representasi parameter gamma pada fungsi kernel

c. Sigmoid

$$k(X, Y) = \tanh(\gamma X, Y + c) \tag{3}$$

X, Y : merupakan vektor di ruang input
 γ : merupakan representasi parameter gamma pada fungsi kernel
 c : merupakan bias parameter pada fungsi kernel

(Al-Mejibli *et al.*, 2020)

2.4 MWMOTE

MWMOTE melakukan pembuatan data sintetik dengan pembobotan dan klasterisasi data minoritas. Pembobotan dan pengklasteran data minoritas ini bertujuan untuk dan klasterisasi data minoritas. Pembobotan dan pengklasteran data minoritas ini bertujuan untuk mendapatkan data sintetik yang representatif. Hasil pengklasteran data minoritas ini bertujuan untuk mendapatkan data sintetik yang representatif. (Untoro, 2021).

MWMOTE mengidentifikasi data minoritas yang sulit dipelajari. Data minoritas yaitu pada data mayoritas, data minoritas berdekatan (*borderline*) dengan data mayoritas dan data minoritas informasi tersebut berada pada *borderline*. Kedua, setiap anggota sampel minoritas yang informatif diberi bobot pilihan sampel yang tertimbang (Untoro *et al.*, 2020).

Metode MWMOTE memiliki 3 tahapan penting dalam memperbaiki permasalahan *imbalance class*. Tahap pertama MWMOTE mengidentifikasi sampel kelas minoritas yang paling penting dari set minoritas yang asli, S_{min} dibuat kedalam set oleh sampel yang teridentifikasi, tahap kedua setiap anggota S_{min} diberikan bobot sesuai data, pada tahap ketiga MWMOTE menghasilkan sampel sintesis dari S_{min} menggunakan S_{ws} dan menghasilkan output S_{omin} dengan menambahkan sampel sintesis ke dalam S_{min} (Syarif & Nugraha, 2023).

Terdapat enam data parameter yang ada pada MWMOTE yang digunakan untuk melakukan generate data sintesis, yaitu:

- a. S_{max} : kumpulan data mayoritas
- b. S_{min} : kumpulan data minoritas
- c. N : jumlah data sistetis yang akan di-generate
- d. K : jumlah node tetangga yang digunakan untuk memprediksi *noise* pada kelas minoritas
- e. K_2 : jumlah node tetangga dari kelas mayoritas yang digunakan untuk membangun informasi mengenai kelas minoritas
- f. K_3 : jumlah node tetangga dari kelas minoritas yang digunakan untuk membangun informasi mengenai kelas minoritas.

(Saputra *et al.*, 2021)

2.5 Improve MWMOTE

Pada penelitiannya Saputra *et al.*, (2021) mengajukan improvisasi MWMOTE agar menghasilkan data sintesis pada area S_{imin} yang berada di *borderline minorities*. Proses ini merupakan

improvisasi metode MWMOTE dengan menambahkan informasi data S_{bmin} dengan penambahan parameter k_4 . S_{bmin} merupakan *mini-cluster* dari *cluster* kelas minoritas, yaitu area antara *centroid cluster* minoritas dengan *borderline minorities* yang menjadi area *generate* data sintesis baru dan k_4 merupakan jumlah note tetangga dari *centroid cluster* minoritas. Data S_{bmin} menjadi kunci pada metode ini agar data sintesis yang dihasilkan tidak terlalu rapat ke tepi (*borderline minorities*) akan tetapi menyebar dalam area *cluster minoritas*.

Generate data sintesis dengan menambahkan parameter k_4 pada MWMOTE untuk menentukan arena *generate* data sintesis (S_{bmin}). Tahapan *generate* dilakukan dengan cara :

1. Menentukan *cluster - cluster* minoritas (L_M) dari kelas minoritas S_{min}
2. Inialisasi *oversample* data $S_{omin} = S_{min}$
3. lakukan perhitungan sebanyak N data sinetis sampling, dengan mengacu pada persamaan.

$$S = x + \alpha \times (y - x) \tag{4}$$

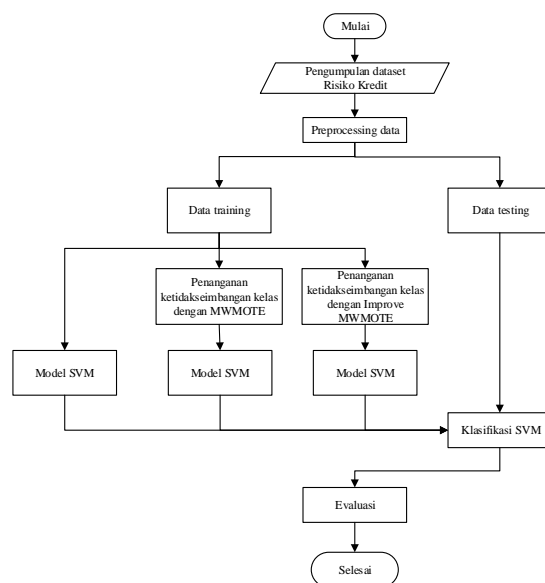
Dimana : x adalah data sampel random dari S_{imin} dalam cluster $L_k, 1 \leq k \leq M$

y : adalah data sampel random dari S_{bmin} pada *cluster* yang sama dengan x

α : adalah nilai random [0,1]

3. Metode Penelitian

Perancangan alur penelitian dilukan agar penelitian dapat dilakukan dengan baik. alur penelitian dapat dilihat pada gambar 1.



