

Impact of Normalization Methods on Credit Risk Classification with Imbalanced Data

Ahmad Abil Rizky Ramadhan ^{1*}, Gusti Wiwik Nur Afifa ^{2*}, Mula Agung Barata ³,
Ifnu Wisma Dwi Prastya ⁴, Afril Efan Pajri ⁵

* Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Nahdlatul Ulama Sunan Giri
Ahmadabil527@gmail.com ¹, wiwiknurafifa7@gmail.com ², mula.ab26@gmail.com ³, ifnuprastya@unugiri.ac.id ⁴, afril@unugiri.ac.id ⁵

Article Info	ABSTRACT
Article history: Received 2026-06-16 Revised 2026-09-11 Accepted 2026-10-17 Keyword: ANN, Credit Risk, Imbalanced Data, Normalization, SMOTE, SVM.	Credit risk classification is an essential task in the financial sector to minimize loan default losses and maintain financial system stability. One of the main challenges in building accurate credit risk models is class imbalance, where the number of non-default cases significantly outnumbers default cases, causing classifiers to be biased toward the majority class. This study investigates the effect of two normalization methods, Z-Score and Min-Max, on the performance of Support Vector Machine (SVM) and Artificial Neural Network (ANN) classifiers for imbalanced credit risk prediction. Six experimental scenarios were designed, incorporating the Synthetic Minority Over-sampling Technique (SMOTE), Chi-Square feature selection, and GridSearchCV hyperparameter tuning. Experiments were conducted on a dataset of 4,801 loan records with a class imbalance ratio of 77:23, divided into 80% training and 20% testing using stratified sampling. Evaluation metrics include Accuracy, Precision, Recall, F1-Score, AUC-ROC, Matthews Correlation Coefficient (MCC), and Balanced Accuracy, with F1-Score as the primary metric given the imbalanced class distribution. The results demonstrate that Z-Score normalization, when applied in isolation, consistently outperforms Min-Max across both classifiers, with ANN achieving an accuracy of 95.42% and F1-Score of 0.9332 on the Z-Score scenario without SMOTE. However, the fully optimized pipeline combining Min-Max normalization, Chi-Square feature selection, SMOTE, and hyperparameter tuning surpassed this result, yielding the best overall performance with an ANN accuracy of 95.73%, F1-Score of 0.9393, and AUC-ROC of 0.9610. This indicates that Min-Max's sensitivity to outliers can be effectively mitigated through appropriate feature selection and hyperparameter tuning. These findings provide empirical guidance for selecting optimal preprocessing strategies in imbalanced financial classification tasks.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Sektor keuangan, khususnya perbankan dan *fintech*, menghadapi tantangan serius akibat meningkatnya risiko kredit macet yang dapat mengganggu stabilitas sistem keuangan secara keseluruhan [1]. Prediksi risiko kredit secara otomatis menggunakan *machine learning* menjadi solusi yang semakin diminati karena kemampuannya menganalisis pola kompleks dari data peminjam dalam jumlah besar secara efisien [2]. Di antara berbagai algoritma *machine learning*, *Support Vector Machine* (SVM) dan *Artificial Neural*

Network (ANN) merupakan dua metode yang telah terbukti efektif untuk tugas klasifikasi data keuangan karena kemampuannya menangani data berdimensi tinggi serta hubungan nonlinier antarvariabel [3]. Selain itu, Z-Score dan Min-Max dipilih karena keduanya merupakan metode normalisasi yang paling banyak digunakan dalam literatur klasifikasi data keuangan, sehingga perbandingan sistematis antara keduanya memberikan kontribusi praktis yang langsung relevan bagi para peneliti dan praktisi di bidang ini. Kebaruan penelitian ini terletak pada pengujian interaksi

antara pilihan normalisasi dengan seleksi fitur Chi-Square, SMOTE, dan hyperparameter tuning secara terintegrasi dalam satu pipeline sebuah aspek yang belum pernah dikaji secara sistematis pada konteks klasifikasi risiko kredit dengan data tidak seimbang.

Salah satu tantangan utama dalam klasifikasi risiko kredit adalah ketidakseimbangan kelas (*class imbalance*), di mana jumlah debitur yang gagal bayar jauh lebih sedikit dibandingkan dengan debitur yang lancar [4]. Kondisi ini menyebabkan model *machine learning* cenderung bias terhadap kelas mayoritas, sehingga prediksi didominasi oleh kelas yang lebih banyak [5]. Padahal, kelas minoritas, yaitu debitur yang berpotensi gagal bayar, justru merupakan kelompok yang paling krusial untuk diidentifikasi secara akurat dalam konteks manajemen risiko kredit. Untuk mengatasi permasalahan tersebut, *Synthetic Minority Over-sampling Technique* (SMOTE) telah terbukti menjadi salah satu teknik *oversampling* yang efektif dengan cara membangkitkan sampel sintesis pada kelas minoritas berdasarkan interpolasi antara data yang ada [6].

Selain penanganan ketidakseimbangan kelas, tahap *preprocessing* data, khususnya normalisasi, juga berpengaruh signifikan terhadap kinerja model klasifikasi [7]. Dua metode yang paling umum digunakan adalah *Z-Score Normalization*, yang mentransformasi data ke distribusi standar dengan rata-rata nol dan standar deviasi 1, serta *Min-Max Normalization*, yang menskalakan nilai data ke rentang [0,1] [8]. Pemilihan metode normalisasi yang tepat sangat krusial karena SVM berbasis perhitungan jarak dan ANN yang menggunakan *gradient descent* memiliki sensitivitas yang berbeda terhadap skala fitur masukan [9].

