DATAMATIKA

Tersedia secara daring di:
https://conference.ipb.ac.id/datamatika

PREDIKSI CREDIT DEFAULT PADA DATA TIDAK SEIMBANG MENGGUNAKAN REGRESI LOGISTIK DAN ARTIFICIAL NEURAL NETWORK

CREDIT DEFAULT PREDICTION ON IMBALANCED DATA USING LOGISTIC REGRESSION AND ARTIFICIAL NEURAL NETWORK

Nyayu Azzahra Nabila[1‡], Dewi Muliani[1], Raihanah Nurkhalishah[1], Raden Roro Jessica Faaris[1] and Septian Rahardiantoro[1]

[1,2]Study Program of Statistics and Data Science, IPB University, Indonesia
[‡]corresponding author: nyayunabila@apps.ipb.ac.id

Abstract

Credit default prediction is an important component of credit risk management, particularly in imbalanced datasets dominated by the non-default class. This study aimed to compare the performance of Logistic Regression, Artificial Neural Network (ANN), and ANN with hyperparameter tuning in predicting credit default on imbalanced data using sensitivity and Area Under Curve (AUC) as the primary performance metrics. The study used the Default of Credit Card Clients Dataset from the UCI Machine Learning Repository consisting of 30,000 observations. Data preprocessing included One-Hot Encoding, Min-Max Scaling, and class imbalance handling using the Synthetic Minority Oversampling Technique (SMOTE) on the training data. Hyperparameter tuning for ANN was performed using Grid Search and early stopping. Model performance was evaluated using sensitivity, AUC, F1-score, balanced accuracy, and G-Means. The results showed that ANN with hyperparameter tuning achieved the highest sensitivity of 0.741 and an AUC of 0.769 on the testing data. Although Logistic Regression produced a relatively higher F1-score, ANN was more effective in detecting potential default cases. All models also demonstrated good generalization ability without significant overfitting. Therefore, ANN with hyperparameter tuning is more suitable as an early screening model for detecting credit default risk in imbalanced data.

Keywords: Artificial Neural Network, Class Imbalance, Credit Default, Hyperparameter Tuning, Logistic Regression

Abstrak

Prediksi gagal bayar kredit merupakan bagian penting dalam manajemen risiko kredit, terutama pada data tidak seimbang yang didominasi kelas tidak gagal bayar. Penelitian ini bertujuan membandingkan performa regresi logistik, Artificial Neural Network (ANN), dan ANN dengan hyperparameter tuning dalam memprediksi gagal bayar kredit pada data tidak seimbang menggunakan sensitivity dan Area Under Curve (AUC) sebagai ukuran utama performa. Data yang digunakan adalah Default of Credit Card Clients Dataset dari UCI Machine Learning Repository sebanyak 30,000 observasi. Prapemrosesan data meliputi One-Hot Encoding, Min-Max Scaling, dan penanganan ketidakseimbangan kelas menggunakan Synthetic Minority Oversampling Technique (SMOTE) pada data latih. Hyperparameter tuning pada ANN dilakukan menggunakan Grid Search dan early stopping. Evaluasi model dilakukan menggunakan sensitivity, AUC, F1-score, balanced accuracy, dan G-Means. Hasil penelitian menunjukkan bahwa ANN dengan hyperparameter tuning menghasilkan sensitivity tertinggi sebesar 0,741 dan AUC sebesar 0,769 pada data uji. Meskipun regresi logistik menghasilkan F1-score yang relatif lebih tinggi, model ANN lebih efektif dalam mendeteksi nasabah berisiko gagal bayar. Seluruh model juga menunjukkan kemampuan generalisasi yang baik tanpa indikasi overfitting yang signifikan. Dengan demikian, ANN dengan hyperparameter tuning lebih sesuai digunakan sebagai model screening awal dalam deteksi risiko gagal bayar pada data tidak seimbang.

