

PERBANDINGAN RANDOM FOREST DAN XGBOOST DALAM PREDIKSI STATUS PINJAMAN MENGGUNAKAN CROSS VALIDATION DAN ANALISIS FEATURE IMPORTANCE

Josin Bahaswan, Sopingi, Herliyani Hasanah

Teknik Informatika, Universitas Duta Bangsa Surakarta

Jl. Bhayangkara No. 55-57, Tipes, Serengan, Kota Surakarta, Jawa Tengah 57154

220103208@mhs.udb.ac.id

ABSTRAK

Peningkatan penyaluran kredit diikuti oleh meningkatnya risiko gagal bayar yang berdampak pada stabilitas lembaga keuangan. Namun, banyak penelitian komparasi algoritma klasifikasi kredit belum menerapkan skema validasi yang sistematis untuk meminimalisir risiko *data leakage*. Penelitian ini bertujuan membandingkan kinerja Random Forest dan XGBoost dalam memprediksi status pinjaman menggunakan *pipeline* yang dirancang untuk mengurangi risiko *data leakage*. Pendekatan CRISP-DM diterapkan pada *dataset* 50.000 sampel dengan 17 fitur yang diperoleh dari Kaggle. Seluruh tahapan *preprocessing* (imputasi, *One-Hot Encoding*, SMOTE) dan *hyperparameter tuning* (GridSearchCV) dilakukan secara per-fold dalam kerangka *Nested Cross Validation*. Evaluasi menggunakan *Stratified 10-Fold Cross Validation* dengan metrik *Accuracy*, *F1-Score*, *F2-Score*, dan *AUC*. Hasil menunjukkan XGBoost secara konsisten mengungguli Random Forest pada semua metrik (*Accuracy*: 0,9167 vs 0,9013; *F1-Score*: 0,9249 vs 0,9108; *AUC*: 0,9745 vs 0,9649), dengan perbedaan yang signifikan secara statistik berdasarkan *paired t-test* ($p < 0,001$). Estimasi finansial menunjukkan potensi penghematan sebesar 7.564.966 (11,8%), sementara skor_kredit dan rasio_hutang_terhadap_pendapatan teridentifikasi sebagai fitur paling dominan pada kedua model.

Kata kunci : *Random Forest, XGBoost, Cross Validation, Feature Importance, SMOTE, Prediksi Status Pinjaman*

1. PENDAHULUAN

Peningkatan penyaluran kredit oleh lembaga keuangan dalam beberapa tahun terakhir diikuti oleh meningkatnya risiko kredit bermasalah yang berdampak langsung terhadap stabilitas institusi keuangan[1].

Untuk mengatasi hal tersebut, proses evaluasi kelayakan kredit semakin mengadopsi pendekatan *machine learning* guna meningkatkan akurasi prediksi sekaligus efisiensi dalam pengambilan keputusan[2].

Di antara berbagai metode yang digunakan, algoritma berbasis *ensemble* seperti Random Forest dan XGBoost menunjukkan kinerja yang unggul dalam klasifikasi data *tabular* terstruktur[3].

Selain itu, kedua algoritma ini menyediakan informasi *feature importance* secara bawaan, sehingga memungkinkan identifikasi faktor-faktor yang paling berpengaruh terhadap risiko gagal bayar[4].

Berbagai penelitian terdahulu juga menunjukkan bahwa algoritma *ensemble* memiliki kemampuan klasifikasi yang baik pada permasalahan *credit scoring*, meskipun kinerjanya dapat bervariasi tergantung karakteristik data dan skema validasi yang digunakan[5].

Random Forest dan XGBoost merepresentasikan dua pendekatan *ensemble* yang berbeda, yaitu *bagging* dan *boosting*, sehingga perbandingan keduanya dapat memberikan wawasan yang lebih komprehensif terkait *trade-off* dalam prediksi kredit. Random Forest dikenal mampu menangani data berdimensi tinggi serta mengurangi *overfitting* melalui mekanisme

bootstrap aggregating, sehingga relatif stabil terhadap variasi data[6], [7].

Namun, performanya dapat menurun pada data dengan pola non-linear yang kompleks.