Gambar 1. Alur penelitian

3.1 Pengumpulan Data

Data yang digunakan pada penelitian ini adalah dataset *German Credit Risk Dataset* dari situs *UCI*. Dataset memiliki 1000 record terdiri atas 20 fitur

dengan dua kelas kelas yaitu kredit buruk atau kredit baik.

3.2 Preprocessing Data

Sebelum data dapat digunakan, perlu dilakukan penyesuaian data agar tidak memberikan dampak buruk pada kinerja klasifikasi. *Preprocessing* yang dilakukan diantaranya adalah merubah data kategorik yang tidak dapat diproses oleh klasifikasi menjadi data numerik menggunakan *label encoding*.

Label encoding merepresentasikan setiap kategori pada data kedalam suatu integer sehingga dapat dipahami oleh model. Namun, karena output yang dihasilkan metode ini berupa integer, maka *output* yang dihasilkan akan memiliki natural *ordered relationship*, yang berarti komputer secara otomatis akan memberikan bobot yang lebih besar kepada kategori yang memiliki nilai *categorical* yang lebih besar (Winata *et al.*, 2020).

Kemudian untuk data numerik yang memiliki skala nilai yang berbeda – beda dinormalisasi menggunakan normalisasi penskalaan *Min-Max* yang menskalakan semua nilai numerik dari fitur numerik kekisaran tertentu melalui perhitungan *Min-Max* sebagai berikut:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (5)$$

X : merupakan nilai dari fitur yang akan dinormalisasi

X_{min} : merupakan nilai minimum dari fitur yang akan dinormalisasi

X_{max} : merupakan nilai maksimum dari fitur yang akan dinormalisasi

(Alam *et al.*, 2020).

3.3 Pembagian Data

Data akan dibagi kedalam dua bagian yaitu training dan testing, 80% dari keseluruhan dataset menjadi data training dan 20% data dari keseluruhan dataset menjadi data testing.

Pada percobaan pertama data training yang merupakan 80% bagian dari data asli diambil dengan mempertahankan proporsi data asli, kemudian dilakukan evaluasi klasifikasi menggunakan metode SVM. Sedangkan untuk percobaan kedua, dilakukan *oversampling* menggunakan metode MWMOTE pada data training dengan jumlah yang sama yaitu 80% dan dengan mempertahankan proporsi data asli, setelahnya dilakukan evaluasi untuk melihat kinerja dari metode *oversampling* menggunakan 20% data testing menggunakan klasifikasi SVM. Untuk percobaan ketiga data *oversampling* yang digunakan pada data training adalah *oversampling* menggunakan metode *Improve MWMOTE* dan kemudian di evaluasi menggunakan klasifikasi SVM. Hal ini dilakukan untuk membandingkan hasil dari klasifikasi data training tanpa *oversampling* dengan

data training yang menggunakan *oversampling MWMOTE* serta data training yang menggunakan *oversampling Improve MWMOTE*.

3.4 Oversampling

Data sintetis dibentuk melalui proses *oversampling*, yaitu melakukan *oversampling* pada data minoritas untuk mendapatkan data sintetis. Data sintetis sendiri berasal dari data minoritas pada data latih, sedangkan data testing tanpa melakukan penanganan imbalance bertujuan untuk mengevaluasi performa dari MWMOTE dan *Improve MWMOTE* dalam melakukan *oversampling*.

3.5 Klasifikasi SVM

Data yang telah ditraining dan ditesting akan diklasifikasikan menggunakan metode *Support Vector Machine (SVM)* kernel RBF dan klasifikasi *Support Vector Machine (SVM)* dengan penanganan ketidakseimbangan kelas menggunakan MWMOTE dan *Improve MWMOTE*.

3.6 Evaluasi

Untuk evaluasi Model yaitu untuk mengetahui performa klasifikasi SVM dengan membandingkan hasil kerja. Evaluasi dilakukan dengan menghitung akurasi, *recall*, presisi dan *F-Measure (F1-score)* menggunakan confusion matrix. Pada pengukuran kinerja *confusion matrix*, ada 4 (empat) istilah sebagai representatif dari hasil proses klasifikasi, yaitu *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)* (Hadianto *et al.*, 2019).

Tabel 1. *Confusion Matrix*

kelas	Terklasifikasi Positif	Terklasifikasi Negatif
Positif	TP	FN
Negatif	FP	TN

Berdasarkan nilai *confusion matrix* yang didapat dapat diperoleh nilai akurasi, presisi dan recall.

a. Akurasi adalah seberapa akurat sistem dapat mengklasifikasikan data secara benar.

$$akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (5)$$

b. Presisi adalah jumlah data positif yang diklasifikasikan secara benar dibagi dengan total data yang diklasifikasi positif.

$$precision = \frac{TP}{FP + TP} \times 100\% \quad (6)$$

c. *Recall* menunjukkan seberapa besar data positif yang terklasifikasi dengan benar oleh sistem.

$$recall = \frac{TP}{FN + TP} \times 100\% \quad (7)$$

d. *F1 score* merupakan rata – rata harmonik antara presisi dan recall.

$$F1 - score = 2 \times \frac{\text{presisi} \times \text{recall}}{\text{presisi} + \text{recall}} \quad (8)$$

(Satria et al., 2020).

