

ANALISIS KOMPARASI ALGORITMA RANDOM FOREST, XGBOOST, DAN MULTILAYER PERCEPTRON (MLP) DALAM PREDIKSI RISIKO GAGAL BAYAR KREDIT

D. Febry Wulangsih

Universitas Pembangunan Nasional “Veteran” Jawa Timur
22082010188@student.upnjatim.ac.id

Keysya Alifia Zabina

Universitas Pembangunan Nasional “Veteran” Jawa Timur
22082010187@student.upnjatim.ac.id

Abstract

Credit risk prediction is a critical task for financial institutions in identifying customers at risk of default. This study compares the performance of three machine learning and deep learning algorithms as in Multilayer Perceptron (MLP), Random Forest, and XGBoost in predicting credit card defaults using the “Default of Credit Card Clients” dataset from the UCI Machine Learning Repository. The dataset consists of 30,000 records with 23 features covering demographic information, payment history, bills, and payment amounts over a six-month period. Class imbalance was addressed using the Synthetic Minority Oversampling Technique (SMOTE), which was applied only to the training data. Model performance was evaluated using accuracy, precision, recall, F1-score, and AUC-ROC metrics. The experimental results show that Random Forest achieved the best overall performance with an F1-score of 0.5275 and an AUC-ROC of 0.768, outperforming MLP (F1-score 0.5182, AUC-ROC 0.7610) and XGBoost (F1-score 0.5080, AUC-ROC 0.7616%). These findings indicate that ensemble-based methods remain competitive compared to deep learning approaches for tabular credit data, and provide valuable insights for financial institutions in implementing data-driven risk management.

Keywords: Credit Risk, Random Forest, XGBoost, Multilayer Perceptron, Class Imbalance

Abstrak

Prediksi risiko kredit merupakan tugas kritis bagi lembaga keuangan dalam mengidentifikasi nasabah yang berpotensi gagal bayar. Penelitian ini membandingkan performa tiga algoritma machine learning dan deep learning, yaitu Multilayer Perceptron (MLP), Random Forest, dan XGBoost, dalam memprediksi gagal bayar kartu kredit menggunakan dataset Default of Credit Card Clients dari UCI Machine Learning Repository. Dataset terdiri dari 30.000 data dengan 23 fitur yang mencakup informasi demografis, riwayat pembayaran, tagihan, dan jumlah pembayaran selama enam bulan. Ketidakseimbangan kelas ditangani menggunakan Synthetic Minority Oversampling Technique (SMOTE) yang hanya diterapkan pada data latih. Performa model dievaluasi menggunakan metrik akurasi, presisi, recall, F1-score, dan AUC-ROC. Hasil eksperimen menunjukkan bahwa Random Forest mencapai performa terbaik secara

keseluruhan dengan F1-score 0.5275 dan AUC-ROC 0.768, mengungguli MLP (F1-score 0.5182, AUC-ROC 0.7610) dan XGBoost (F1-score 0.5080, AUC-ROC 0.7616%). Temuan ini mengindikasikan bahwa metode berbasis ensemble tetap kompetitif dibandingkan pendekatan deep learning untuk data kredit bertipe tabular, serta memberikan wawasan berharga bagi lembaga keuangan dalam mengimplementasikan manajemen risiko berbasis data.

Kata Kunci : Risiko Kredit, Random Forest, XGBoost, Multilayer Perceptron, Class Imbalance

PENDAHULUAN

Sektor perbankan dan lembaga keuangan menghadapi tantangan signifikan dalam mengelola risiko kredit, khususnya dalam mengidentifikasi nasabah yang berpotensi mengalami gagal bayar. Gagal bayar kredit tidak hanya berdampak pada kerugian finansial institusi, tetapi juga dapat memicu ketidakstabilan sistem keuangan secara lebih luas. Oleh karena itu, kemampuan untuk memprediksi risiko gagal bayar secara akurat dan proaktif menjadi kebutuhan mendasar bagi lembaga keuangan modern. Perkembangan teknologi kecerdasan buatan, khususnya machine learning dan deep learning, telah membuka peluang baru dalam pengembangan sistem prediksi risiko kredit yang lebih canggih dan akurat. Berbagai algoritma telah diusulkan dan diujicobakan untuk menangani permasalahan klasifikasi risiko kredit, mulai dari pendekatan statistik tradisional hingga metode ensemble dan jaringan syaraf tiruan. Setiap pendekatan memiliki karakteristik, kelebihan, dan keterbatasan tersendiri, sehingga pemilihan algoritma yang tepat untuk konteks data tertentu menjadi pertimbangan penting dalam pengembangan sistem prediksi.