Kata Kunci: Artificial Neural Network, Class Imbalance, Credit Default, Hyperparameter Tuning, Logistic Regression

1. Pendahuluan

Risiko kredit merupakan salah satu ancaman utama bagi lembaga keuangan karena kegagalan debitur dalam memenuhi kewajiban pembayaran dapat menimbulkan kerugian finansial yang signifikan. Oleh karena itu, kemampuan memprediksi kemungkinan gagal bayar (*credit default*) secara akurat menjadi bagian penting dalam manajemen risiko kredit modern. Sistem *credit scoring* bertujuan mengklasifikasikan debitur ke dalam kategori baik (non-*default*) atau buruk (*default*) sehingga lembaga keuangan dapat mengambil keputusan kredit secara lebih tepat (Lessmann *et al.,* 2015).

Salah satu tantangan utama dalam prediksi gagal bayar adalah ketidakseimbangan kelas (*class imbalance*), yaitu kondisi ketika jumlah debitur gagal bayar jauh lebih sedikit dibandingkan debitur yang memenuhi kewajibannya. Kondisi ini menyebabkan model klasifikasi cenderung bias terhadap kelas mayoritas sehingga akurasi menjadi kurang representative

(Lessmann *et al.,* 2015). Akibatnya, kemampuan model mendeteksi debitur gagal bayar menjadi terbatas, padahal kesalahan mengklasifikasikan debitur gagal bayar sebagai non-*default* dapat menimbulkan kerugian yang lebih besar dibandingkan kesalahan sebaliknya.

Regresi logistik merupakan metode statistik klasik yang banyak digunakan dalam industri perbankan karena bersifat interpretatif dan stabil (Dumitrescu *et al.,* 2022). Namun, metode ini memiliki keterbatasan dalam menangkap hubungan nonlinier pada data kredit yang kompleks. Oleh karena itu, algoritma *machine learning* seperti Artificial Neural Network (ANN) banyak digunakan karena mampu memodelkan pola nonlinier dan hubungan kompleks antarfitur(El Khair Ghoujdam *et al.,* 2024; Louzada *et al.,* 2016). Meskipun demikian, performa ANN sangat dipengaruhi oleh pemilihan hyperparameter sehingga berpotensi mengalami *overfitting*.

Penelitian terdahulu menunjukkan bahwa konfigurasi *hyperparameter* memengaruhi performa model deep learning dalam prediksi gagal bayar kredit (Wahab *et al.,* 2024). Namun, studi yang secara khusus membandingkan regresi logistik, ANN, dan ANN dengan *hyperparameter* tuning pada data tidak seimbang menggunakan sensitivity dan AUC sebagai metrik utama evaluasi masih perlu dikaji lebih lanjut.

Berdasarkan uraian tersebut, penelitian ini bertujuan membandingkan performa regresi logistik, ANN, dan ANN dengan *hyperparameter* tuning dalam memprediksi gagal bayar kredit pada data tidak seimbang, serta mengevaluasi kemampuan masing-masing model dalam mendeteksi kelas gagal bayar menggunakan sensitivity dan AUC sebagai ukuran utama performa.

2. Metodologi

2.1 Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang bersumber dari UCI Machine Learning Repository. Dataset yang digunakan adalah *Default of Credit Card Clients Dataset* yang berkaitan dengan kasus gagal bayar nasabah kartu kredit di Taiwan. Tujuan klasifikasi pada dataset ini adalah memprediksi status gagal bayar nasabah kartu kredit (*default payment*) berdasarkan karakteristik demografi, riwayat pembayaran, tagihan, dan pembayaran sebelumnya.

Dataset yang digunakan terdiri atas 30.000 unit observasi dengan 24 peubah. Dari 24 peubah tersebut, 23 peubah merupakan peubah independen (fitur/peubah X) dan 1 peubah merupakan peubah dependen (target/peubah Y). Peubah target yang digunakan adalah *default.payment.next.month*, yaitu status gagal bayar nasabah pada bulan berikutnya. Nilai 1 menunjukkan nasabah mengalami gagal bayar, sedangkan nilai 0 menunjukkan nasabah tidak mengalami gagal bayar.

Tabel 1: Peubah yang digunakan.