4. Hasil dan Pembahasan

4.1 Pengumpulan Data

Data yang digunakan adalah German Credit Risk yang diambil dari situs UCI machine learning repository. Dataset yang digunakan merupakan dataset lengkap, dataset ini memiliki total 1000 data dan terdapat 21 atribut atau fitur didalamnya, terdiri atas dua kelas yaitu kelas good dan bad credit. Kelas bad credit atau kredit dengan risiko tinggi yang terdiri atas 300 data dan kelas good credit atau kredit dengan risiko rendah terdiri atas 700 data. Adapun deskripsi dari dataset dapat dilihat pada tabel 2.

Tabel 2. Deskripsi dataset

No	Nama Variabel	Deskripsi
1	Status (kategori)	Status akun debitur pada bank untuk pengeluaran sehari hari
2	Duration (kuantitatif)	Durasi kredit dalam hitungan bulan
3	Credit history (kategori)	Riwayat kontrak kredit sebelumnya atau saat ini
4	Purpose (kategori)	Alasan mengambil kredit
5	Amount (kuantitatif)	Jumlah total kredit
6	Savings (kategori)	Jumlah tabungan debitur
7	Employment duration (kategori)	Durasi kerja debitur dengan perusahaan saat ini
8	Installment rate (kuantitatif)	Angsuran kredit sekali pakai debitur
9	Personal status sex (kategori)	Informasi mengenai jenis kelamin dan status perkawinan
10	Other debtors (kategori)	Debitur lain untuk kredit
11	Present residence (kuantitatif)	Durasi tinggal ditempat tinggal saat ini
12	Property (kategori)	Peringkat properti debitur apabila diurutkan
13	Age (kuantitatif)	Umur dari debitur
14	Other installment plans (kategori)	Beban kredit atau angsuran selain yang diberikan bank
15	Housing (kategori)	Status tempat tinggal saat ini
16	Number credits (kuantitatif)	Riwayat lengkap mengenai kredit yang diambil
17	Job (kategori)	Tingkat pekerjaan debitur
18	People liable (kuantitatif)	Jumlah total orang yang bergantung pada debitur secara finansial
19	Telephone (kategori)	Status telepon rumah yang terdaftar anas nama debitur
20	Foreign worker (kategori)	Apakah debitur pekerja asing

No	Nama Variabel	Deskripsi
21	Credit risk (binary)	Baik atau buruk

Dari tabel 2 dapat dilihat bahwa 13 dari 21 fitur merupakan fitur dengan nilai kategorik, kemudian ada 7 fitur dengan nilai kuantitatif dan fitur kelas dalam bentuk binary.

4.2 Preprocessing Data

Pada penelitiannya Valentim et al (2019) menurunkan atribut sex dari atribut personal status sex pada dataset asli, hal ini dilakukan dengan menghapus atribut personal status sex yang berisikan beberapa kategori status pernikahan dan gender kemudian digantikan atribut sex yang sudah diturunkan dari atribut sebelumnya menjadi hanya pria dan wanita.

Untuk Label Encoding dilakukan penambahan bilangan bulat pada setiap kategori serta tidak menambahkan kolom baru kedalam data. Didalam dataset german credit risk, terdapat beberapa fitur dengan nilai kategorik yaitu status, credit history, purpose, savings, employment, sex, property, other installment plan, housing, job, telephone, foreign worker. Proses preprocessing ini dilakukan agar data dapat digunakan machine learning.

Pada perhitungan normalisasi min-max perlu diketahui nilai maksimum dan minimum dari fitur yang dinormalisasi. Misalnya pada fitur duration nilai maksimumnya adalah 72 sedangkan nilai minimumnya adalah empat, pada fitur amount nilai maksimumnya adalah 18429 dan nilai minimumnya 230, sedangkan pada fitur age nilai maksimalnya adalah 75 dan nilai minimumnya adalah 19. Contoh perhitungan fitur duration:

$$X_{norm} = \frac{6 - 4}{72 - 4} = 0,02941$$

4.3 Pembagian Data

Sebelum diklasifikasin, data akan dibagi terlebih dahulu menjadi data training dan juga data testing menggunakan split validation. Pembagian dataset dilakukan dengan mempertahankan proporsi yang sama dari training setiap kelas sama seperti dataset awal, pembagian data dibagi dengan besaran 80% untuk data training dan 20% untuk data testing.

4.4 Oversampling MWMOTE

Pada data training data sintetik dibentuk melalui proses oversampling, yaitu melakukan oversampling pada data minoritas dari data training untuk mendapatkan data sintetik. Data sintetik sendiri berasal dari data minoritas pada data latih, sedangkan data testing tanpa melakukan penanganan imbalance bertujuan untuk mengevaluasi performa dari MWMOTE dalam melakukan oversampling.

Oversampling dilakukan dua jenis yaitu Oversampling menggunakan MWMOTE dan Oversampling menggunakan Improve MWMOTE.