Kajian mengenai prediksi risiko kredit menggunakan SVM dan ANN telah banyak dilakukan dan memberikan hasil yang menjanjikan [1], [3]. Mestiri (2024) membandingkan SVM dan ANN untuk *credit scoring* dan menunjukkan keunggulan ANN pada data tidak seimbang, namun tidak mengkaji pengaruh metode normalisasi secara sistematis [3]. Demikian pula, Suhadolnik et al. (2023) mengevaluasi berbagai pendekatan *machine learning* untuk penilaian risiko kredit, tetapi memperlakukan normalisasi sebagai langkah baku tanpa membandingkan dampaknya secara eksplisit [10]. Mayoritas penelitian yang ada lebih menitikberatkan pada pemilihan algoritma dibandingkan dengan mengkaji dampak teknik normalisasi secara mendalam, terutama ketika *dataset* mengalami ketidakseimbangan kelas [7], [9]. Hal ini menjadi permasalahan signifikan karena pilihan normalisasi dapat memperburuk atau memperbaiki bias model terhadap kelas minoritas secara berbeda pada SVM dan ANN.

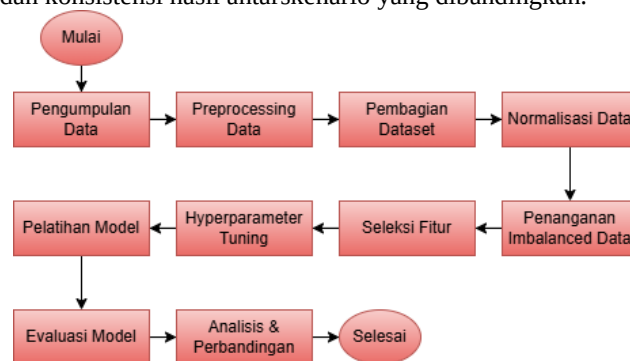
Hingga saat ini, penelitian yang secara spesifik menganalisis perbandingan *Z-Score* dan *Min-Max* dalam satu *pipeline* lengkap yang melibatkan SMOTE, seleksi fitur *Chi-Square*, dan penyesuaian hiperparameter pada SVM dan ANN untuk klasifikasi risiko kredit masih sangat terbatas [5], [8]. Studi yang ada umumnya hanya menguji normalisasi secara terisolasi tanpa mempertimbangkan interaksinya

dengan teknik penanganan data *imbalanced* seperti SMOTE [7].

Penelitian ini bertujuan untuk mengisi kesenjangan tersebut dengan membandingkan pengaruh metode normalisasi *Z-Score* dan *Min-Max* terhadap performa SVM dan ANN dalam klasifikasi risiko kredit pada data yang tidak seimbang. Sebanyak enam skenario eksperimen dirancang secara sistematis dengan mengintegrasikan SMOTE, seleksi fitur berbasis *Chi-Square*, dan penyesuaian hiperparameter menggunakan *GridSearchCV*. Bagian berikut menjelaskan metodologi eksperimen secara rinci, mencakup sumber data, tahap *preprocessing*, desain skenario, konfigurasi model, serta metrik evaluasi yang digunakan.

II. METODE

Penelitian ini menggunakan pendekatan eksperimental komparatif untuk membandingkan pengaruh dua metode normalisasi terhadap kinerja SVM dan ANN dalam klasifikasi risiko kredit. Keseluruhan alur metodologi penelitian disajikan dalam bentuk *flowchart* pada Gambar 1. Setiap tahapan dirancang untuk memastikan validitas eksperimen dan konsistensi hasil antarskenario yang dibandingkan.



Gambar 1. Flowchart Alur Metodologi Penelitian

A. Persiapan dan Sumber Data

Penelitian ini menggunakan *dataset* prediksi risiko kredit yang diperoleh dari platform Kaggle dengan judul *Loan Risk Prediction Dataset* [11], [12]. *Dataset* asli terdiri dari 5.000 baris data dengan 10 fitur, yaitu lima fitur numerik (Age, Income, LoanAmount, CreditScore, YearsExperience) dan empat fitur kategorikal (Gender, Education, City, EmploymentType), serta satu variabel target (LoanApproved) [13]. Analisis awal menunjukkan adanya *missing value* pada fitur Income (196 nilai), CreditScore (194 nilai), dan Education (198 nilai), serta terdapat 3 baris data dengan nilai Income negatif yang dihapus karena tidak valid secara logis [13]. Setelah proses pembersihan data, *dataset* bersih terdiri dari 4.801 sampel dengan distribusi kelas tidak seimbang, yaitu 3.698 sampel kelas 0 (tidak disetujui) dan 1.103 sampel kelas 1 (disetujui), sehingga menghasilkan rasio ketidakseimbangan sebesar 77:23 [4]. Rincian fitur *dataset* disajikan pada Tabel I.

TABEL I

DESKRIPSI FITUR DATASET LOAN RISK PREDICTION

No	Fitur	Tipe Data	Keterangan
1	Age	Numerik	Usia peminjam (tahun)
2	Income	Numerik	Pendapatan tahunan peminjam
3	LoanAmount	Numerik	Jumlah pinjaman yang diajukan
4	CreditScore	Numerik	Skor kredit peminjam
5	YearsExperience	Numerik	Lama pengalaman kerja (tahun)
6	Gender	Kategorikal	Jenis kelamin peminjam
7	Education	Kategorikal	Tingkat pendidikan terakhir
8	City	Kategorikal	Kota domisili peminjam
9	EmploymentType	Kategorikal	Jenis pekerjaan peminjam
10	LoanApproved	Target	0 = Ditolak, 1 = Disetujui

Sumber: Kaggle Loan Risk Prediction Dataset (diolah, 2024)