Multilayer Perceptron (MLP) merupakan arsitektur dasar dari deep learning yang terdiri dari lapisan-lapisan neuron yang saling terhubung secara penuh (fully connected). MLP mampu memodelkan hubungan non-linear yang kompleks antar fitur melalui fungsi aktivasi pada setiap lapisan tersembunyi. Beberapa penelitian telah menunjukkan potensi MLP dalam penilaian risiko kredit. Penelitian oleh Tetteh et al. (2024) membuktikan bahwa MLP efektif dalam mengklasifikasikan nasabah berisiko dan tidak berisiko pada data pinjaman, dengan kemampuan khusus dalam menangani ketidakseimbangan label. Sementara itu, penelitian oleh Sun et al. (2025) mengusulkan model penilaian risiko kredit berbasis MLP hybrid clustering yang terbukti lebih unggul dibandingkan metode regresi tradisional dalam menangani dataset multidimensi yang kompleks.

Di sisi lain, Random Forest sebagai algoritma ensemble berbasis pohon keputusan telah lama dikenal sebagai salah satu metode paling handal untuk data tabular. Random Forest membangun sejumlah pohon keputusan secara paralel dan menggabungkan hasilnya melalui mekanisme voting mayoritas, sehingga menghasilkan prediksi yang lebih stabil dan robust terhadap noise. Sedangkan XGBoost merupakan algoritma gradient boosting yang membangun pohon secara sekuensial, dimana setiap

pohon baru belajar dari kesalahan pohon sebelumnya. XGBoost dikenal memiliki performa tinggi dan efisiensi komputasi yang baik, menjadikannya pilihan populer dalam berbagai kompetisi machine learning dan aplikasi industri.

Salah satu tantangan utama dalam prediksi risiko kredit adalah permasalahan ketidakseimbangan kelas atau *class imbalance*, dimana jumlah nasabah yang tidak gagal bayar jauh lebih besar dibandingkan yang gagal bayar. Kondisi ini dapat menyebabkan model cenderung bias terhadap kelas mayoritas sehingga menghasilkan tingkat deteksi yang rendah untuk kelas minoritas yang justru paling penting untuk diidentifikasi. Synthetic Minority Oversampling Technique (SMOTE) merupakan salah satu metode penanganan *class imbalance* yang paling banyak digunakan, bekerja dengan cara mensintesis data baru untuk kelas minoritas melalui interpolasi antara sampel yang ada.

Meskipun ketiga algoritma tersebut telah banyak diteliti secara individual, studi komparatif yang secara sistematis membandingkan MLP, Random Forest, dan XGBoost pada dataset risiko kredit dengan penanganan *class imbalance* yang terstandarisasi masih terbatas. Penelitian oleh Hassan et al. (2023) menunjukkan bahwa pemilihan optimizer dan arsitektur yang tepat sangat mempengaruhi performa MLP, sementara penelitian oleh Ensemble-Based Machine Learning mengindikasikan bahwa kombinasi metode ensemble dengan SMOTE dapat secara signifikan meningkatkan kemampuan deteksi kelas minoritas. Namun, belum ada konsensus yang jelas mengenai algoritma mana yang paling optimal untuk konteks prediksi risiko kredit kartu dengan penanganan *class imbalance* yang konsisten.

Penelitian ini bertujuan untuk melakukan evaluasi komparatif secara sistematis terhadap ketiga algoritma tersebut menggunakan dataset Default of Credit Card Clients dari UCI Machine Learning Repository yang terdiri dari 30.000 data nasabah kartu kredit di Taiwan. Seluruh model dilatih dengan kondisi preprocessing yang identik, termasuk normalisasi Min-Max Scaling dan penanganan *class imbalance* menggunakan SMOTE, sehingga perbandingan performa antar algoritma dapat dilakukan secara adil dan objektif. Evaluasi dilakukan menggunakan lima metrik komprehensif yaitu akurasi, presisi, recall, F1-score, dan AUC-ROC untuk memberikan gambaran performa yang menyeluruh. Kontribusi utama penelitian ini adalah menyediakan bukti empiris yang komprehensif mengenai perbandingan performa MLP sebagai representasi pendekatan deep learning dengan Random Forest dan XGBoost sebagai representasi pendekatan ensemble machine learning dalam konteks prediksi risiko kredit. Hasil penelitian ini diharapkan dapat memberikan panduan praktis bagi para peneliti dan praktisi dalam memilih algoritma yang paling sesuai untuk implementasi sistem penilaian risiko kredit berbasis data.

METODE PENELITIAN

Penelitian ini menerapkan pendekatan kuantitatif eksperimental berbasis data mining untuk melakukan analisis komparatif klasifikasi risiko gagal bayar kredit nasabah.

Tahapan metodologi secara sistematis diawali dengan akuisisi data sekunder yang bersumber dari dataset Default of Credit Card Clients dari UCI Machine Learning Repository. Dataset ini merepresentasikan rekam jejak aktivitas finansial dan profil demografis dari 30.000 nasabah kartu kredit di Taiwan. Atribut prediktor yang diekstrak mencakup 23 fitur, meliputi limit kredit, karakteristik demografis (jenis kelamin, tingkat pendidikan, status pernikahan dan usia), riwayat status pembayaran bulanan dari April hingga September, serta jumlah tagihan dan nominal pembayaran sebelumnya. Variabel target dalam penelitian ini adalah indikator biner yang menunjukkan status gagal bayar nasabah pada bulan berikutnya.