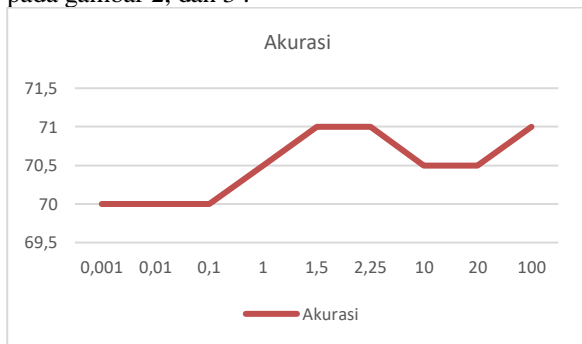
4.5 Klasifikasi SVM

Klasifikasi yang dilakukan pada penelitian ini menggunakan metode klasifikasi SVM dengan trik kernel RBF untuk mengklasifikasi dataset german credit risk untuk mengklasifikasikan good dan bad credit dari data yang ada. Untuk mendapatkan hasil terbaik dari klasifikasi SVM maka observasi dilakukan dengan mencoba parameter dengan rentang nilai parameter uji yang dapat dilihat pada tabel :

Tabel 3. Nilai Parameter Uji

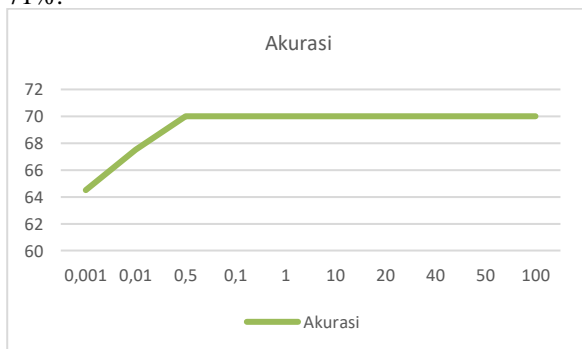
Parameter	Nilai
C	0,001, 0,01, 0,1, 1, 1,5, 2,25, 10, 20, 100
Gamma	0,001, 0,01, 0,1, 0,5, 1, 10, 20, 40, 50, 100

Dari tabel 3 dapat dilihat bahwa pengujian dilakukan dengan menguji nilai parameter C dan gamma, hal ini dilakukan dengan mengubah parameter c pada nilai yang sudah ditentukan dengan parameter gamma pada nilai default, begitu pula sebaliknya melakukan percobaan parameter gamma dengan nilai parameter yang sudah ditentukan dengan nilai parameter c pada nilai default. Hasilnya didapatkan grafik performa SVM yang dapat dilihat pada gambar 2, dan 3 :



Gambar 2. Grafik Performa Parameter C

Pada gambar 2 dapat dilihat bahwa dari pengujian 9 parameter, nilai tertinggi dari akurasi klasifikasi SVM dengan parameter C adalah sebesar 71%.

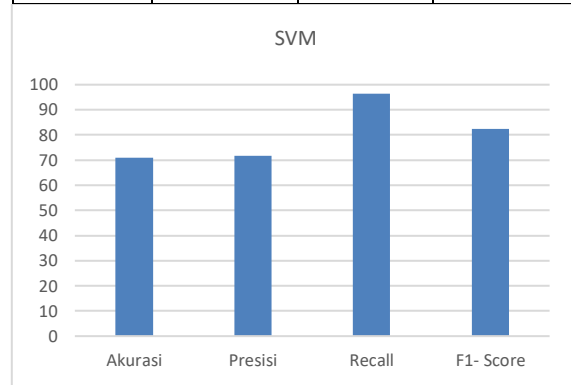


Gambar 3. Hasil uji parameter Gamma

Pada gambar 3 bisa dilihat bahwa dari 11 percobaan parameter nilai akurasi mencapai angka 70% sebagai nilai tertinggi dan setelahnya tidak mengalami kenaikan nilai akurasi sama sekali.

Tabel 4. Performa Model Terbaik SVM

Akurasi	Presisi	Recall	F1-Score
71	71,81	96,43	82,32



Gambar 4. grafik performa model SVM terbaik

Tabel 4 dan gambar 4 memperlihatkan performa terbaik yang didapatkan dari percobaan parameter C dan Gamma pada klasifikasi SVM menggunakan kernel RBF, hasilnya dapat dilihat bahwa nilai akurasi tertinggi pada SVM RBF dengan nilai sebesar 71% dengan nilai presisi 71,81%, recall 96,43%, dan f1 score 82,32%.

Klasifikasi kedua yang dilakukan pada penelitian ini adalah klasifikasi menggunakan SVM dengan oversampling menggunakan MWMOTE untuk menangani ketidakseimbangan kelas pada data training dengan mengobservasi performa dari SVM dengan kernel RBF. Observasi dilakukan dengan mengambil nilai rata – rata akurasi tertinggi dari klasifikasi SVM sebanyak sepuluh kali dengan pengujian nilai parameter berdasarkan tabel 5.