Di sisi lain, XGBoost merupakan pengembangan metode *boosting* yang mengoptimalkan fungsi *loss* secara iteratif dan dilengkapi mekanisme regularisasi, sehingga sering menghasilkan performa prediksi yang lebih tinggi, meskipun dengan kompleksitas komputasi yang lebih besar dan sensitivitas terhadap parameter[8], [9].

Meskipun beberapa penelitian sebelumnya telah membandingkan kedua metode tersebut, masih terbatas penelitian yang menerapkan *pipeline Nested Cross Validation* yang dirancang untuk meminimalisir risiko *data leakage* secara sistematis, menggunakan *F2-Score* sebagai metrik evaluasi tambahan yang lebih relevan dalam konteks risiko kredit, serta mengonfirmasi perbedaan performa model melalui uji statistik *paired t-test*.

Penelitian ini bertujuan untuk membandingkan kinerja Random Forest dan XGBoost dalam prediksi status pinjaman, menganalisis fitur-fitur yang paling berpengaruh terhadap risiko gagal bayar, serta memvalidasi perbedaan kinerja kedua model secara statistik menggunakan *paired t-test*.

Berdasarkan karakteristik permasalahan klasifikasi status kredit yang melibatkan variabel numerik dan kategorikal dengan potensi ketidakseimbangan kelas, penelitian ini menggunakan pendekatan komparatif antara Random Forest dan

XGBoost dengan skema evaluasi *Stratified 10-Fold Cross Validation* [10], [11].

Untuk mengatasi ketidakseimbangan data, diterapkan teknik SMOTE pada data pelatihan di setiap *fold* tanpa memengaruhi data pengujian guna mencegah *data leakage* [12], [13].

Selain itu, dilakukan proses *hyperparameter tuning* secara sistematis untuk memperoleh konfigurasi model yang optimal. Pendekatan ini diharapkan mampu menghasilkan evaluasi yang lebih objektif, stabil, dan representatif dalam mencerminkan kemampuan sebenarnya dari masing-masing algoritma [14], [15].

2. TINJAUAN PUSTAKA

Penelitian sebelumnya yang dilakukan oleh Ulandari yang meneliti prediksi risiko gagal bayar Kredit Kepemilikan Rumah (KPR) menggunakan Random Forest dengan variasi parameter *mtry* dan *ntree*. Model terbaik menghasilkan *Accuracy* 84,75% dan AUC 84,32% pada proporsi data *training* 60%. Namun penelitian ini hanya menggunakan satu algoritma tanpa perbandingan dengan XGBoost maupun pengujian signifikansi statistik antar model [16].

Penelitian lain menerapkan *Logistic Regression* dengan SMOTE untuk prediksi status persetujuan pinjaman nasabah dan memperoleh *Accuracy* sebesar 80,25% serta AUC sebesar 0,8616. Meskipun mampu menangani ketidakseimbangan kelas dengan baik, pendekatan ini menggunakan model tunggal sehingga tidak memungkinkan adanya perbandingan antar algoritma *ensemble* [17].

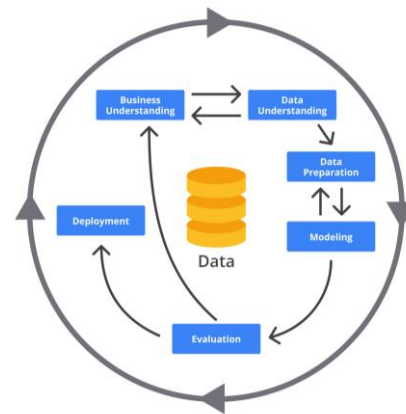
Penelitian membandingkan XGBoost dan Random Forest untuk prediksi potensi banjir di DKI Jakarta menggunakan *GridSearchCV* dan uji *McNemar*. XGBoost menghasilkan *Accuracy* sebesar 82,61%, sedangkan Random Forest sebesar 80,43%, namun perbedaan tersebut tidak signifikan secara statistik. Selain itu, domain penelitian yang digunakan adalah prediksi cuaca, yang memiliki karakteristik berbeda dengan prediksi risiko kredit yang umumnya melibatkan fitur yang lebih kompleks [18].