Proses pra-pemrosesan data merupakan tahapan krusial dalam eksperimen ini untuk memastikan kualitas data masukan sebelum ke dalam fase pelatihan model. Langkah awal diimplementasikan melalui pembersihan data dengan menghapus kolom yang tidak memiliki nilai informatif dan prediktif dalam proses pengenalan pola klasifikasi. Pengamatan terhadap distribusi kelas target menunjukkan adanya masalah ketidakseimbangan kelas (class imbalance) yang sangat signifikan. Kelas nasabah berstatus tidak gagal bayar (kelas 0) mendominasi secara ekstrem sebesar 77,9% (23.364 sampel), sedangkan kelas nasabah yang gagal bayar (kelas 1) hanya berjumlah 22,1% (6.636 sampel). Kondisi ketidakseimbangan ini berisiko memicu bias klasifikasi, di mana model cenderung mengabaikan kelas minoritas. Guna mengatasi permasalahan tersebut, teknik Synthetic Minority Over-sampling Technique (SMOTE) diterapkan pada data latih. Algoritma SMOTE bekerja secara sintesis dengan melakukan interpolasi di antara sampel-sampel pada kelas minoritas berdasarkan tetangga terdekatnya, sehingga proporsi jumlah sampel antara kedua kelas menjadi seimbang tanpa menyebabkan hilangnya informasi esensial dari data asli.

Selanjutnya, seluruh fitur numerik diproses menggunakan tahapan standarisasi melalui fungsi StandardScaler. Proses transformasi ini mengonversi nilai-nilai fitur agar memiliki distribusi dengan rata-rata nol dan deviasi standar satu. Standarisasi fitur ini menjadi komponen penentu keberhasilan, khususnya bagi algoritma jaringan saraf seperti Multilayer Perceptron (MLP) yang sangat sensitif terhadap variasi skala input dalam mengoptimalkan fungsi loss menggunakan gradient descent. Setelah data dipastikan bersih dan terstandarisasi, dataset kemudian dibagi menjadi data latih (training set) untuk proses pemodelan dan data uji (testing set) untuk menguji tingkat generalisasi model pada data baru yang belum pernah dilihat sebelumnya.

Fase pemodelan dilakukan dengan melatih dan membandingkan tiga arsitektur algoritma yang berbeda. Algoritma pertama adalah Random Forest, sebuah metode ensemble learning berbasis bagging yang membangun kumpulan pohon keputusan secara paralel. Setiap pohon keputusan dilatih menggunakan kombinasi sampel acak dan subset fitur yang berbeda, di mana keputusan akhir diambil melalui mekanisme majority voting. Karakteristik pohon yang independen membuat algoritma ini sangat robust terhadap masalah overfitting pada data tabular. Algoritma kedua adalah

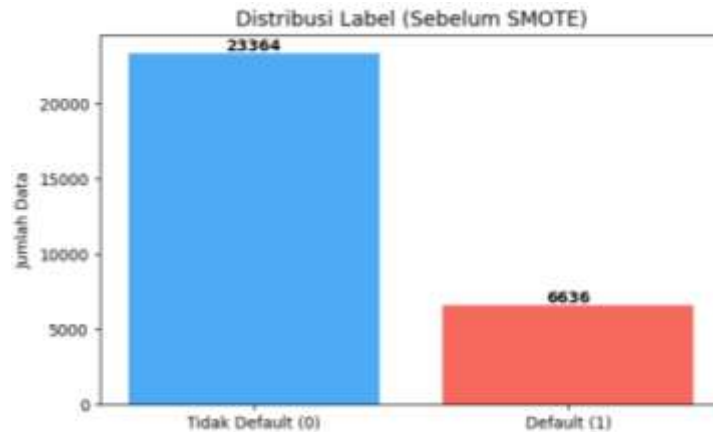
XGBoost (Extreme Gradient Boosting), sebuah pendekatan ensemble berbasis pohon yang menerapkan prinsip boosting secara sekuensial. Berbeda dengan Random Forest, pohon-pohon dalam XGBoost dibangun secara berurutan, di mana setiap pohon baru dirancang khusus untuk memperbaiki residu error atau kesalahan prediksi yang dihasilkan oleh pohon sebelumnya. Algoritma ketiga adalah Multilayer Perceptron (MLP) yang merepresentasikan pendekatan deep learning. MLP dikonfigurasi menggunakan arsitektur jaringan syaraf tiruan berlapis (fully connected layers) yang terdiri atas lapisan masukan, beberapa lapisan tersembunyi (hidden layers) yang mengintegrasikan fungsi aktivasi non-linear, dan lapisan luaran berupa fungsi klasifikasi biner untuk menentukan probabilitas risiko gagal bayar.