B. Tahap Preprocessing Data

Tahap preprocessing dilakukan secara bertahap. Pertama, 3 baris data dengan nilai Income negatif dihapus karena tidak valid secara logis. Kedua, missing value ditangani menggunakan imputasi rata-rata (mean imputation) untuk fitur numerik dan imputasi modus untuk fitur kategorikal [13]. Ketiga, normalisasi data dilakukan menggunakan dua metode: Z-Score Normalization yang mentransformasi data ke distribusi standar dengan rata-rata nol dan standar deviasi satu, serta Min-Max Normalization yang menskalakan nilai ke rentang [0,1] [8]. Secara matematis:

$$z = (x - \mu) / \sigma \quad (1)$$

$$x' = (x - x_m^{ln}) / (x_m^{ax} - x_m^{ln}) \quad (2)$$

Di mana z adalah hasil Z-Score, x' adalah hasil Min-Max, x adalah nilai asli, μ adalah rata-rata, σ adalah standar deviasi, x_m^{ln} adalah nilai minimum, dan x_m^{ax} adalah nilai maksimum fitur [7]. Keempat, fitur kategorikal dikonversi menggunakan One-Hot Encoding [13]. Kelima, dataset dibagi menjadi data latih (80%) dan data uji (20%) menggunakan stratified split dengan random_state = 42 untuk mempertahankan proporsi kelas [14]. SMOTE hanya diterapkan pada data latih untuk menghindari data leakage [6]. Urutan implementasi preprocessing secara lengkap adalah: (1) train-test split dilakukan terlebih dahulu dengan stratified sampling; (2) normalisasi di-fit hanya pada data latih kemudian diterapkan ke data uji menggunakan parameter yang sama; (3) seleksi fitur Chi-Square di-fit hanya pada data latih; (4) SMOTE diterapkan hanya pada data latih setelah seleksi fitur; dan (5) GridSearchCV dijalankan dengan 3-fold cross-validation pada data latih yang telah di-oversampling. Pendekatan ini

memastikan tidak ada informasi dari data uji yang bocor ke proses pelatihan model. [6].

Hasil preprocessing menghasilkan total 22 fitur setelah One-Hot Encoding diterapkan pada empat fitur kategorikal. Distribusi fitur ini menjadi dasar perancangan enam skenario eksperimen berikut.

C. Skenario Eksperimen

Enam skenario eksperimen dirancang secara sistematis sebagaimana disajikan pada Tabel II. Skenario 1 (Baseline) tidak menerapkan normalisasi sebagai pembandingan awal. Skenario 2 menerapkan Z-Score Normalization menggunakan StandardScaler. Skenario 3 menerapkan Min-Max Normalization menggunakan MinMaxScaler. Skenario 4 mengombinasikan Z-Score dengan SMOTE. Skenario 5 mengombinasikan Min-Max dengan SMOTE. Skenario 6 (Final) merupakan pipeline lengkap yang mengintegrasikan Min-Max Normalization, seleksi fitur Chi-Square ($k = 10$ fitur terpilih berdasarkan analisis elbow), SMOTE, dan hyperparameter tuning menggunakan GridSearchCV [15], [16].

TABEL II
SKENARIO EKSPERIMEN PENELITIAN

N o	Skenario	Nor m.	SMOT E	Fitur	Tuning
1	Baseline	—	—	—	—
2	Z-Score	Z-Score	—	—	—
3	Min-Max	Min-Max	—	—	—
4	ZS+SMOTE	Z-Score	✓	—	—
5	MM+SMOTE	Min-Max	✓	—	—
6	Final	Min-Max	✓	Chi-Square	GridSearchCV

Keterangan: ✓ = diterapkan, — = tidak diterapkan

D. Konfigurasi Model SVM dan ANN

Model Support Vector Machine (SVM) dikonfigurasi menggunakan kernel Radial Basis Function (RBF) dengan parameter C dioptimalkan melalui GridSearchCV pada rentang {0.1, 1, 10} dan cross-validation 3-fold dengan metrik optimasi F1-Score [15]. Konfigurasi terbaik SVM diperoleh pada $C = 10$ dengan kernel RBF [3]. Pada Skenario 1 hingga 5, SVM menggunakan nilai $C = 1,0$ (default) dengan gamma = scale. Ruang pencarian GridSearchCV {C: 0.1, 1, 10} dipilih untuk mencakup rentang dari margin lebar (C kecil, bias lebih besar) hingga margin sempit (C besar, fitting lebih ketat), sehingga model dapat menemukan keseimbangan optimal antara bias dan variansi.

Model Artificial Neural Network (ANN) dibangun menggunakan arsitektur Sequential dengan dua hidden layer berukuran 64-32 neuron yang dipilih berdasarkan panduan pengurangan bertahap untuk mendorong representasi fitur

yang semakin abstrak sesuai dengan 22 fitur masukan [17]. Kedua *layer* menggunakan fungsi aktivasi *ReLU*, diikuti oleh teknik *Dropout* sebesar 0,3 untuk mencegah *overfitting* [18]. *Layer output* menggunakan fungsi aktivasi *sigmoid* untuk klasifikasi biner. Model dilatih menggunakan *optimizer* Adam dengan fungsi *loss binary cross-entropy* dan *early stopping* (*patience* = 5), dengan maksimum *epoch* 50 (Skenario 1–3) dan 100 (Skenario 4–6) untuk mengakomodasi ukuran *training set* yang lebih besar setelah *oversampling* [19]. Pada Skenario 6 (Final), setelah seleksi fitur Chi-Square menghasilkan 10 fitur masukan, grid search manual dilakukan dengan ruang pencarian neurons $\in \{64, 128\}$ dan epochs $\in \{10, 30\}$, menghasilkan konfigurasi terbaik neurons = 128 dan epochs = 30 dengan arsitektur Input(10) -> Dense(128, ReLU) -> Dropout(0,3) -> Dense(64, ReLU) -> Dropout(0,3) -> Dense(1, Sigmoid). Learning rate yang digunakan adalah 0,001 (default Adam) dengan batch size 32.

E. Metrik Evaluasi

Evaluasi kinerja model menggunakan lima metrik berdasarkan *confusion matrix*: Accuracy, Precision, Recall, F1-Score, dan AUC-ROC [17]. F1-Score dipilih sebagai metrik utama karena mengukur keseimbangan antara Precision dan Recall secara bersamaan, sehingga lebih representatif pada kondisi data *imbalanced* [5]. AUC-ROC digunakan untuk mengukur kemampuan diskriminasi model secara keseluruhan tanpa bergantung pada nilai *threshold* [19]. F1-Score dihitung menggunakan pendekatan *macro-average* agar setiap kelas mendapat bobot yang sama [17]. Sebagai metrik pendukung, Matthews Correlation Coefficient (MCC) dan Balanced Accuracy juga dihitung untuk memberikan evaluasi yang lebih komprehensif pada kondisi data tidak seimbang — MCC mempertimbangkan semua empat kuadran *confusion matrix* secara seimbang, sedangkan Balanced Accuracy menghitung rata-rata *sensitivity* dan *specificity* tanpa terpengaruh ketidakseimbangan kelas. Persamaan metrik utama yang digunakan adalah sebagai berikut:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (3)$$

$$Precision = TP / (TP + FP) \quad (4)$$

$$Recall = TP / (TP + FN) \quad (5)$$

$$F1-Score = (2 \times Precision \times Recall) / (Precision + Recall) \quad (6)$$