Penelitian lain juga ada yang menggunakan *Gradient Boosting* dengan SMOTE untuk prediksi risiko gagal bayar premi asuransi dan memperoleh *Accuracy* sebesar 0,89 serta AUC sebesar 0,95. Penelitian ini tidak secara spesifik membandingkan Random Forest dan XGBoost, serta berada pada domain asuransi, bukan kredit [19].

Sementara itu, penelitian yang mengimplementasikan *ensemble* XGBoost dan Random Forest untuk prediksi waktu penggantian baterai menggunakan metrik regresi (MAPE, MAE, RMSE), yang berbeda dengan permasalahan klasifikasi kredit dalam penelitian ini [20].

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yang terdiri dari enam fase *iterative Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment*. Diagram alur CRISP-DM ditampilkan pada Gambar 1.



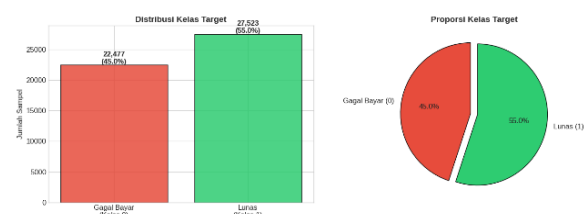
Gambar 1. Diagram alur CRISP-DM

3.1. Business Understanding

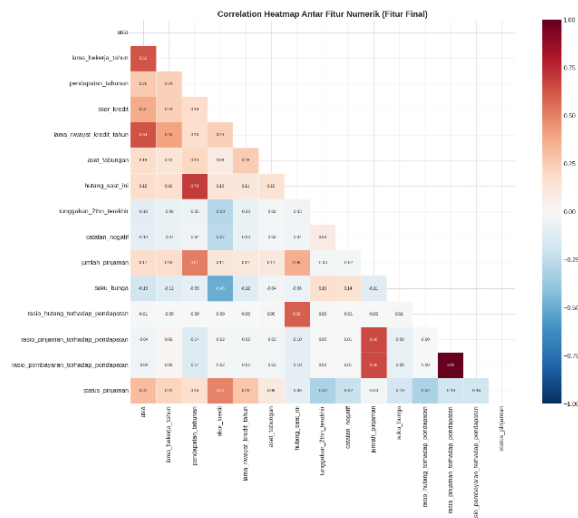
Pada fase ini ditetapkan tujuan penelitian yaitu membandingkan performa Random Forest dan XGBoost dalam memprediksi status pinjaman (Lunas/Gagal Bayar) serta mengidentifikasi fitur-fitur paling berpengaruh. Permasalahan utama adalah bagaimana memperoleh model klasifikasi yang paling akurat, valid secara statistik, dan bermanfaat secara praktis untuk mendukung keputusan pemberian kredit.

3.2. Data Understanding

Dataset diperoleh dari Kaggle (<https://www.kaggle.com/datasets/ardava/prediksi-status-pinjaman>), terdiri dari 50.000 sampel dengan 20 kolom. Variabel target adalah status pinjaman dengan distribusi 27.523 sampel Lunas (55,0%) dan 22.477 sampel Gagal Bayar (45,0%). Ditemukan *missing values* pada kolom usia (142 baris, 0,28%) dan status pekerjaan (158 baris, 0,32%). Distribusi kelas target ditampilkan pada Gambar 2 dan analisis korelasi antar fitur numerik pada Gambar 3.



Gambar 2. Distribusi dan proporsi kelas target

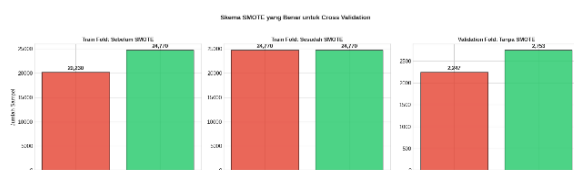


Gambar 3. Correlation heatmap antar fitur numerik (setelah seleksi fitur)

Korelasi tertinggi terhadap variabel target dimiliki oleh skor_kredit (+0,50), diikuti usia (+0,31) dan lama_risiko_kredit_tahun (+0,28). Sebaliknya, tunggakan_2thn_terakhir (-0,32) dan rasio_hutang_terhadap_pendapatan (-0,32) memiliki korelasi negatif, mengindikasikan bahwa peningkatan kedua variabel tersebut berkaitan dengan meningkatnya risiko gagal bayar.