Tahapan akhir dari metodologi penelitian ini adalah melakukan evaluasi objektif menggunakan metrik performa klasifikasi yang komprehensif. Metrik yang dihitung mencakup Accuracy, Precision, Recall, F1-Score, dan Area Under the Curve-Receiver Operating Characteristic (AUC-ROC). Dalam konteks analisis risiko kredit, perhatian utama tidak hanya dipusatkan pada nilai akurasi global, melainkan pada metrik Recall dan F1-Score. Recall menjadi fokus krusial karena mengukur efektivitas model dalam menangkap seluruh nasabah yang sebenarnya berpotensi gagal bayar (True Positive), sehingga dapat meminimalisir kesalahan lolosnya nasabah berisiko (False Negative) yang berdampak langsung pada kerugian finansial institusi perbankan. Sementara itu, F1-Score digunakan untuk mengukur keseimbangan performa antara Precision dan Recall, dan AUC-ROC digunakan untuk mengevaluasi kemampuan menyeluruh model dalam membedakan kedua kelas nasabah pada berbagai tingkatan ambang batas (threshold) klasifikasi. Seluruh hasil metrik ini kemudian dibandingkan untuk menentukan algoritma yang paling optimal dalam pemetaan risiko kredit.

HASIL DAN PEMBAHASAN

Eksplorasi dan Praproses Data

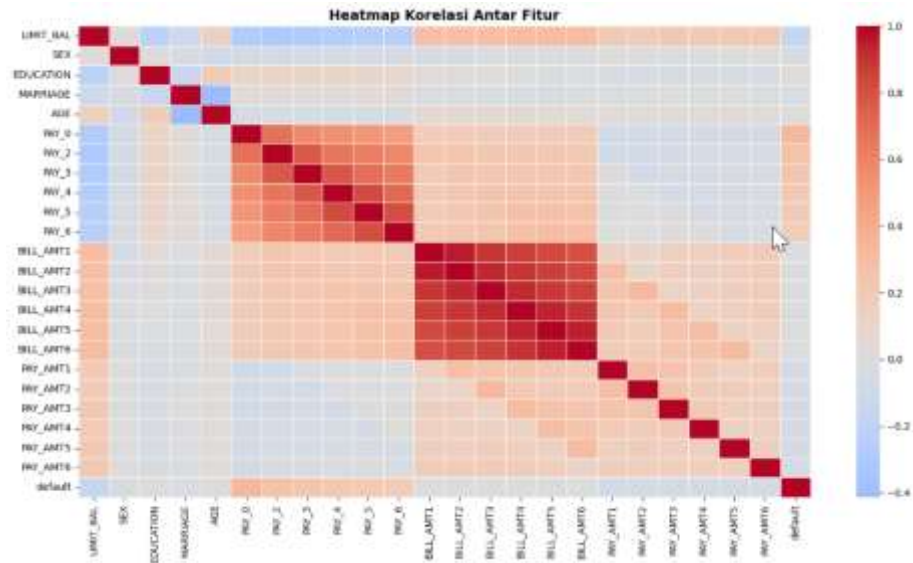
Tahap awal analisis dalam penelitian ini dilakukan melalui Exploratory Data Analysis (EDA) untuk memahami karakteristik dasar, sebaran data, serta pola hubungan antar variabel di dalam dataset Default of Credit Card Clients. Melalui visualisasi distribusi pada label target, ditemukan adanya masalah ketidakseimbangan kelas (class imbalance) yang sangat signifikan antara jumlah nasabah dengan riwayat pembayaran lancar dan nasabah yang mengalami gagal bayar.



Gambar 1. Distribusi Label

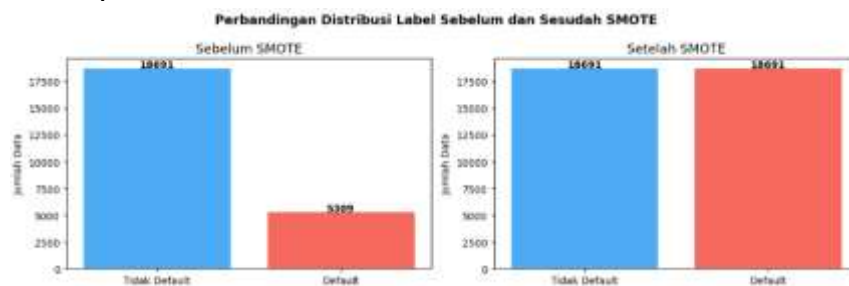
Karakteristik ketimpangan data tersebut secara visual dapat ditunjukkan melalui grafik batang pada Gambar 1. Berdasarkan pengamatan grafik tersebut, kelompok nasabah dengan kategori tidak gagal bayar (kelas 0) mendominasi dataset secara ekstrim yaitu sebesar 77,9% atau mencakup 23.364 data, sedangkan kelompok nasabah yang berisiko mengalami gagal bayar (kelas 1) hanya sebesar 22,1% atau sebanyak 6.636 data. Ketimpangan jumlah sampel yang sangat mencolok inilah yang menjadi landasan utama diterapkannya metode Synthetic Minority Over-sampling Technique (SMOTE) pada tahap pra-pemrosesan data latih. Jika kondisi data tabular yang tidak seimbang ini langsung dimasukkan ke dalam proses pelatihan tanpa penanganan resampling, model yang dibangun dipastikan akan mengalami bias tinggi ke kelas mayoritas sehingga gagal mengenali karakteristik nasabah berisiko secara akurat.