Peubah	Nama Peubah	Deskripsi Peubah	Tipe Data
Y	*default.payment.next.month*	Status gagal bayar bulan berikutnya	Kategorik
X1	*LIMIT_BAL*	Jumlah kredit yang diberikan	Numerik
X2	*SEX*	Jenis Kelamin	Kategorik
X3	*EDUCATION*	Tingkat pendidikan	Kategorik
X4	*MARRIAGE*	Status pernikahan	Kategorik
X5	*AGE*	Usia nasabah	Numerik
X6 – X11	*PAY_6 – PAY_0*	Status pembayaran April – September 2005	Numerik
X12-X17	BILL_AMT6 – BILL_AMT1	Jumlah tagihan April – September 2005	Numerik
X18-X23	PAY_AMT6 – PAY_AMT1	Jumlah pembayaran April –September 2005	Numerik

2.2 Metode Penelitian

Prosedur analisis penelitian sebagai berikut:

1. Prapemrosesan dan Eksplorasi Data
   Prosedur Prapemrosesan data dibuat seragam pada regresi logistik dan Artificial Neural Network (ANN) untuk menjaga konsistensi dan validitas perbandingan performa model
   - Tahapan awal meliputi pemeriksaan struktur dataset, identifikasi tipe data, penghapusan peubah yang tidak relevan, pemeriksaan *missing value*, pemeriksaan multikolinearitas, mendeteksi *outlier* serta validasi kategori pada peubah kategorik (Côté *et al.,* 2024; Koshti *et al.,* 2025). Jika ditemukan data hilang, penanganan dilakukan menggunakan *mean imputation* untuk peubah numerik dan modus untuk peubah kategorik
   - Tahapan selanjutnya adalah pembagian dataset. Dataset kemudian dibagi menggunakan *stratified random sampling* dengan proporsi 80% data latih dan 20% data uji untuk mempertahankan distribusi kelas pada kedua kelompok data. Peubah kategorik dikonversi menggunakan *One-Hot Encoding*, sedangkan peubah numerik ditransformasikan menggunakan *Min-Max Scaling* ke rentang 0-1 . Selanjutnya penyeimbangan kelas pada data latih dilakukan menggunakan *Synthetic Minority Oversampling Technique* (SMOTE) .
2. Pemodelan Regresi Logistik
   Pemodelan regresi logistik dilakukan menggunakan data latih yang telah

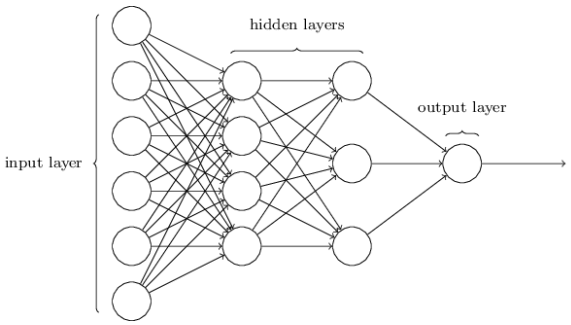
melalui tahap prapemrosesan untuk memodelkan terjadinya gagal pembayaran. Regresi logistik digunakan sebagai metode klasifikasi biner yang menghasilkan probabilitas prediksi pada rentang 0 hingga 1 sebagaimana ditunjukkan oleh Persamaan (1) .

$$y = \frac{1}{1 + e^{-(b_0 + b_1 x)}} \quad (1)$$

dengan y adalah nilai probabilitas *output* yang diprediksi, b_0 adalah konstanta atau *bias term,* dan b_1 adalah koefisien untuk peubah *input* x. Probabilitas hasil prediksi selanjutnya diklasifikasikan ke dalam kelas gagal bayar dan berhasil bayar menggunakan nilai *cutoff* sebesar 0.5 .

3. Pemodelan Artificial Neural Network

   Artificial Neural Network (ANN) merupakan model komputasi yang terinspirasi oleh cara kerja jaringan saraf biologis. Model ini memproses informasi melalui serangkaian lapisan yang saling terhubung. Secara umum, arsitektur ANN terdiri atas tiga komponen utama, yaitu *input layer*, satu atau lebih *hidden layer*, dan *output layer*, yang ditunjukkan pada Gambar 1.