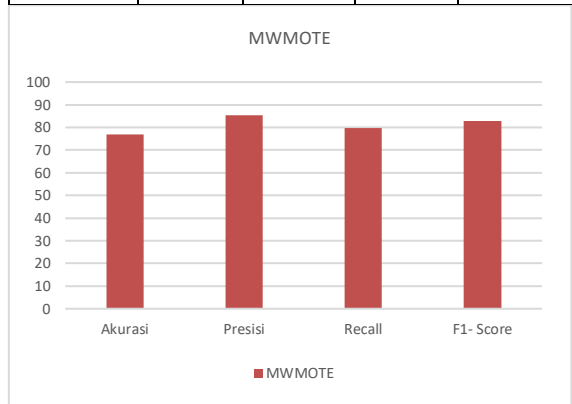
Tabel 5. Performa 10 Akurasi Tertinggi Model MWMOTE-SVM

No	Akurasi	Presisi	Recall	F1-Score
1	77	85,6	80,7	83,1
2	76	84,8	80	82,4
3	77	84,1	72,9	83,5
4	77	85,6	80,7	83,1
5	77,5	86,3	80,7	83,4
6	75,5	85,8	77,9	81,6
7	76,5	85,5	80	82,7
8	77	85,6	80,7	83,1
9	76,5	83,9	82,1	83
10	77,5	86,8	80	83,3

Pada tabel 5 terlihat bahwa performa model untuk mengklasifikasikan kelas memiliki nilai akurasi tertinggi yang berbeda – beda tiap kali percobaan. Nilai akurasi tertinggi pada percobaan ini adalah 77,5% dan nilai akurasi tertinggi dengan nilai performa terendah terendah sebesar 75,5%.

Tabel 6. Rata – Rata Akurasi Tertinggi

No	Akurasi	Presisi	Recall	F1-Score
Rata - rata	76,75	85,4	79,57	82,92



Gambar 5. grafik nilai rata – rata nilai akurasi

Pada tabel 6 dan gambar 5 dapat dilihat bahwa rata – rata nilai akurasi tertinggi dari model klasifikasi SVM dengan *oversampling* MWMOTE adalah 76,75% kemudian rata rata presisi sebesar 85,4%, rata – rata *recall* 79,57%, dan rata – rata *F1-score* sebesar 82,92%.

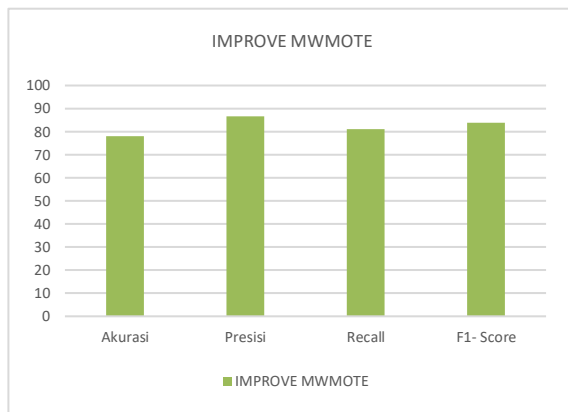
Klasifikasi ketiga yang dilakukan pada penelitian ini adalah klasifikasi menggunakan SVM dengan *oversampling* menggunakan *Improve* MWMOTE untuk menangani ketidakseimbangan kelas pada data training dengan melihat performa dengan menggunakan klasifikasi SVM dengan kernel RBF. Observasi dilakukan dengan mengambil nilai rata – rata akurasi tertinggi dari klasifikasi SVM dengan pengujian nilai parameter berdasarkan tabel 7. Tabel 7. Performa Akurasi Terbaik Model *Improve* MWMOTE-SVM

No	Akurasi	Presisi	Recall	F1-Score
1	78,5	85,9	82,9	84,4
2	78	88,1	79,3	83,5
3	78,5	87,6	80,7	84
4	77,5	86,3	80,7	83,4
5	79	87,7	81,4	84,4
6	77	83,6	83,6	83,6
7	77	86,7	79,3	82,8
8	78,5	87	81,4	84,1
9	77,5	86,3	80,7	83,4
10	78	86,9	80,7	83,7

Pada tabel 7, terlihat bahwa performa model untuk mengklasifikasikan kelas memiliki nilai akurasi tertinggi yang berbeda – beda tiap kali percobaan. Nilai akurasi tertinggi pada percobaan adalah 79% dan nilai akurasi terendah terendah sebesar 77%.

Tabel 8. Rata – Rata Akurasi Terbaik Model *Improve* MWMOTE-SVM

No	Akurasi	Presisi	Recall	F1-Score
Rata - rata	77,95	86,61	81,07	83,73



Gambar 6. grafik rata – rata akurasi

Pada tabel 8 dan gambar 6 dapat dilihat bahwa rata – rata nilai akurasi tertinggi dari model klasifikasi SVM dengan *oversampling* *Improve* MWMOTE adalah 77,95% kemudian rata rata presisi sebesar 86,61%, rata – rata *recall* 81,07%, dan rata – rata *F1-score* sebesar 83,73%.

4.6 Evaluasi

Evaluasi dilakukan dengan menghitung akurasi, presisi, *recall*, dan *F-Measure (F1-score)* untuk mendapatkan hasil evaluasi dari model terbaik klasifikasi SVM.