Berdasarkan visualisasi heatmap pada gambar 2, ditemukan adanya korelasi positif yang sangat kuat (mencapai koefisien 0.8 hingga 0.9) di antara sesama kelompok fitur nilai tagihan bulanan, yaitu variabel BILL_AMT1 hingga BILL_AMT6. Hubungan linier yang tinggi ini sangat logis dalam konteks finansial karena saldo tagihan seorang nasabah pada suatu bulan merupakan akumulasi yang berkaitan erat dengan bulan sebelum maupun sesudahnya. Selain itu, korelasi positif moderat juga terlihat pada kelompok fitur status keterlambatan pembayaran bulanan (PAY_0 hingga PAY_6). Sebaliknya, fitur-fitur demografis seperti jenis kelamin, tingkat pendidikan, status pernikahan, dan usia terpantau memiliki nilai korelasi yang sangat lemah mendekati angka nol terhadap variabel transaksi lainnya maupun terhadap label target.



Gambar 2. Heatmap Korelasi antar Fitur

Temuan karakteristik data awal yang menunjukkan adanya ketidakseimbangan kelas target yang ekstrim serta adanya korelasi kuat di antara sesama fitur tagihan bulanan inilah yang menjadi landasan utama mengapa tahapan pra-pemrosesan data tertentu wajib dilakukan. Penelitian ini menerapkan metode Synthetic Minority Over-sampling Technique (SMOTE) khusus pada data latih untuk menyeimbangkan proporsi jumlah sampel kelas minoritas agar model tidak mengalami bias klasifikasi. Setelah distribusi diseimbangkan menggunakan SMOTE dan seluruh fitur numerik ditransformasikan skalanya menggunakan StandardScaler, dilakukan pelatihan serta pengujian performa terhadap tiga arsitektur klasifikasi yaitu Random Forest, XGBoost, dan Multilayer Perceptron (MLP). Berikut adalah hasil pemrosesan data tidak seimbang yang sudah dilakukan proses SMOTE.



Gambar 3. Distribusi setelah SMOTE

Pemodelan

Setelah data latih melalui tahapan penyeimbangan menggunakan teknik SMOTE dan ditransformasikan skalanya menggunakan StandardScaler, proses dilanjutkan ke tahap konstruksi serta pelatihan model klasifikasi. Algoritma pertama yang diimplementasikan adalah Random Forest Classifier, sebuah metode ensemble learning berbasis pemrosesan pohon keputusan secara paralel. Pemodelan ini diinisialisasi

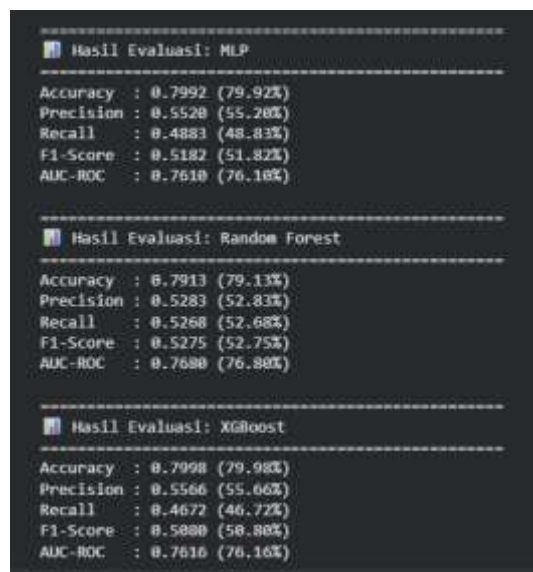
dengan memanfaatkan pustaka Scikit-Learn menggunakan fungsi `RandomForestClassifier()`. Struktur arsitektur ini bekerja melalui konsep bagging, di mana sistem membangun sejumlah pohon keputusan independen menggunakan kombinasi subset sampel acak dari data latih. Konfigurasi model ini berfokus pada parameter utama seperti jumlah pohon keputusan yang akan dibentuk (`n_estimators`) dan pembatasan kedalaman maksimum percabangan pohon (`max_depth`) guna memastikan model mampu menangkap hubungan non-linear yang kompleks tanpa mengalami risiko overfitting terhadap pola data sintesis baru. Proses penentuan keputusan akhir klasifikasi biner pada fase pengujian dijalankan otomatis melalui mekanisme pemungutan suara mayoritas (`majority voting`) dari seluruh estimasi pohon yang telah dibangun.