Gambar 1: Ilustrasi arsitektur ANN dengan dua hidden layer (Nielsen, 2015)

   Informasi mengalir secara searah dari lapisan *input* menuju lapisan *output* melalui *hidden layer* dalam arsitektur yang disebut *feedforward neural network*. Setiap neuron menerima *input* berupa kombinasi linear dari neuron-neuron pada lapisan sebelumnya, yang kemudian ditransformasi oleh fungsi aktivasi untuk menghasilkan *output* yang bersifat nonlinier (Nielsen, 2015b). Proses pembelajaran ANN dilakukan melalui algoritma *backpropagation*, yang secara iteratif menyesuaikan bobot (*weight*) dan bias pada setiap koneksi antarneuron berdasarkan gradien fungsi *loss* terhadap parameter model.

4. Pemodelan Artificial Neural Network dengan *Hyperparameter Tuning*

   Hyperparameter tuning adalah proses optimasi *hyperparameter* model yang ditentukan sebelum proses pelatihan untuk memperoleh performa model yang lebih optimal. Ruang pencarian *hyperparameter* yang digunakan pada proses optimasi tertera pada Tabel 2.

Tabel 2: *Hyperparameter* yang akan dioptimasi.

Hyperparameter	Nilai yang diuji	Sumber
Hidden layer	[1, 2]	(Borisov *et al.,* 2024)
Neuron hidden layer	[32, 64, 128]	(Borisov *et al.,* 2024)
Learning rate	[0.001; 0.01; 0.0001]	(AbdelAziz *et al.,* 2025)
Dropout rate	[0.0; 0.2; 0.5]	(Salehin & Kang, 2023)

Metode yang digunakan adalah *Grid Search*, yaitu pencarian sistematis dengan mengevaluasi seluruh kombinasi *hyperparameter* pada ruang pencarian yang telah ditentukan (Belete & Huchaiah, 2021). Setiap kombinasi dilatih dengan menggunakan maksimum 100 *epoch*, *batch size* sebesar 32, *optimizer* Adam, fungsi aktivasi ReLU pada *hidden layer,* dan fungsi aktivasi Sigmoid pada *output layer.* Selain itu, diterapkan mekanisme *early stopping* untuk mencegah terjadinya *overfitting* (Hussein & Shareef, 2024).

5. Evaluasi performa model dilakukan pada kondisi *imbalanced* data karena penggunaan akurasi saja dapat menghasilkan evaluasi yang bias terhadap kelas mayoritas (Tharwat, 2018). Oleh karena itu, penelitian ini menggunakan metrik Sensitivity (Recall), F1-Score, Area Under Curve (AUC), Balanced Accuracy, dan G-Means. Sensitivity digunakan untuk mengukur kemampuan model mendeteksi kelas positif, F1-Score untuk mengukur keseimbangan antara precision dan recall, sedangkan AUC digunakan untuk mengukur kemampuan model membedakan kelas positif dan negatif pada berbagai nilai *threshold*. Selain itu, Balanced Accuracy dihitung sebagai rata-rata sensitivity dan spesificity:

$$Balanced\ Accuracy = \frac{Sensitivity + Specificity}{2} \quad (2)$$

Sementara itu, G-Means digunakan untuk mengukur keseimbangan performa klasifikasi antara kelas mayoritas dan minoritas dengan rumus:

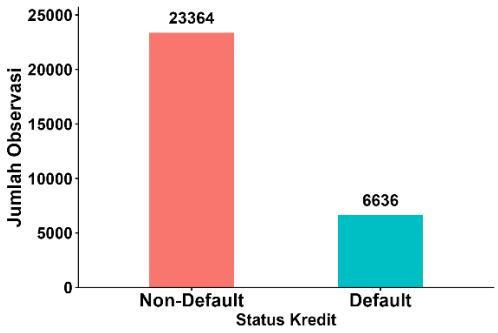
$$G - Means = \sqrt{Sensitifity \times Specificity} \quad (3)$$

Evaluasi model dilakukan pada data latih dan uji untuk menilai kemampuan generalisasi serta mendeteksi *overfitting* melalui selisih performa (*gap*) antara data latih dan uji (Goodfellow *et al.,* 2016). Model dikategorikan mengalami *overfitting* apabila nilai gap melebihi 0,05, sedangkan model dengan selisih performa yang relatif seimbang dianggap memiliki kemampuan generalisasi yang baik.