Tabel 9. *Confusion Matrix* Nilai Model SVM

<i>Confusion matrix</i>		Prediksi	
		<i>Good credit</i>	<i>Bad credit</i>
Aktual	<i>Good credit</i>	135	5
	<i>Bad credit</i>	53	7

Evaluasi akurasi SVM dilakkan dengan menghitung nilai didapatkan dari jumlah prediksi kelas yang benar dari kelas positif dan negatif kemudian dibagi dengan keseluruhan data perhitungan dilakukan dengan menggunakan persamaan 5.

$$akurasi = \frac{135+7}{127+7+53+5} = \frac{162}{200} = 71\%$$

Evaluasi presisi didapatkan dengan membagi jumlah benar kelas positif (*true positive*) dibagi dengan jumlah benar kelas positif (*true positive*) ditambah dengan jumlah prediksi kelas positif yang salah (*false positive*). perhitungan dilakukan dengan menggunakan persamaan 6.

$$presisi = \frac{135}{135+53} = \frac{135}{188} = 71,81\%$$

Hasil evaluasi *recall* didapatkan dari jumlah prediksi benar kelas positif (*true positive*) dibagi dengan jumlah prediksi benar kelas positif (*true positive*) ditambahkan dengan prediksi kelas negatif

yang salah (*false negative*) perhitungan dilakukan dengan menggunakan persamaan 7.

$$recall = \frac{135}{5+135} = \frac{135}{140} = 96,43\%$$

Evaluasi *f1 score* pada model algoritma SVM didapatkan dari nilai presisi yang dikalikan dengan nilai *recall* kemudian dibagi dengan nilai presisi yang ditambah dengan nilai *recall* kemudian hasil pembagian tersebut dikalikan dua.

$$f1 - score = 2 \times \frac{71,81 \times 96,43}{71,81 + 96,43} = 82,32\%$$

Pada model SVM didapatkan performa terbaik berupa akurasi sebesar 71%, presisi sebesar 96,43%, *recall* sebesar 71,81 %, dan *F1-score* sebesar 82,32%.

Evaluasi kedua dilakukan pada model SVM dengan *oversampling* MWMOTE yang digunakan untuk menangani ketidakseimbangan kelas. Evaluasi dilakukan dengan menghitung nilai akurasi, *recall*, presisi, dan *F1-score* pada model berdasarkan tabel *confusion matrix*.

Tabel 10. *Confusion Matrix* Model MWMOTE-SVM

<i>Confusion matrix</i>		Prediksi	
		<i>Good credit</i>	<i>Bad credit</i>
Aktual	<i>Good credit</i>	112	28
	<i>Bad credit</i>	17	43

Evaluasi akurasi klasifikasi SVM dengan perbaikan data menggunakan *oversampling* MWMOTE dengan menghitung nilai akurasi yang didapatkan dari jumlah prediksi kelas yang benar dari kelas positif dan negatif kemudian dibagi dengan keseluruhan data prediksi. Perhitungan dilakukan dengan menggunakan persamaan 5.

$$akurasi = \frac{112+43}{112+43+28+17} = \frac{155}{200} = 77,5\%$$

Evaluasi presisi pada model terbaik algoritma klasifikasi SVM dengan perbaikan data menggunakan *oversampling* MWMOTE dengan hasil yang didapatkan dengan membagi jumlah benar kelas positif (*true positive*) dibagi dengan jumlah benar kelas positif (*true positive*) ditambah dengan jumlah prediksi kelas positif yang salah (*false positive*). Perhitungan dilakukan dengan menggunakan persamaan 6.

$$presisi = \frac{112}{112+17} = \frac{112}{129} = 86,8\%$$

Evaluasi *recall* pada klasifikasi SVM dengan perbaikan data menggunakan *oversampling* MWMOTE didapatkan dari jumlah prediksi benar kelas positif (*true positive*) dibagi dengan jumlah prediksi benar kelas positif (*true positive*) ditambahkan dengan prediksi kelas negatif yang salah (*false negative*). Perhitungan dilakukan dengan menggunakan persamaan 7.

$$recall = \frac{112}{112+28} = \frac{112}{140} = 80\%$$

Evaluasi *f1-score* pada model terbaik algoritma klasifikasi SVM dengan perbaikan data menggunakan *oversampling* MWMOTE dengan hasil yang didapatkan dari nilai presisi yang dikalikan dengan nilai *recall* kemudian dibagi dengan nilai presisi yang ditambah dengan nilai *recall* kemudian

hasil pembagian tersebut dikalikan dua. Perhitungan dilakukan dengan menggunakan persamaan 8.

$$f1 - score = 2 \times \frac{86,8 \times 80}{86,8 + 80} = 83,3\%$$

Jadi pada model SVM didapatkan performa terbaik berupa akurasi sebesar 77,5%, presisi sebesar 86,8%, *recall* sebesar 80%, dan *F1-score* sebesar 83,3%.

Evaluasi ketiga dilakukan pada model SVM dengan *oversampling Improve* MWMOTE yang digunakan untuk menangani ketidakseimbangan kelas. Evaluasi dilakukan dengan menghitung nilai akurasi, *recall*, presisi, dan *F1-score* pada model berdasarkan tabel *confusion matrix*.