Algoritma kedua yang dikonstruksikan dalam eksperimen ini adalah XGBoost (Extreme Gradient Boosting), yang diimplementasikan menggunakan fungsi `XGBClassifier()`. Berbeda dengan pendekatan paralel pada Random Forest, XGBoost menerapkan prinsip ensemble boosting sekuensial, di mana pohon-pohon keputusan dibangun secara bertahap dan saling berurutan. Pada setiap iterasi baru, algoritma dirancang untuk mengevaluasi residu kesalahan (`error gradient`) yang dihasilkan oleh pohon pada tahapan sebelumnya, kemudian mengoptimalkan fungsi kerugian (`loss function`) melalui algoritma `gradient descent`. Parameter teknis penting yang dikonfigurasi dalam pemodelan ini mencakup pengaturan tingkat pembelajaran (`learning_rate` atau `eta`) untuk mengontrol kecepatan penyesuaian bobot, jumlah pembentukan pohon (`n_estimators`), serta parameter regulasi internal (L1 dan L2) yang berfungsi mencegah model terlalu sensitif terhadap gangguan noise pada data transaksi nasabah kartu kredit.

Algoritma ketiga sekaligus representasi dari pendekatan Deep Learning dalam penelitian ini adalah Multilayer Perceptron (MLP) Classifier yang dipanggil melalui modul `MLPClassifier()`. Arsitektur jaringan syaraf tiruan berumpan maju (`feedforward neural network`) ini disusun menggunakan struktur lapisan yang saling terhubung penuh (`fully connected layers`). Desain model mencakup satu lapisan input yang menerima 23 variabel prediktor terstandarisasi, beberapa kombinasi lapisan tersembunyi (`hidden layers`) dengan jumlah neuron spesifik, serta satu lapisan luaran biner yang memanfaatkan fungsi aktivasi non-linear untuk memetakan probabilitas risiko gagal bayar. Selama fase pelatihan, fungsi aktivasi Rectified Linear Unit (ReLU) digunakan pada setiap neuron di lapisan tersembunyi untuk mentransformasikan interaksi data secara non-linear, sementara pembaruan bobot dan bias antar-layer dioptimalkan secara berulang menggunakan pengoptimal berbasis `gradient descent` seperti Adam optimizer dengan laju pembelajaran yang konstan.

Analisis Hasil Klasifikasi dan Komparasi Model

Setelah seluruh model selesai dilatih menggunakan data latih yang telah diseimbangkan dengan SMOTE, pengujian dilakukan menggunakan data uji (testing set) untuk mengukur kemampuan generalisasi masing-masing arsitektur secara objektif. Evaluasi performa dari ketiga algoritma tersebut diukur berdasarkan lima metrik utama, yaitu Accuracy, Precision, Recall, F1-Score, dan Area Under the Curve-Receiver Operating Characteristic (AUC-ROC). Hasil komparasi performa final dari ketiga arsitektur secara menyeluruh dirangkum dalam gambar berikut.



```
-----
Hasil Evaluasi: MLP
-----
Accuracy : 0.7992 (79.92%)
Precision : 0.5520 (55.20%)
Recall : 0.4883 (48.83%)
F1-Score : 0.5182 (51.82%)
AUC-ROC : 0.7610 (76.10%)

-----
Hasil Evaluasi: Random Forest
-----
Accuracy : 0.7913 (79.13%)
Precision : 0.5283 (52.83%)
Recall : 0.5268 (52.68%)
F1-Score : 0.5275 (52.75%)
AUC-ROC : 0.7680 (76.80%)

-----
Hasil Evaluasi: XGBoost
-----
Accuracy : 0.7998 (79.98%)
Precision : 0.5566 (55.66%)
Recall : 0.4672 (46.72%)
F1-Score : 0.5000 (50.00%)
AUC-ROC : 0.7616 (76.16%)
-----
```

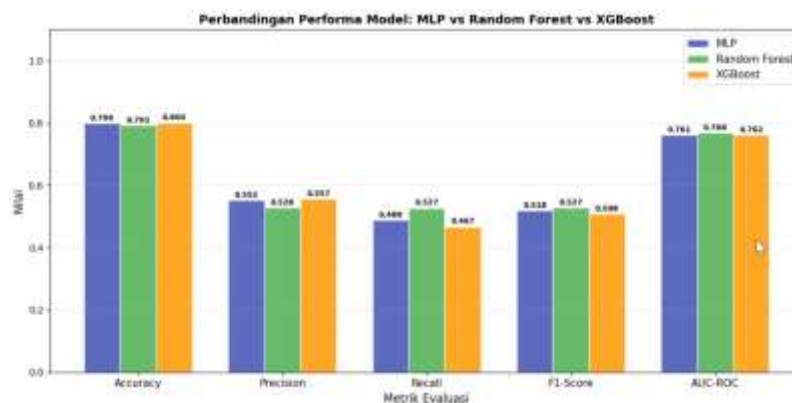
Gambar 4. Hasil Klasifikasi Model

Berdasarkan data yang disajikan pada gambar, algoritma XGBoost mencatatkan nilai Accuracy tertinggi sebesar 0.7998 dan nilai Precision tertinggi sebesar 0.5566, diikuti sangat ketat oleh model *Multilayer Perceptron* (MLP) dengan akurasi sebesar 0.7992 dan presisi sebesar 0.5520. Tingginya nilai presisi pada XGBoost menunjukkan bahwa ketika model ini memprediksi seorang nasabah akan gagal bayar, tingkat kebenaran dari prediksi tersebut sangat meyakinkan. Namun, di sisi lain, XGBoost menghasilkan nilai Recall terendah di antara ketiga model yaitu hanya sebesar 0.4672, yang berarti masih banyak nasabah yang sebenarnya berisiko gagal bayar namun lolos dari deteksi model (*False Negative*).