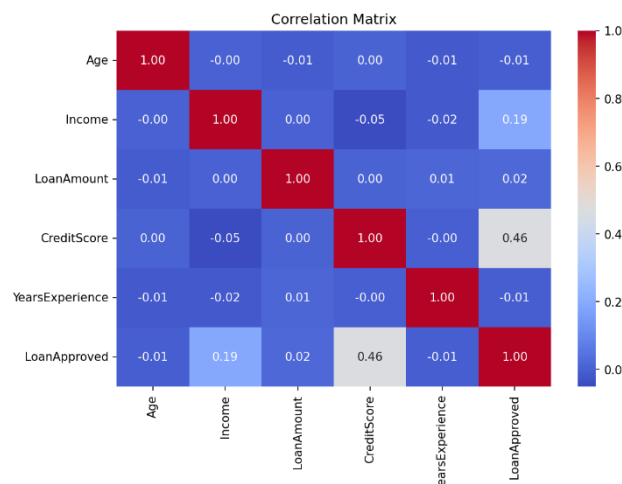
Di mana TP (*True Positive*) adalah data positif yang diprediksi benar, TN (*True Negative*) adalah data negatif yang diprediksi benar, FP (*False Positive*) adalah data negatif yang diprediksi positif, dan FN (*False Negative*) adalah data positif yang diprediksi negatif.

III. HASIL DAN PEMBAHASAN

A. Analisis Dataset dan Hasil Preprocessing

Setelah tahap *preprocessing*, *dataset* bersih yang digunakan dalam eksperimen terdiri dari 4.801 sampel

dengan distribusi kelas tidak seimbang, yaitu 3.698 sampel kelas 0 (tidak disetujui) dan 1.103 sampel kelas 1 (disetujui), menghasilkan rasio 77:23. Kondisi ini mengonfirmasi perlunya penanganan *imbalanced data* sebelum pelatihan model agar *classifier* tidak bias terhadap kelas mayoritas. Analisis korelasi antarvariabel divisualisasikan pada Gambar 2.



Gambar 2. Matriks Korelasi Antarvariabel Dataset

Berdasarkan Gambar 2, fitur *CreditScore* memiliki korelasi tertinggi terhadap variabel target *LoanApproved* dengan nilai 0,46. *Income* berada di posisi berikutnya dengan nilai korelasi 0,19. Fitur *Age*, *LoanAmount*, dan *YearsExperience* menunjukkan korelasi yang sangat rendah, mendekati nol. Ini mengindikasikan minimnya kontribusi langsung terhadap keputusan persetujuan kredit. Tidak ditemukan *multicollinearity* yang signifikan antarvariabel prediktor. Dengan demikian, dataset layak digunakan untuk pelatihan model klasifikasi. SMOTE pada data latih berhasil menyeimbangkan distribusi kelas minoritas. Seleksi fitur Chi-Square pada Skenario 6 menghasilkan 10 fitur terpilih yang berkontribusi paling signifikan terhadap prediksi risiko kredit. Fitur-fitur tersebut meliputi *EmploymentType_Unemployed* (skor Chi-Square tertinggi = 288,00), *CreditScore* (131,73), *EmploymentType_Salaried* (85,95), *EmploymentType_Self-Employed* (63,82), *Income* (6,05), *Education_PhD* (4,95), *Education_High School* (2,46), *Education_Bachelors* (1,63), *City_Houston* (0,98), dan *Gender_Female* (0,64). Rincian skor Chi-Square seluruh fitur terpilih disajikan pada Tabel IV.

Dominasi *CreditScore* dan *Income* sebagai prediktor utama memberikan konteks penting untuk memahami mengapa seleksi fitur pada Skenario 6 berkontribusi pada peningkatan performa. Temuan ini memiliki implikasi praktis bagi institusi keuangan: sistem credit scoring dapat difokuskan pada fitur-fitur yang paling informatif ini untuk mengoptimalkan akurasi prediksi sekaligus mengurangi kompleksitas komputasi. Analisis perbandingan performa seluruh skenario disajikan pada subbagian berikut.

B. Perbandingan Performa Berdasarkan Metode Normalisasi

Tabel III menyajikan perbandingan performa SVM dan ANN berdasarkan Accuracy, Precision, Recall, F1-Score, AUC-ROC, Matthews Correlation Coefficient (MCC), dan Balanced Accuracy pada keenam skenario eksperimen.

TABEL III

PERBANDINGAN PERFORMA SVM DAN ANN PADA ENAM SKENARIO EKSPERIMEN

N o	Skenari o SVM	Acc	Pre c	Rec	F1	AU C	MC C	Bal. Acc
1	Baseline	0.7700	0.3850	0.5000	0.4350	0.8870	0.0000	0.5000
2	Z-Score	0.9428	0.9372	0.8978	0.9155	0.9484	0.8340	0.8978
3	Min-Max	0.8949	0.8581	0.8381	0.8474	0.9379	0.6959	0.8381
4	ZS+SM OTE	0.9220	0.8833	0.9033	0.8927	0.9465	0.7864	0.9033
5	MM+S MOTE	0.8595	0.8007	0.8659	0.8227	0.9379	0.6634	0.8659
6	Final	0.9022	0.8509	0.9048	0.8723	0.9587	0.7537	0.9048

N o	Skenari o ANN	Acc	Pre c	Rec	F1	AU C	MC C	Bal. Acc
1	Baseline	0.7700	0.3850	0.5000	0.4350	0.5020	0.0000	0.5000
2	Z-Score	0.9542	0.9491	0.9195	0.9332	0.9541	0.8681	0.9195
3	Min-Max	0.8907	0.8585	0.8212	0.8375	0.9385	0.6787	0.8212
4	ZS+SM OTE	0.9553	0.9362	0.9376	0.9369	0.9574	0.8739	0.9376
5	MM+S MOTE	0.8772	0.8210	0.8885	0.8445	0.9474	0.7063	0.8885
6	Final	0.9573	0.9429	0.9358	0.9393	0.9610	0.8787	0.9393