Tabel 11. *Confusion Matrix Improve* MWMOTE-SVM

<i>Confusion matrix</i>		Prediksi	
		<i>Good credit</i>	<i>Bad credit</i>
Aktual	<i>Good credit</i>	114	36
	<i>Bad credit</i>	16	44

Evaluasi klasifikasi SVM dengan perbaikan data menggunakan *oversampling Improve* MWMOTE dengan menghitung nilai akurasi yang didapatkan dari jumlah prediksi kelas yang benar dari kelas positif dan negatif kemudian dibagi dengan keseluruhan data prediksi. Perhitungan dilakukan dengan menggunakan persamaan 5.

$$akurasi = \frac{114+44}{114+44+16+26} = \frac{158}{200} = 79\%$$

Evaluasi presisi pada model terbaik algoritma klasifikasi SVM dengan perbaikan data menggunakan *oversampling Improve* MWMOTE hasil yang didapatkan dengan membagi jumlah benar kelas positif (*true positive*) dibagi dengan jumlah benar kelas positif (*true positive*) ditambah dengan jumlah prediksi kelas positif yang salah (*false positive*). Perhitungan dilakukan dengan menggunakan persamaan 6.

$$presisi = \frac{114}{16+114} = \frac{114}{130} = 87,7\%$$

Evaluasi *recall* pada klasifikasi SVM dengan perbaikan data menggunakan *oversampling Improve* MWMOTE didapatkan dari jumlah prediksi benar kelas positif (*true positive*) dibagi dengan jumlah prediksi benar kelas positif (*true positive*) ditambahkan dengan prediksi kelas negatif yang salah (*false negative*). Perhitungan dilakukan dengan menggunakan persamaan 7.

$$recall = \frac{114}{26+114} = \frac{114}{140} = 81,42\%$$

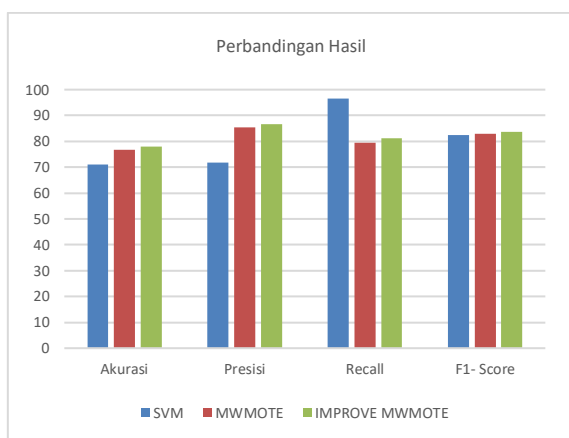
Evaluasi *F1-score* pada model terbaik algoritma klasifikasi SVM dengan perbaikan data menggunakan *oversampling Improve* MWMOTE yang didapatkan dari nilai presisi yang dikalikan dengan nilai *recall* kemudian dibagi dengan nilai presisi yang ditambah dengan nilai *recall* kemudian hasil pembagian tersebut dikalikan dua. Perhitungan dilakukan dengan menggunakan persamaan 8.

$$f1 - score = 2 \times \frac{87,7 \times 81,42}{87,7 + 81,42} = 83,6\%$$

Jadi pada model SVM didapatkan performa terbaik berupa akurasi sebesar 79%, presisi sebesar 87,7%, *recall* sebesar 81,42,86 %, dan *F1-score* sebesar 83,6%.

Tabel 12. Perbandingan Hasil Klasifikasi

	Akurasi	Presisi	Recall	F1-Score
SVM	71	71,81	96,43	82,32
MWMOTE-SVM	76,75	85,4	79,57	82,92
IMPROVE MWMOTE-SVM	77,95	86,61	81,07	83,73



Gambar 7. Perbandingan hasil klasifikasi

Berdasarkan perbandingan hasil klasifikasi yang ada pada tabel 12 dan gambar 7. Nilai rata - rata akurasi dari SVM dengan MWMOTE serta rata - rata akurasi SVM dengan *Improve* MWMOTE memiliki nilai yang lebih besar jika dibandingkan dengan hasil akurasi dengan menggunakan klasifikasi SVM. Begitu pula dengan nilai presisi dan *F1-score*, meskipun begitu nilai presisi dari klasifikasi SVM dengan MWMOTE serta klasifikasi SVM dengan *Improve* MWMOTE memiliki nilai yang lebih rendah jika dibandingkan dengan nilai *recall* dari klasifikasi SVM.

5. Kesimpulan dan Saran

Hasil dari penelitian ini adalah nilai akurasi, presisi, *recall*, dan *f1-score* dari klasifikasi SVM pada dataset *german credit* adalah sebesar 71%, 71,81%, 96,43%, 82,32%. Apabila dibandingkan dengan hasil klasifikasi SVM, hasil klasifikasi MWMOTE-SVM meningkatkan hasil akurasi, presisi, dan *F1-score* pada dataset *german credit* menjadi 76,75%, 85,4%, 82,92% dan penurunan pada nilai *recall* menjadi 79,57 % Dibandingkan dengan hasil klasifikasi SVM dan MWMOTE SVM, hasil klasifikasi *Improve* MWMOTE-SVM meningkatkan hasil akurasi, presisi, *recall* dan *F1-score* pada dataset *german credit* menjadi 77,95%, 86,61%, 81,07%, dan 83,73%

Adapun saran – saran yang dapat diberikan dari penelitian ini adalah. Untuk menerapkan

oversampling MWMOTE dan klasifikasi SVM pada studi kasus berbeda atau pada dataset dengan kelas lebih dari dua.

Daftar Pustaka:

Al-Mejibli, I. S., Alwan, J. K., & Abd, D. H. (2020). The Effect Of Gamma Value On Support Vector Machine Performance With Different Kernels. *International Journal Of Electrical And Computer Engineering*, 10(5), 5497–5506. <https://doi.org/10.11591/IJECE.V10I5.PP5497-5506>

Alam, T. M., Shaukat, K., Hameed, I. A., Luo, S., Sarwar, M. U., Shabbir, S., Li, J., & Khushi, M. (2020). An Investigation Of Credit Card Default Prediction In The Imbalanced Datasets. *IEEE Access*, 8, 201173–201198. <https://doi.org/10.1109/ACCESS.2020.3033784>

Brahamana, N. S., Miftahuddin, D. H., & Se, A. P. (2020). Analisis Profitabilitas Dalam Pemberian Kredit Pada Koperasi Kredit Unam. *Jurnal Mahasiswa UMA*.