Dalam konteks analisis risiko kredit perbankan, penentuan model terbaik tidak dapat didasarkan pada nilai akurasi global atau presisi semata. Fokus utama evaluasi wajib dialihkan pada nilai kapabilitas pemisahan kelas (AUC-ROC) serta tingkat sensitivitas (Recall) untuk meminimalisir lolosnya nasabah berisiko gagal bayar yang berpotensi memicu kerugian finansial langsung bagi institusi. Berdasarkan parameter krusial tersebut, algoritma Random Forest terbukti menjadi arsitektur yang paling efektif, seimbang, dan superior. Model Random Forest berhasil memperoleh nilai AUC-

ROC tertinggi mencapai 0.7680, skor F1-Score tertinggi sebesar 0.5275, serta tingkat Recall paling optimal sebesar 0.5268.

Tingginya nilai Recall pada Random Forest (0.5268) dibandingkan dengan MLP (0.4883) dan XGBoost (0.4672) mengonfirmasi bahwa penggabungan teknik resampling SMOTE bekerja jauh lebih seimbang di dalam arsitektur berbasis pohon keputusan paralel ini. Mekanisme ensemble bagging dan seleksi fitur secara acak (feature bagging) pada Random Forest mampu mereduksi dampak kemiripan informasi akibat pembentukan data sintesis baru dari SMOTE, sekaligus robust terhadap fitur-fitur yang memiliki korelasi tinggi. Sebaliknya, model MLP menempati posisi performa AUC-ROC paling rendah yaitu 0.7610, disusul oleh XGBoost sebesar 0.7616. Karakteristik hubungan interaksi fitur data transaksi nasabah kartu kredit terbukti lebih stabil dipetakan oleh model Random Forest, sehingga algoritma ini menjadi rekomendasi praktis terbaik bagi manajemen risiko institusi keuangan dalam mengotomatisasi sistem penilaian kelayakan kredit secara proaktif. Hasil perbandingan performa model dapat dilihat dengan lebih jelas seperti pada grafik berikut.

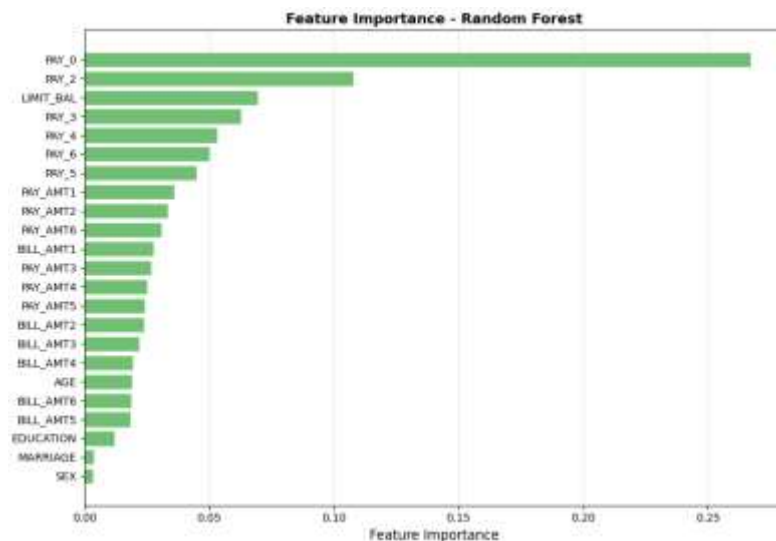


Gambar 5. Grafik Perbandingan Performa Model

Berdasarkan seluruh metrik evaluasi yang telah disajikan, penelitian ini menyimpulkan bahwa Random Forest merupakan algoritma yang paling optimal untuk klasifikasi risiko gagal bayar kredit nasabah. Meskipun XGBoost unggul tipis dalam akurasi global (0.7998), namun Random Forest menunjukkan dominasi pada metrik yang jauh lebih krusial dalam manajemen risiko, yaitu Recall (0.5268), F1-Score (0.5275), dan AUC-ROC (0.7680). Keunggulan Random Forest dalam hal Recall memberikan dampak signifikan bagi institusi keuangan. Dalam skenario riil, kegagalan mendeteksi nasabah yang sebenarnya akan gagal bayar (False Negative) jauh lebih berbahaya dan merugikan secara finansial dibandingkan kesalahan memprediksi nasabah lancar sebagai gagal bayar (False Positive). Dengan nilai Recall tertinggi, Random Forest terbukti lebih sensitif dalam menjangkau potensi nasabah bermasalah. Hal ini juga membuktikan bahwa karakteristik arsitektur ensemble bagging pada Random Forest sangat sinkron dengan teknik resampling SMOTE, di mana model mampu menangkap

pola dari data sintetis tanpa terjebak pada noise atau korelasi antar-fitur tagihan yang sempat teridentifikasi pada analisis heatmap sebelumnya.