Keterangan: Acc=Accuracy, Prec=Precision, Rec=Recall, F1=F1-Score (macro-avg), AUC=AUC-ROC, MCC=Matthews Correlation Coefficient, Bal.Acc=Balanced Accuracy; baris kuning=performa terbaik

TABEL IV

10 FITUR TERPILIH BERDASARKAN SELEKSI CHI-SQUARE (SKENARIO 6)

No	Fitur	Chi-Square Score
1	EmploymentType_Unemployed	288.0031
2	CreditScore	131.7318
3	EmploymentType_Salaried	85.9504
4	EmploymentType_Self-Employed	63.8221
5	Income	6.0543
6	Education_PhD	4.9495
7	Education_High School	2.4573
8	Education_Bachelors	1.6299
9	City_Houston	0.9848
10	Gender_Female	0.6453

Sumber: Hasil seleksi Chi-Square pada data latih (diolah, 2024)

Berdasarkan Tabel III, Skenario 1 (*Baseline*) tanpa normalisasi menghasilkan Accuracy 77,00% pada kedua model dengan F1-Score yang sangat rendah sebesar 0,4350 dan Precision 0,3850. Kondisi ini terjadi karena tanpa normalisasi, distribusi fitur yang tidak seragam menyebabkan kedua model memprediksi hampir seluruh sampel sebagai kelas mayoritas, sehingga F1-Score kolaps meskipun Accuracy tampak tidak terlalu buruk akibat dominasi kelas 0.

Penerapan *Z-Score Normalization* pada Skenario 2 meningkatkan performa secara signifikan: SVM mencapai Accuracy 94,28%, Precision 0,9372, Recall 0,8978, dan F1-Score 0,9155 (AUC 0,9484), sementara ANN mencapai Accuracy 95,42%, Precision 0,9491, Recall 0,9195, F1-Score 0,9332, dan AUC 0,9541. Peningkatan yang besar ini menunjukkan bahwa normalisasi adalah prasyarat kritis bagi kedua *classifier*. Sebaliknya, *Min-Max Normalization* pada Skenario 3 menghasilkan performa yang lebih rendah: SVM mencapai Accuracy 89,49%, Precision 0,8581, Recall 0,8381 dan F1-Score 0,8474 (AUC 0,9379), sedangkan ANN mencapai Accuracy 89,07%, Precision 0,8585, Recall 0,8212 dan F1-Score 0,8375 (AUC 0,9385). Selisih F1-Score antara *Z-Score* dan *Min-Max* mencapai sekitar 0,08 poin untuk kedua model, selisih yang cukup substansial untuk aplikasi deteksi risiko kredit. Perbedaan performa ini dikonfirmasi secara statistik melalui uji Wilcoxon signed-rank test berbasis bootstrap ($n=1.000$) antara ANN *Z-Score* dan ANN *Min-Max*, yang menunjukkan bahwa perbedaan F1-Score antara kedua metode normalisasi signifikan secara statistik ($p < 0,001$, Wilcoxon statistic = 0,000) pada $\alpha = 0,05$, mengindikasikan bahwa keunggulan *Z-Score* bukan merupakan hasil kebetulan.

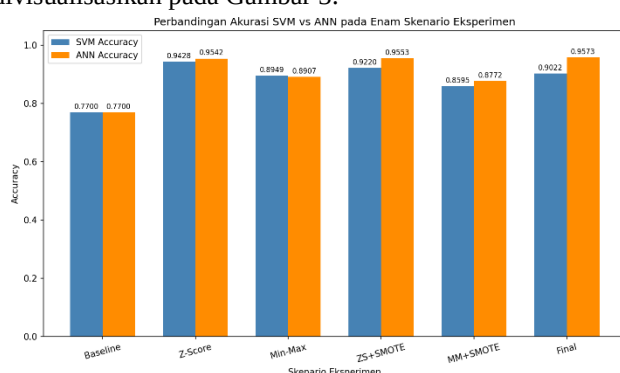
Perbedaan ini dapat dijelaskan melalui karakteristik masing-masing metode: *Z-Score* mentransformasi data ke distribusi standar sehingga lebih tahan terhadap *outlier* dan distribusi yang tidak normal, dua kondisi yang umum pada data keuangan nyata. *Min-Max* menskalakan data ke rentang tetap $[0,1]$ sehingga sangat sensitif terhadap nilai ekstrem. Satu *outlier* dapat menekan seluruh distribusi fitur ke rentang yang sempit. Hal ini berdampak pada SVM yang berbasis perhitungan jarak dan ANN yang menggunakan *gradient descent*, karena skala fitur yang terdistorsi mengganggu konvergensi optimasi kedua algoritma tersebut.

C. Pengaruh SMOTE terhadap Performa Model

Penerapan SMOTE menghasilkan dampak yang berbeda tergantung pada metode normalisasi yang digunakan. Pada Skenario 4 (*Z-Score+SMOTE*), SMOTE menghasilkan performa yang bervariasi dibandingkan dengan Skenario 2 tanpa SMOTE: Accuracy SVM turun dari 94,28% menjadi 92,20% (Precision 0,8833, Recall 0,9033, F1-Score 0,8927, AUC 0,9465), sedangkan Accuracy ANN sedikit meningkat dari 95,42% menjadi 95,53% (Precision 0,9362, Recall 0,9376, F1-Score 0,9369, AUC 0,9574). Hal ini mengindikasikan bahwa sampel sintesis yang dibangkitkan

SMOTE dalam ruang fitur *Z-Score* kurang optimal untuk SVM berbasis jarak pada konfigurasi ini.