Dana, P., Ketiga, P., Kredit, R., Pasar, R., Ekonomi, F., Manajemen, J., & Ratulangi, U. S. (2019). *Pengaruh Dana Pihak Ketiga, Risiko Kredit, Risiko Pasar Dan Risiko Operasional Terhadap Profitabilitas Pada Bank Kategori Buku 2 Periode 2014-2017*. 7(3), 2751–2760.

Desda, M. M., & Yurasti, Y. (2019). Analisis Penerapan Manajemen Risiko Kredit Dalam Meminimalisir Kredit Bermasalah Pada PT. BPR Swadaya Anak Nagari Bandarejo Simpang Empat Periode 2013-2018. *Mbia*, 18(1), 94–106. <https://doi.org/10.33557/Mbia.V18i1.351>

Hadianto, N., Novitasari, H. B., Rahmawati, A., & Komputer, I. (2019). *Klasifikasi Peminjaman Nasabah Bank Menggunakan Metode Neural Network*. 15(2), 163–170. <https://doi.org/10.33480/Pilar.V15i2.658>

Lusiyanti, D. (2018). *Sistem Sederhana Untuk Memprediksi Risiko Pemberian Kredit*. 15.

Nuryanto, U. W., Salam, A. F., Sari, R. P., & Suleman, D. (2020). *Pengaruh Rasio Kecukupan Modal , Likuiditas , Risiko Kredit Dan Efisiensi Biaya Terhadap Profitabilitas Pada Bank Go Public*. 7(1), 1–9.

Saputra, P. Y., Abdullah, M. Z., & Kirana, A. P. (2021). *Improvisasi Teknik Oversampling MWMOTE Untuk Penanganan Data Tidak Seimbang*. 5(April), 398–405. <https://doi.org/10.30865/Mib.V5i2.2811>

Satria, F., Zamhariri, Z., & Syaripudin, M. A. (2020). Prediksi Ketepatan Waktu Lulus Mahasiswa Menggunakan Algoritma C4.5 Pada Fakultas Dakwah Dan Ilmu Komunikasi UIN Raden Intan Lampung. *Jurnal Ilmiah Matrik*, 22(1), 28–35.

- <https://doi.org/10.33557/jurnalmatrik.v22i1.836>
- Setiyani, L. (2020). *Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes : Systematic Review*. 13(1), 35–43. <https://doi.org/10.30998/faktorexacta.v13i1.5548>
- Silalahi, D. K., Murfi, H., & Satria, Y. (2017). *Studi Perbandingan Pemilihan Fitur Untuk Support Vector Machine Pada Klasifikasi Penilaian Risiko Kredit*. 1(2), 119–136.
- Syarif, M., & Nugraha, W. (2023). Mwmote Dalam Mengatasi Ketidakseimbangan Kelas Pada Prediksi Churn Menggunakan Klasifikasi C4.5. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 54–62. <https://doi.org/10.36040/jati.v7i1.6070>
- Untoro, M. C. (2021). *Optimasi MWMOTE Pada Data Tidak Seimbang Menggunakan Complete Linkage MWMOTE Optimization For Imbalanced Data Using Complete Linkage*. 9(January), 77–82. <https://doi.org/10.14710/jtsiskom.2021.13748>
- Untoro, M. C., & Buliali, J. L. (2018). Penanganan Imbalance Class Data Laboratorium Kesehatan Dengan Majority Weighted Minority Oversampling Technique. *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, 4(1), 23–29. <https://doi.org/10.26594/register.v4i1.1184>
- Untoro, M. C., Praseptiawan, M., Widianingsih, M., Ashari, I. F., Afriansyah, A., & Oktafianto. (2020). Evaluation Of Decision Tree, K-NN, Naive Bayes And SVM With MWMOTE On UCI Dataset. *Journal Of Physics: Conference Series*, 1477(3). <https://doi.org/10.1088/1742-6596/1477/3/032005>
- Utomo, D. P. (2020). *Analisis Komparasi Metode Klasifikasi Data Mining Dan Reduksi Atribut Pada Data Set Penyakit Jantung*. 4(April), 437–444. <https://doi.org/10.30865/mib.v4i2.2080>
- UU Nomor 10 Tahun 1998. (1998). Undang-Undang Republik Indonesia Nomor 10 Tahun 1998 Tentang Perbankan. *Bank Indonesia*, 1–65.
- Valentim, I., Lourenco, N., & Antunes, N. (2019). The Impact Of Data Preparation On The Fairness Of Software Systems. *Proceedings - International Symposium On Software Reliability Engineering, ISSRE, 2019-October*, 391–401. <https://doi.org/10.1109/ISSRE.2019.00046>
- Winata, W., Dewi, L. P., & Tjondrowiguno, A. N. (2020). Prediksi Skor Pertandingan Sepak Bola Menggunakan Neuroevolution Of Augmenting Topologies Dan Backpropagation. *Infra*.