3. Hasil dan Pembahasan

3.1 Prapemrosesan dan Eksplorasi Data

Hasil pemeriksaan awal menunjukkan bahwa dataset tidak memiliki data hilang dan seluruh kategori pada peubah kategorik telah divalidasi sesuai metadata dataset. Distribusi peubah target menunjukkan bahwa data bersifat tidak seimbang, di mana proporsi kelas berhasil bayar lebih dominan dibandingkan kelas gagal bayar yang ditunjukkan pada Gambar 2. Kondisi ini berpotensi menyebabkan model lebih cenderung memprediksi kelas mayoritas sehingga diperlukan penanganan ketidakseimbangan kelas pada tahap prapemrosesan yang dilakukan menggunakan SMOTE hanya pada data latih.



Gambar 2: Distribusi kelas target

Identifikasi distribusi data menunjukkan adanya *skewness* positif dan nilai ekstrem pada beberapa peubah numerik finansial. *Outlier* tetap dipertahankan karena masih merepresentasikan kondisi finansial yang valid, tetapi pengaruhnya dikurangi melalui transformasi distribusi. Transformasi akar kuadrat diterapkan pada peubah dengan *skewness* moderat, sedangkan transformasi akar kubik digunakan pada peubah dengan *skewness* tinggi.

3.2 Hasil Pemodelan

Model regresi logistik dibangun menggunakan data latih yang telah melalui tahap prapemrosesan, serta menggunakan seluruh 26 peubah prediktor hasil *encoding* dengan nilai cutoff sebesar 0,5 pada proses klasifikasi. Model Artificial Neural Network (ANN) tanpa *hyperparameter tuning* dibangun menggunakan arsitektur sederhana sebagai *baseline* sebelum proses optimasi model. Arsitektur ANN terdiri atas satu input layer dengan 26 neuron yang merepresentasikan seluruh peubah prediktor hasil prapemrosesan, dua *hidden layer* dengan masing-masing 16 dan 8 neuron, serta satu *output layer* dengan satu neuron. Jumlah neuron pada input layer disesuaikan dengan jumlah fitur masukan yang digunakan dalam pemodelan. Penggunaan dua hidden layer didasarkan pada pendapat Nielsen (2015) yang menyatakan bahwa jaringan dengan dua atau lebih hidden layer mampu membangun hierarki representasi data secara bertingkat, mulai dari pola sederhana hingga pola yang lebih kompleks.

Pada output layer digunakan fungsi aktivasi sigmoid untuk menghasilkan probabilitas prediksi pada rentang [0,1]. Model ANN dilatih menggunakan fungsi loss binary crossentropy, optimizer Adam dengan *learning rate* 0,001, *batch size* 32, dan 30 *epoch* pelatihan.

Proses *hyperparameter tuning* dilakukan menggunakan metode Grid Search dengan mengevaluasi seluruh kombinasi dari ruang pencarian yang telah ditetapkan, menghasilkan total 54 kombinasi yang diuji secara sistematis. Setiap kombinasi dilatih dengan maksimum 100 epoch, *batch size* 32, optimizer Adam, fungsi aktivasi ReLU pada setiap hidden layer, dan fungsi aktivasi Sigmoid pada output layer. Mekanisme *early stopping* dengan *patience* 10 diterapkan untuk menghentikan pelatihan apabila *validation loss* tidak membaik selama 10 *epoch* berturut-turut.