3.3. Data Preparation

Preprocessing dilakukan secara per-fold di dalam *pipeline Nested Cross Validation* untuk mencegah data leakage. Tahapan yang dilakukan meliputi beberapa langkah utama. Pertama, penghapusan kolom, yaitu id_pelanggan karena hanya berfungsi sebagai identifier unik, serta gagal_bayar_tercatat karena bersifat data *leakage* (memiliki hubungan prediktif terhadap target). Kedua, imputasi nilai hilang, di mana kolom usia diisi menggunakan median dari data pada *training fold*, sedangkan status_pekerjaan diisi menggunakan modus dari *training fold*. Ketiga, transformasi fitur kategorikal menggunakan *One-Hot Encoding* (OHE) pada tiga variabel, yaitu status_pekerjaan, tipe_produk, dan tujuan_pinjaman, dengan *OneHotEncoder* yang di-fit hanya pada *X_train_fold*, sehingga menghasilkan total 26 fitur. Keempat, penerapan SMOTE pada data *training fold* untuk menyeimbangkan distribusi kelas dari 20.230:24.770 menjadi 24.770:24.770. Skema penerapan SMOTE pada setiap *fold* ditunjukkan pada Gambar 4.



Gambar 4. Diagram SMOTE

OHE dipilih menggantikan *Label Encoding* karena ketiga variabel tersebut bersifat nominal tanpa urutan/tingkatan. *Label Encoding* berisiko menciptakan hubungan ordinal semu yang dapat disalah interpretasikan oleh algoritma, sementara OHE merepresentasikan setiap kategori sebagai fitur biner tersendiri.

3.4. Modeling

Pemodelan menggunakan dua algoritma dengan mekanisme *ensemble* berbeda: Random Forest (*bagging*) dan XGBoost (*boosting*). *Hyperparameter tuning* dilakukan menggunakan *GridSearchCV* yang dijalankan di dalam setiap *training fold* (*Nested CV*) dengan *scoring F1* dan *Stratified 3-Fold inner*. Rentang parameter yang diujikan dapat dilihat pada Tabel 1 dan 2.

Tabel 1. Rentang hyperparameter random forest

Parameter	Rentang Pencarian	Nilai Terbaik
n_estimators	[100, 200]	200
max_depth	[10, 15]	15
min_samples_split	[2, 5]	2
min_samples_leaf	[1, 2]	1

Tabel 2. Rentang hyperparameter xgboost

Parameter	Rentang Pencarian	Nilai Terbaik
n_estimators	[100, 200]	200
max_depth	[4, 6]	4
learning_rate	[0.05, 0.1]	0.1
subsample	[0.8, 1.0]	0.8
colsample_bytree	[0.8]	0.8

3.5. Evaluation

Evaluasi performa model dilakukan dengan metode *Stratified 10-Fold Cross Validation* menggunakan *random_state=42* guna memastikan stabilitas hasil pengujian. Pengukuran kinerja mencakup metrik *Accuracy*, *Precision*, *Recall*, *F1-Score*, *F2-Score*, dan *AUC*, di mana *F2-Score* (beta=2) diberikan penekanan khusus karena memberikan bobot dua kali lebih besar pada *Recall*. Hal ini sangat penting dalam domain risiko kredit untuk meminimalisir *False Positive*, karena meloloskan peminjam yang gagal bayar (diprediksi Lunas padahal Gagal Bayar) dianggap jauh lebih merugikan bagi lembaga keuangan dibandingkan menolak peminjam yang sebenarnya layak.

3.6. Deployment

Model XGBoost terbaik diimplementasikan sebagai prototipe aplikasi web menggunakan *Streamlit*. Staf kredit dapat memasukkan data peminjam secara interaktif dan sistem mengeluarkan prediksi status pinjaman beserta probabilitasnya sebagai alat bantu pengambilan keputusan kredit berbasis data.