Pada Skenario 5 (*Min-Max*+SMOTE), penerapan SMOTE justru menurunkan performa kedua model dibandingkan dengan Skenario 3: Accuracy SVM turun dari 89,49% menjadi 85,95% (Precision 0,8007, Recall 0,8659, F1-Score 0,8227, AUC 0,9379), dan Accuracy ANN turun dari 89,07% menjadi 87,72% (Precision 0,8210, Recall 0,8885, F1-Score 0,8445, AUC 0,9474). Temuan ini menunjukkan bahwa efektivitas SMOTE sangat bergantung pada metode normalisasi yang mendahuluinya: *Z-Score*+SMOTE meningkatkan performa ANN, namun menurunkan SVM, sementara *Min-Max*+SMOTE menurunkan performa kedua model. Oleh karena itu, pemilihan kombinasi normalisasi dan SMOTE perlu mempertimbangkan karakteristik algoritma yang digunakan. Perbandingan akurasi seluruh skenario divisualisasikan pada Gambar 3.



Gambar 3. Perbandingan Akurasi SVM dan ANN pada Enam Skenario Eksperimen

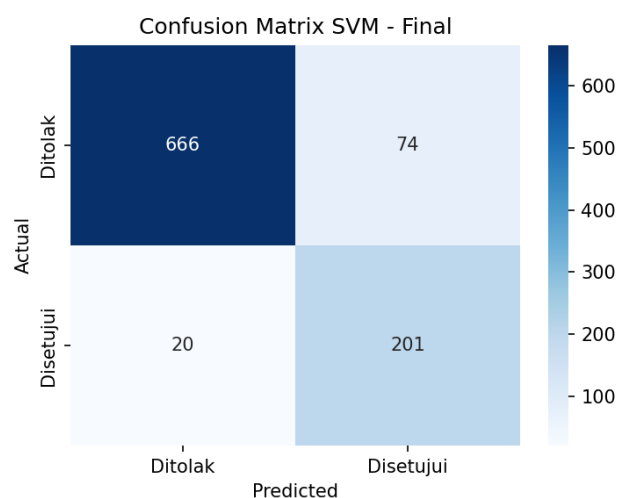
D. Perbandingan Performa SVM dan ANN serta Model Terbaik

Secara keseluruhan, ANN menunjukkan performa yang lebih unggul dibandingkan dengan SVM pada hampir seluruh skenario, terutama pada skenario yang melibatkan normalisasi dan SMOTE. Hal ini sejalan dengan karakteristik ANN yang mampu menangkap hubungan nonlinier yang lebih kompleks dalam data keuangan. Skenario 6 (Final) menghasilkan performa terbaik secara keseluruhan dengan ANN mencapai Accuracy 95,73%, Precision 0,9429, Recall 0,9358, F1-Score 0,9393, dan AUC-ROC 0,9610; sementara SVM mencapai Accuracy 90,22%, Precision 0,8509, Recall 0,9048, dan F1-Score 0,8723 (AUC 0,9587, MCC 0,7537, dan Balanced Accuracy 0,9048). Nilai MCC ANN Final sebesar 0,8787 mengindikasikan korelasi yang sangat kuat antara prediksi model dengan label aktual, mengonfirmasi reliabilitas model meskipun data tidak seimbang. Pendekatan ini sejalan dengan studi Chang et al. (2024) yang menunjukkan efektivitas kombinasi ML dan DL dalam prediksi risiko kredit.

Hasil ini menarik dicermati karena *Min-Max* yang secara individual lebih lemah dibandingkan *Z-Score* justru menghasilkan performa tertinggi ketika dikombinasikan dengan seleksi fitur, SMOTE, dan *tuning*. Dua mekanisme menjelaskan hal ini: pertama, seleksi fitur *Chi-Square*

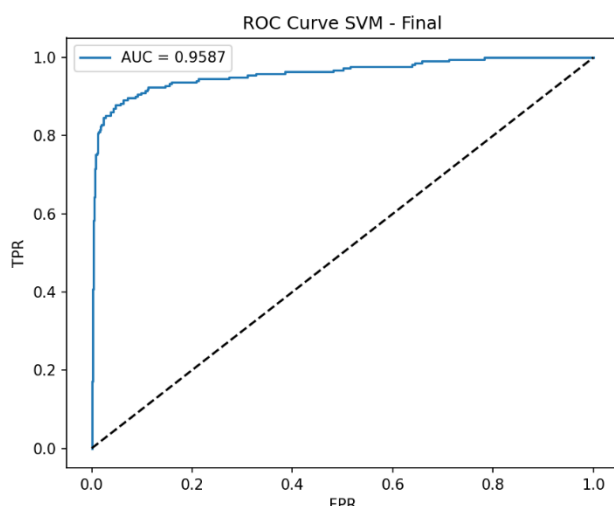
mengeliminasi fitur yang mengandung *outlier* ekstrem justru pada kondisi di mana *Min-Max* paling rentan sehingga kelemahan utamanya berhasil dimitigasi; kedua, *hyperparameter tuning* melalui GridSearchCV mengoptimalkan parameter ANN secara spesifik untuk distribusi *Min-Max*, menghasilkan konfigurasi yang lebih baik.

Hasil *confusion matrix* model ANN Final pada Gambar 4 menunjukkan bahwa dari 961 data uji, model berhasil mengklasifikasikan 722 sampel kelas 0 dengan benar (True Negative), 198 sampel kelas 1 dengan benar (True Positive), 18 sampel kelas 0 salah diprediksi sebagai kelas 1 (False Positive), dan 23 sampel kelas 1 salah diprediksi sebagai kelas 0 (False Negative).



Gambar 4. Confusion Matrix Model ANN Final

Nilai *False Negative* sebesar 23 menunjukkan bahwa model hanya melewatkan 23 peminjam berpotensi *default* dari total 221 peminjam *default* — setara dengan error rate 10,41% untuk kelas *default* — pencapaian yang sangat baik dalam konteks manajemen risiko kredit, di mana *False Negative* merupakan kesalahan yang paling merugikan karena setiap peminjam berisiko tinggi yang tidak terdeteksi berpotensi mengakibatkan kerugian kredit macet bagi institusi keuangan. Sebaliknya, *False Positive* sebesar 18 berarti hanya 18 peminjam layak yang ditolak — menimbulkan *opportunity cost* namun tidak mengancam stabilitas portofolio kredit secara langsung. Kemampuan diskriminasi model dikonfirmasi melalui kurva ROC pada Gambar 5 dengan AUC sebesar 0,9610, yang mengindikasikan kemampuan membedakan kelas yang sangat tinggi, sehingga konfigurasi model final ini dinilai sesuai untuk implementasi praktis dalam sistem credit scoring.