Tabel 3: Konfigurasi hyperparameter terbaik hasil Grid Search

Hyperparameter	Nilai Terpilih
Hidden layer	2
Neuron hidden layer	32 (layer 1) & 16 (layer 2)
Learning rate	0.01
Dropout rate	0.0
Epoch	14 (*early stopping*)

Berdasarkan hasil tuning dengan Grid Search, konfigurasi *hyperparameter* terbaik yang terpilih disajikan pada Tabel 3. Proses pelatihan berhenti pada *epoch* ke-14 melalui mekanisme *early stopping*, yang mengindikasikan bahwa model mencapai konvergensi lebih awal dari batas maksimum yang ditetapkan.

3.3 Perbandingan Performa Model

Perbandingan performa ketiga model dilakukan untuk mengevaluasi kemampuan masing-masing model dalam mendeteksi gagal bayar kredit pada data tidak seimbang. Evaluasi dilakukan menggunakan metrik sensitivity, F1-score, balanced accuracy, G-Means, dan AUC karena akurasi tunggal kurang representatif pada kondisi class imbalance. Hasil evaluasi pada data latih dan uji disajikan pada Tabel 4.

Berdasarkan Tabel 4, model ANN secara umum menunjukkan performa yang lebih baik dibandingkan regresi logistik pada sebagian besar metrik evaluasi, baik pada data latih maupun data uji. Pada data uji, ANN dengan *hyperparameter tuning* menghasilkan nilai sensitivity tertinggi sebesar 0.741, diikuti ANN tanpa *hyperparameter tuning* sebesar 0.664, dan regresi logistik sebesar 0.598. Hasil tersebut menunjukkan bahwa model ANN memiliki kemampuan yang lebih baik dalam mendeteksi nasabah yang berpotensi mengalami gagal bayar. Pola serupa juga terlihat pada nilai AUC, dengan ANN dengan *hyperparameter tuning* memperoleh nilai tertinggi sebesar 0.769, diikuti ANN tanpa *hyperparameter tuning* sebesar 0.766, dan regresi logistik sebesar 0.742.

Tabel 4: Metrik evaluasi regresi logistik pada data latih dan uji.

Metrik Evaluasi	Data	Regresi Logistik	ANN*	ANN**
Sensitivity	Latih	0.589	0.695	0.756
	Uji	0.598	0.664	0.741
F1-score	Latih	0.646	0.698	0.701
	Uji	0.517	0.511	0.501
Balanced accuracy	Latih	0.694	0.722	0.708
	Uji	0.697	0.699	0.698
G-Means	Latih	0.686	0.721	0.706
	Uji	0.690	0.698	0.696
AUC	Latih	0.755	0.798	0.787
	Uji	0.742	0.766	0.769

* : tanpa *hyperparameter tuning*
** : dengan *hyperparameter tuning*

Meskipun model ANN menunjukkan sensitivity dan AUC yang lebih tinggi, ketiga model mengalami penurunan nilai F1-score pada data uji dibandingkan data latih. Regresi logistik menghasilkan nilai F1-score tertinggi pada data uji sebesar 0.517, sedangkan ANN dengan *hyperparameter tuning* memperoleh nilai sebesar 0.501. Penurunan tersebut mengindikasikan adanya *trade-off* antara sensitivity dan ketepatan prediksi positif pada data tidak seimbang. Sensitivity yang lebih tinggi pada model ANN menunjukkan bahwa model lebih agresif dalam mengidentifikasi nasabah berisiko gagal bayar. Namun, peningkatan kemampuan deteksi tersebut mengindikasikan kemungkinan peningkatan *false positive*, yaitu nasabah yang sebenarnya layak menerima kredit tetapi diprediksi sebagai gagal bayar.