4. HASIL DAN PEMBAHASAN

4.1. Perbandingan Kinerja Model dengan Cross Validation

Hasil *Stratified 10-Fold Cross Validation* untuk Random Forest ditampilkan pada Gambar 5 dan XGBoost pada Gambar 6. Gambar ini menunjukkan nilai metrik pada setiap *fold* beserta *mean* dan standar deviasi.

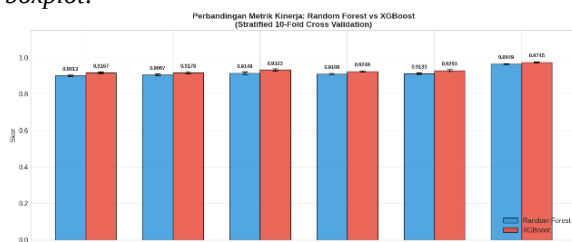
Fold	Accuracy	Precision	Recall	F1-Score	F2-Score	AUC
1	0.9040	0.9107	0.9154	0.9130	0.9144	0.9643
2	0.9024	0.9090	0.9143	0.9116	0.9132	0.9666
3	0.8936	0.8931	0.9165	0.9046	0.9117	0.9607
4	0.9032	0.9059	0.9197	0.9127	0.9169	0.9683
5	0.8958	0.9106	0.8990	0.9047	0.9013	0.9632
6	0.9074	0.9157	0.9161	0.9159	0.9160	0.9680
7	0.9060	0.9107	0.9193	0.9150	0.9176	0.9652
8	0.9016	0.9044	0.9182	0.9113	0.9154	0.9657
9	0.9022	0.9028	0.9215	0.9121	0.9177	0.9647
10	0.8970	0.9039	0.9095	0.9067	0.9084	0.9619
Mean	0.9013	0.9067	0.9149	0.9108	0.9133	0.9649
Std	0.0042	0.0059	0.0062	0.0038	0.0049	0.0023

Gambar 5. Hasil evaluasi random forest per fold (Stratified 10-Fold CV)

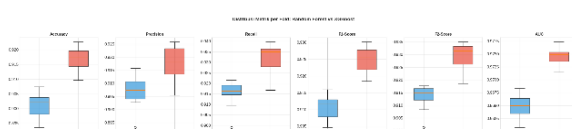
Fold	Accuracy	Precision	Recall	F1-Score	F2-Score	AUC
1	0.9158	0.9129	0.9364	0.9245	0.9316	0.9735
2	0.9226	0.9243	0.9361	0.9302	0.9337	0.9765
3	0.9108	0.9054	0.9357	0.9203	0.9295	0.9718
4	0.9198	0.9167	0.9397	0.9280	0.9350	0.9774
5	0.9138	0.9259	0.9168	0.9213	0.9186	0.9754
6	0.9160	0.9219	0.9259	0.9239	0.9251	0.9752
7	0.9214	0.9238	0.9342	0.9290	0.9321	0.9748
8	0.9188	0.9195	0.9342	0.9268	0.9313	0.9750
9	0.9186	0.9153	0.9390	0.9270	0.9341	0.9738
10	0.9096	0.9122	0.9248	0.9184	0.9222	0.9715
Mean	0.9167	0.9178	0.9323	0.9249	0.9293	0.9745
Std	0.0041	0.0062	0.0070	0.0037	0.0052	0.0018

Gambar 6. Hasil Evaluasi xgboost per Fold (Stratified 10-Fold CV)

Untuk memvisualisasikan perbedaan secara lebih jelas, Gambar 7 menampilkan perbandingan metrik dengan *error bar* (standar deviasi) dan Gambar 8 menampilkan distribusi metrik per *fold* dalam bentuk *boxplot*.