Gambar 5. ROC Curve Model ANN Final (AUC = 0,9610)

Perbandingan dengan penelitian terdahulu menunjukkan bahwa hasil penelitian ini kompetitif. Zhang et al. (2024) menerapkan pendekatan *imbalanced data* berbasis SMOTE yang ditingkatkan, namun tanpa mengintegrasikan analisis normalisasi secara sistematis dalam *pipeline*-nya. Mestiri (2024) juga menunjukkan keunggulan ANN dalam *credit scoring* tanpa membandingkan dampak pilihan normalisasi secara eksplisit. Selain itu, studi terdahulu seperti Chang et al. [20], Muhindo et al. [21], Chen et al. [22], Aruleba dan Sun [23], serta Ghoudam et al. [24] juga mengonfirmasi efektivitas berbagai pendekatan machine learning untuk prediksi risiko kredit. Temuan penelitian ini menegaskan bahwa pemilihan strategi *preprocessing* yang tepat, khususnya kombinasi normalisasi, seleksi fitur, SMOTE, dan *hyperparameter tuning*, memberikan kontribusi signifikan terhadap peningkatan performa model klasifikasi risiko kredit pada data tidak seimbang [25], [26].

IV. KESIMPULAN

Penelitian ini telah membandingkan pengaruh metode normalisasi Z-Score dan Min-Max terhadap performa *Support Vector Machine* (SVM) dan *Artificial Neural Network* (ANN) dalam klasifikasi risiko kredit pada data yang tidak seimbang menggunakan enam skenario eksperimen. Hasil penelitian menunjukkan bahwa Z-Score Normalization secara individual, tanpa kombinasi seleksi fitur dan tuning tambahan, secara konsisten menghasilkan performa yang lebih baik dibandingkan dengan Min-Max Normalization pada kedua *classifier*, dengan ANN mencapai *Accuracy* tertinggi sebesar 95,42% dan F1-Score sebesar 0,9332 pada skenario Z-Score tanpa SMOTE. Temuan ini mengonfirmasi bahwa Z-Score Normalization lebih tahan terhadap distribusi data yang tidak normal dan keberadaan *outlier* dibandingkan dengan Min-Max Normalization ketika diterapkan secara langsung tanpa proses optimasi lanjutan.

Namun, *pipeline* eksperimen final yang mengintegrasikan Min-Max Normalization, seleksi fitur berbasis Chi-Square, SMOTE, dan *hyperparameter tuning* menggunakan

GridSearchCV justru melampaui hasil tersebut, menghasilkan performa terbaik secara keseluruhan dengan ANN mencapai *Accuracy* 95,73%, F1-Score 0,9393, AUC-ROC 0,9610, MCC 0,8787, dan *Balanced Accuracy* 0,9358. Temuan ini menunjukkan bahwa kelemahan Min-Max Normalization terhadap *outlier* dapat dimitigasi secara efektif melalui seleksi fitur dan *hyperparameter tuning* yang tepat, sehingga kombinasi strategi *preprocessing* yang menyeluruh berkontribusi secara signifikan terhadap peningkatan performa klasifikasi, terutama pada kondisi *imbalanced data* di mana F1-Score menjadi metrik evaluasi yang lebih representatif dibandingkan *Accuracy*.

Secara keseluruhan, ANN terbukti lebih unggul dibandingkan dengan SVM pada hampir seluruh skenario eksperimen, mengindikasikan kemampuan ANN yang lebih baik dalam menangkap pola nonlinier pada data risiko kredit. Penelitian ini memberikan panduan empiris bagi praktisi dan peneliti dalam memilih strategi *preprocessing* yang optimal untuk tugas klasifikasi data keuangan yang tidak seimbang. Meskipun demikian, penelitian ini memiliki beberapa keterbatasan yang membuka peluang pengembangan lebih lanjut. Pertama, normalisasi yang diuji hanya Z-Score dan Min-Max, sehingga metode lain seperti Robust Scaling yang lebih tahan terhadap *outlier* ekstrem belum dieksplorasi. Kedua, penggunaan satu dataset tunggal dari Kaggle dengan rasio ketidakseimbangan 77:23 membatasi kemampuan generalisasi hasil; temuan ini mungkin tidak berlaku secara universal pada dataset kredit dari institusi lain yang memiliki karakteristik distribusi data berbeda, sehingga validasi pada dataset eksternal diperlukan sebelum penerapan sistem ini di lingkungan operasional yang berbeda. Ketiga, model tunggal seperti ANN dan SVM berpotensi ditingkatkan lebih jauh melalui teknik *ensemble learning* yang menggabungkan kekuatan berbagai *classifier*.

DAFTAR PUSTAKA

- [1] A. Sapaatullah, R. Rahmat, and M. Darip, "Analisis Performa Support Vector Machine untuk Klasifikasi Risiko Kredit Nasabah pada Perbankan Daerah," *Bull. Inf. Technol.*, vol. 7, no. 1, pp. 84–91, 2026, doi: 10.47065/bit.v7i1.2603.
- [2] J. P. Noriega, L. A. Rivera, and J. A. Herrera, "Machine Learning for Credit Risk Prediction: A Systematic Literature Review," *Data*, vol. 8, no. 11, p. 169, 2023, doi: 10.3390/data8110169.
- [3] S. Mestiri, "Credit scoring using machine learning and deep Learning-Based models," *Data Sci. Financ. Econ.*, vol. 4, no. 2, pp. 236–248, 2024, doi: 10.3934/dsfe.2024009.
- [4] Z. Gu, J. Lv, B. Wu, Z. Hu, and X. Yu, "Credit risk assessment of small and micro enterprise based on machine learning," *Heliyon*, vol. 10, no. 5, p. e27096, 2024, doi: 10.1016/j.heliyon.2024.e27096.
- [5] B. Nemade, V. Bharadi, S. S. Alegavi, and B. Marakarkandy, "A Comprehensive Review: SMOTE-Based Oversampling Methods for Imbalanced Classification Techniques, Evaluation, and Result Comparisons," *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 9s, pp. 378–390, 2023.
- [6] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.
- [7] R. R. Laska and A. M. Yolanda, "A Comparative Study of Z-Score and Min-Max Normalization for Rainfall Classification in Pekanbaru," *J. Data Sci. Informatics*, vol. 2024, no. 4, 2024, doi:

- 10.61453/jods.v2024no04.
- [8] A. Demircioğlu, "The effect of feature normalization methods in radiomics," *Insights Imaging*, vol. 15, no. 1, p. 2, 2024, doi: 10.1186/s13244-023-01575-7.
- [9] S. Sinsomboonthong, "Performance Comparison of New Adjusted Min-Max with Decimal Scaling and Statistical Column Normalization Methods for Artificial Neural Network Classification," *Int. J. Math. Math. Sci.*, vol. 2022, p. 3584406, 2022, doi: 10.1155/2022/3584406.
- [10] N. Suhadolnik, J. Ueyama, and S. Da Silva, "Machine Learning for Enhanced Credit Risk Assessment: An Empirical Approach," *J. Risk Financ. Manag.*, vol. 16, no. 12, p. 496, 2023, doi: 10.3390/jrfm16120496.
- [11] T. Berhane, T. Melese, and A. M. Seid, "Performance Evaluation of Hybrid Machine Learning Algorithms for Online Lending Credit Risk Prediction," *Appl. Artif. Intell.*, vol. 38, no. 1, p. 2358661, 2024, doi: 10.1080/08839514.2024.2358661.
- [12] taweilo, "Loan Approval Classification Data," Kaggle. [Online]. Available: <https://www.kaggle.com/datasets/taweilo/loan-approval-classification-data>
- [13] D. Breskuvienė and G. Dzemyda, "Categorical Feature Encoding Techniques for Improved Classifier Performance when Dealing with Imbalanced Data of Fraudulent Transactions," *Int. J. Comput. Commun. Control*, vol. 18, no. 3, p. 5433, 2023, doi: 10.15837/ijccc.2023.3.5433.
- [14] A. A. Khan, "Balanced Split: A New Train-Test Data Splitting Strategy for Imbalanced Datasets," 2022, *arXiv*. [Online]. Available: <http://arxiv.org/abs/2212.11116>
- [15] A. Abdo, R. Mostafa, and L. Abdel-Hamid, "An Optimized Hybrid Approach for Feature Selection Based on Chi-Square and Particle Swarm Optimization Algorithms," *Data*, vol. 9, no. 2, p. 20, 2024, doi: 10.3390/data9020020.
- [16] R. A. N. Diaz, I. K. G. D. Putra, M. Sudarma, I. M. Sukarsa, and N. Jawas, "Comparison of Gain Ratio and Chi-Square Feature Selection Methods in Improving SVM Performance on IDS," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 15, no. 1, p. 64, 2024, doi: 10.24843/lkjiti.2024.v15.i01.p06.
- [17] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Sci. Rep.*, vol. 14, no. 1, p. 6086, 2024, doi: 10.1038/s41598-024-56706-x.
- [18] I. Salehin and D.-K. Kang, "A Review on Dropout Regularization Approaches for Deep Neural Networks within the Scholarly Domain," *Electron.*, vol. 12, no. 14, p. 3106, 2023, doi: 10.3390/electronics12143106.
- [19] J. D. Turiel and T. Aste, "Peer-to-Peer Loan Acceptance and Default Prediction with Artificial Intelligence," *R. Soc. Open Sci.*, vol. 7, no. 6, p. 191649, 2020, doi: 10.1098/rsos.191649.
- [20] V. Chang, S. Sivakulasingam, H. Wang, S. T. Wong, M. A. Ganatra, and J. Luo, "Credit Risk Prediction Using Machine Learning and Deep Learning: A Study on Credit Card Customers," *Risks*, vol. 12, no. 11, p. 174, 2024, doi: 10.3390/risks12110174.
- [21] J. Muhindo, K. Mukasa, D. Kitakufe, and J. Kato, "Advancing credit risk assessment and financial decision-making: Integrating modern techniques and insights," *World J. Adv. Res. Rev.*, vol. 23, no. 2, pp. 2019–2027, 2024, doi: 10.30574/wjarr.2024.23.2.2565.
- [22] Y. Chen, R. Calabrese, and B. Martin-Barragan, "Interpretable machine learning for imbalanced credit scoring datasets," *Eur. J. Oper. Res.*, vol. 312, no. 1, pp. 357–372, 2024, doi: 10.1016/j.ejor.2023.06.036.
- [23] I. Aruleba and Y. Sun, "Effective Credit Risk Prediction Using Ensemble Classifiers With Model Explanation," *IEEE Access*, vol. 12, pp. 115015–115025, 2024, doi: 10.1109/ACCESS.2024.3445308.
- [24] M. E. K. Ghoudam, R. Chaabita, O. Elkhalfi, K. Zehraoui, H. Elalaoui, and S. Idamia, "Consumer credit risk analysis through artificial intelligence: a comparative study between the classical approach of logistic regression and advanced machine learning techniques," *Cogent Econ. Financ.*, vol. 12, no. 1, p. 2414926, 2024, doi: 10.1080/23322039.2024.2414926.
- [25] Y. Zhang, L. Deng, and B. Wei, "Imbalanced Data Classification Based on Improved Random-SMOTE and Feature Standard Deviation," *Mathematics*, vol. 12, no. 11, p. 1709, 2024, doi: 10.3390/math12111709.
- [26] J. Pardede and R. Dwianto, "The Effect of Feature Selection on Machine Learning Classification," *Int. J. Informatics Vis.*, vol. 9, no. 4, pp. 1419–1429, 2025, doi: 10.62527/joiv.9.4.2926.