Dalam konteks credit scoring, kedua jenis kesalahan klasifikasi, yaitu *false positive* dan *false negative*, sama-sama memiliki konsekuensi bisnis bagi lembaga keuangan. *False positive* dapat menyebabkan hilangnya peluang pendapatan karena nasabah yang sebenarnya layak kredit diprediksi mengalami gagal bayar, sedangkan *false negative* berpotensi menimbulkan kerugian akibat pemberian kredit kepada nasabah bermasalah. Lessmann et al. (2015) menyatakan bahwa *false negative* umumnya memiliki dampak finansial yang lebih besar dibandingkan *false positive* karena berkaitan langsung dengan risiko kredit macet. Oleh karena itu, model dengan sensitivity yang lebih tinggi tetap relevan dalam konteks deteksi risiko gagal bayar meskipun nilai F1-score cenderung lebih rendah. Dengan demikian, ANN dengan *hyperparameter tuning* lebih sesuai digunakan sebagai alat *screening* awal untuk mendeteksi potensi risiko kredit dibandingkan sebagai satu-satunya dasar dalam pengambilan keputusan kredit.

Selain itu, selisih performa antara data latih dan data uji pada metrik sensitivity, balanced accuracy, G-Means, dan AUC untuk seluruh model relatif kecil, yaitu berada di bawah 0,05. Hasil tersebut menunjukkan bahwa ketiga model memiliki kemampuan generalisasi yang cukup baik dan tidak

mengalami *overfitting* yang signifikan (Thölke *et al.,* 2023). Penurunan nilai F1-score pada data uji diduga dipengaruhi oleh perbedaan distribusi kelas antara data latih dan data uji akibat penerapan SMOTE yang hanya dilakukan pada data latih (Holzmann & Klar, 2024). Meskipun demikian, pola performa pada data uji masih konsisten dengan data latih, terutama pada metrik sensitivity dan AUC. Perbedaan nilai balanced accuracy dan G-Means antar model juga relatif kecil sehingga kemampuan ketiga model dalam menyeimbangkan klasifikasi kelas mayoritas dan minoritas cenderung serupa. Namun, ANN dengan *hyperparameter tuning* tetap unggul pada sensitivity dan AUC yang merupakan metrik utama dalam deteksi risiko gagal bayar (Kim & Yang, 2024). Hasil tersebut menunjukkan bahwa ANN dengan *hyperparameter tuning* memiliki kemampuan deteksi risiko gagal bayar yang lebih baik tanpa penurunan kemampuan generalisasi yang signifikan.

4. Simpulan dan Saran

Berdasarkan hasil penelitian, model ANN dengan *hyperparameter tuning* menunjukkan performa terbaik dalam mendeteksi gagal bayar pada data tidak seimbang dibandingkan regresi logistik dan ANN tanpa *tuning*. Hal tersebut ditunjukkan oleh nilai sensitivity tertinggi sebesar 0,741 dan AUC sebesar 0,769 pada data uji, sehingga model memiliki kemampuan yang lebih baik dalam mengidentifikasi nasabah berisiko gagal bayar. Meskipun regresi logistik menghasilkan nilai F1-score yang relatif lebih tinggi, model ANN lebih efektif dalam mendeteksi kelas gagal bayar yang menjadi fokus utama penelitian. Hasil penelitian juga menunjukkan adanya *trade-off* antara peningkatan sensitivity dan penurunan ketepatan prediksi positif pada model ANN. Selain itu, seluruh model memiliki kemampuan generalisasi yang baik tanpa indikasi *overfitting* yang signifikan. Dengan demikian, ANN dengan *hyperparameter tuning* lebih sesuai digunakan sebagai model *screening* awal dalam deteksi gagal bayar pada data tidak seimbang.

Penelitian selanjutnya disarankan untuk mengeksplorasi metode penanganan data tidak seimbang lainnya, seperti *undersampling*, *hybrid sampling*, maupun *cost-sensitive learning*, serta membandingkan performa ANN dengan algoritma lain seperti Random Forest dan XGBoost. Penelitian lanjutan juga dapat mempertimbangkan analisis biaya kesalahan klasifikasi guna mengevaluasi dampak *false positive* dan *false negative* terhadap keputusan bisnis dalam *credit scoring*.