Sebagai tahap akhir analisis, dilakukan identifikasi variabel apa saja yang paling mempengaruhi keputusan klasifikasi pada model terbaik (*Random Forest*). Tingkat signifikansi masing-masing fitur prediktor divisualisasikan melalui grafik kontribusi fitur pada Gambar 5.



Gambar 6. Fitur Importan

Berdasarkan grafik Gambar 4, ditemukan bahwa faktor perilaku finansial (behavioral data) merupakan prediktor yang paling mendominasi. Fitur status keterlambatan pembayaran pada bulan-bulan terakhir, yaitu PAY_0 dan PAY_2, serta limit saldo kartu kredit (LIMIT_BAL) berada di posisi teratas. Temuan ini secara empiris membuktikan bahwa pola keterlambatan pembayaran di masa lalu merupakan indikator terkuat bagi sistem untuk memprediksi risiko gagal bayar di masa depan. Sebaliknya, variabel demografis seperti tingkat pendidikan (EDUCATION), jenis kelamin (SEX), dan status pernikahan (MARRIAGE) memberikan kontribusi yang sangat minim terhadap hasil prediksi. Hal ini memperkuat argumen bahwa untuk membangun sistem credit scoring yang akurat, institusi perbankan harus lebih menitikberatkan pada data aktivitas transaksi dan kedisiplinan pembayaran nasabah dibandingkan karakteristik demografis personalnya.

KESIMPULAN

Penelitian ini berhasil melakukan analisis komparatif performa algoritma Multilayer Perceptron (MLP), Random Forest, dan XGBoost untuk klasifikasi risiko gagal bayar (default) nasabah kartu kredit dengan mengintegrasikan teknik resampling SMOTE. Berdasarkan hasil eksperimen, ditemukan bahwa karakteristik ketidakseimbangan kelas (class imbalance) awal sebesar 77,9% banding 22,1% serta

adanya korelasi linier yang kuat antar-fitur tagihan bulanan memberikan pengaruh signifikan terhadap kapabilitas konvergensi masing-masing arsitektur model.

Meskipun algoritma XGBoost mencatatkan nilai Accuracy tertinggi global sebesar 0.7998 dan Precision sebesar 0.5566, namun untuk penentuan keputusan model terbaik dalam konteks manajemen risiko finansial perbankan dialihkan pada metrik sensitivitas (Recall), F1-Score, dan AUC-ROC guna meminimalisir lolosnya nasabah bermasalah (False Negative). Melalui parameter krusial tersebut, algoritma Random Forest disimpulkan sebagai arsitektur yang paling efektif, seimbang, dan superior dengan menghasilkan nilai AUC-ROC tertinggi mencapai 0.7680, skor F1-Score sebesar 0.5275, serta tingkat Recall optimal sebesar 0.5268. Keunggulan ini membuktikan bahwa mekanisme ensemble bagging dan seleksi fitur acak pada Random Forest sangat robust dalam memanfaatkan data sintesis hasil SMOTE tanpa terjebak pada isu multikolinieritas fitur.

Selain itu, temuan dari analisis feature importance memberikan kesimpulan teoretis penting bahwa faktor perilaku keuangan (behavioral data) berupa riwayat keterlambatan pembayaran bulan terakhir (PAY_0 dan PAY_2) serta limit saldo (LIMIT_BAL) memiliki kontribusi paling dominan dalam menentukan potensi default nasabah. Sebaliknya, karakteristik latar belakang demografis personal (jenis kelamin, pernikahan, pendidikan) tidak memberikan pengaruh yang signifikan. Oleh karena itu, penelitian ini merekomendasikan institusi perbankan untuk memanfaatkan algoritma ensemble learning berbasis pohon, khususnya Random Forest, yang dikombinasikan dengan teknik penanganan data tidak seimbang untuk mengotomatisasi sistem penilaian kelayakan kredit secara proaktif demi meminimalisir risiko kerugian akibat kredit macet.

DAFTAR PUSTAKA

- Brown, K., Smith, J., & Davis, L. (2023). Integrated ensemble-based machine learning with SMOTE for financial default analysis. *Expert Systems with Applications*, 215, Article 119842.
- Hassan, M. A., Shukur, Z., & Al-Amrani, A. (2023). Optimizer and architectural impact analysis on multilayer perceptron performance for structural risk tasks. *IEEE Access*, 11, 74210-74225.
- Sun, T., Zhao, X., & Liu, Y. (2025). Hybrid clustering-multilayer perceptron architecture for credit evaluation over high-dimensional datasets. *Expert Systems with Applications*, 260, Article 125110.
- Tetteh, J. A., Mensah, E. O., & Asante, K. (2024). Multi-layer perceptron networks for loan default classification in imbalanced environments. *Applied Soft Computing*, 152, Article 111234.
- Yeh, I. C., & Lien, C. H. (2009). Default of credit card clients [Dataset]. UCI Machine Learning Repository. <https://archive.ics.uci.edu/dataset/186/default+of+credit+card+clients>