Daftar Pustaka

AbdelAziz, N. M., Bekheet, M., Salah, A., El-Saber, N., & AbdelMoneim, W. T. (2025). A comprehensive evaluation of machine learning and deep learning models for churn prediction. *Information (Switzerland)*, *16*(7). https://doi.org/10.3390/info16070537

Belete, D. M., & Huchaiah, M. D. (2021). Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results. *International Journal of Computers and Applications*, *44*(9), 875–886. https://doi.org/10.1080/1206212X.2021.1974663

Borisov, V., Leemann, T., Sebler, K., Haug, J., Pawelczyk, M., & Kasneci, G. (2024). Deep Neural Networks and Tabular Data: A Survey. *IEEE Transactions on Neural Networks and Learning Systems*, *35*(6), 7499–7519. https://doi.org/10.1109/TNNLS.2022.3229161

Côté, P. O., Nikanjam, A., Ahmed, N., Humeniuk, D., & Khomh, F. (2024). Data cleaning and machine learning: a systematic literature review. *Automated Software Engineering*, *31*(2). https://doi.org/10.1007/s10515-024-00453-w

Dumitrescu, E., Hué, S., Hurlin, C., & Tokpavi, S. (2022). Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. *European Journal of Operational Research*, *297*(3), 1178–1192. https://doi.org/10.1016/j.ejor.2021.06.053

El Khair Ghoujdam, M., Chaabita, R., Elkhalfi, O., Zehraoui, K., Elalaoui, H., & Idamia, S. (2024). Consumer credit risk analysis through artificial intelligence: a comparative study between the classical approach of logistic regression and advanced machine learning techniques. *Cogent Economics and Finance*, *12*(1). https://doi.org/10.1080/23322039.2024.2414926

Goodfellow, I., Bengio, Y., & Courville, A. (n.d.). *Deep Learning*.

Holzmann, H., & Klar, B. (2024). Robust performance metrics for imbalanced classification problems. *ArXiv Preprint ArXiv:2404.07661*. https://doi.org/https://doi.org/10.48550/arXiv.2404.07661

Hussein, B. M., & Shareef, S. M. (2024). An Empirical Study on the Correlation between Early Stopping Patience and Epochs in Deep Learning. *ITM Web of Conferences*, *64*, 01003. https://doi.org/10.1051/itmconf/20246401003

Kim, S., & Yang, S. Y. (2024). Accuracy improvement in financial sanction screening: is natural language processing the solution? *Frontiers in Artificial Intelligence*, *7*. https://doi.org/10.3389/frai.2024.1374323

Koshti, R. M., Molia, S. J., & Varia, D. J. (2025). Improving Credit Score Classification Using Predictive Analysis and Machine Learning

Techniques. *International Journal on Science and Technology*, *16(2)*. https://doi.org/https://doi.org/10.71097/IJSAT.v16.i2.4759

Lessmann, S., Seow, H.-V., Baesens, B., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: A ten-year update. *European Journal of Operational Research*, *247(1)*, 124–136. https://doi.org/https://doi.org/10.1016/j.ejor.2015.05.030

Louzada, F., Ara, A., & Fernandes, G. B. (2016). Classification methods applied to credit scoring: A systematic review and overall comparison. *Surveys in Operations Research and Management Science*, *21(2)*, 117–134. https://doi.org/https://doi.org/10.1016/j.sorms.2016.10.001

Nielsen, M. (2015a). *Neural Networks and Deep Learning*. Determination Press. http://neuralnetworksanddeeplearning.com

Salehin, I., & Kang, D. K. (2023). A review on dropout regularization approaches for deep neural networks within the scholarly domain. In *Electronics (Switzerland)* (Vol. 12, Number 14). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/electronics12143106

Tharwat, A. (2018). Classification assessment methods. *Applied Computing and Informatics*, *17*(1), 168–192. https://doi.org/10.1016/j.aci.2018.08.003

Thölke, P., Mantilla-Ramos, Y. J., Abdelhedi, H., Maschke, C., Dehgan, A., Harel, Y., Kemtur, A., Mekki Berrada, L., Sahraoui, M., Young, T., Bellemare Pépin, A., El Khantour, C., Landry, M., Pascarella, A., Hadid, V., Combrisson, E., O'Byrne, J., & Jerbi, K. (2023). Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data. *NeuroImage*, *277*. https://doi.org/10.1016/j.neuroimage.2023.120253