Gambar 7. Perbandingan Metrik Kinerja RF vs xgboost



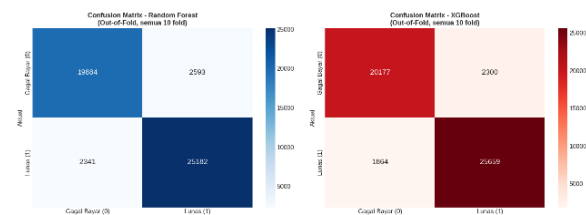
Gambar 8. Boxplot Distribusi Metrik

XGBoost secara konsisten mengungguli Random Forest pada semua metrik yang diuji dalam penelitian ini. Selisih terbesar terdapat pada Recall (+1,74%) dan F2-Score (+1,60%), yang secara khusus penting

karena *F2-Score* memberikan penekanan lebih besar pada *Recall* sehingga kesalahan *False Negative* diperhitungkan lebih berat dalam penilaian performa model. *Boxplot* pada Gambar 8 menunjukkan XGBoost menunjukkan variabilitas yang lebih rendah dibanding Random Forest, khususnya pada metrik *F1-score* (0,0037 vs 0,0038) dan *AUC* (0,0018 vs 0,0023), yang mengindikasikan konsistensi performa yang lebih baik antar *fold*.

4.2. Analisis Confusion Matrix

Confusion matrix agregat (*Out-of-Fold* dari semua 10 *fold*) ditampilkan pada Gambar 9, menunjukkan prediksi pada seluruh 50.000 sampel *dataset*. *Classification report* untuk masing-masing model ditampilkan pada Gambar 10 dan Gambar 11.



Gambar 9. Confusion matrix random forest dan xgboost

Classification Report - Random Forest (Out-of-Fold, semua fold):				
	precision	recall	f1-score	support
Gagal Bayar (0)	0.89	0.88	0.89	22477
Lunas (1)	0.91	0.91	0.91	27523
accuracy			0.90	50000
macro avg	0.90	0.90	0.90	50000
weighted avg	0.90	0.90	0.90	50000

Gambar 10. Hasil evaluasi random forest per fold (Stratified 10-Fold CV)

Classification Report - XGBoost (Out-of-Fold, semua fold):				
	precision	recall	f1-score	support
Gagal Bayar (0)	0.92	0.90	0.91	22477
Lunas (1)	0.92	0.93	0.92	27523
accuracy			0.92	50000
macro avg	0.92	0.91	0.92	50000
weighted avg	0.92	0.92	0.92	50000

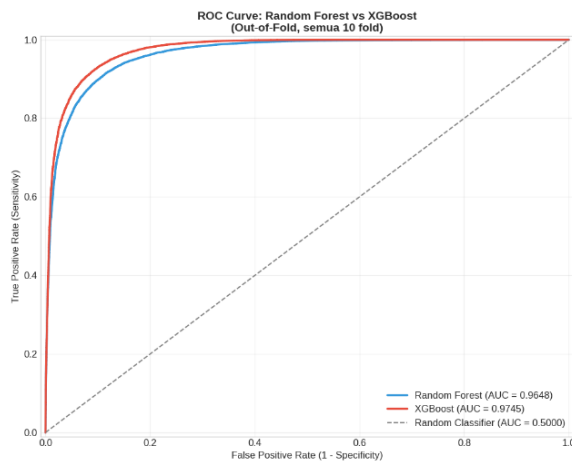
Gambar 11. Hasil evaluasi xgboost per fold (stratified 10-fold cv)

XGBoost menghasilkan jumlah *True Negative* sebesar 20.177 (dibandingkan Random Forest: 19.884) dan *True Positive* sebesar 25.659 (Random Forest: 25.182). Perbedaan paling menonjol terlihat pada *False Positive*, di mana XGBoost menghasilkan nilai yang lebih rendah, yaitu 2.300 dibandingkan 2.593 pada Random Forest (selisih 293 kasus). Dalam konteks penilaian risiko kredit, *False Positive* merepresentasikan peminjam yang sebenarnya berpotensi gagal bayar tetapi diprediksi Lunas oleh model sehingga tetap diberikan persetujuan pinjaman, dan merupakan jenis kesalahan yang paling merugikan bagi lembaga keuangan. Selain itu, XGBoost juga menunjukkan *False Negative* yang lebih rendah, yaitu

1.864 dibandingkan 2.341 pada Random Forest (selisih 477 kasus).

4.3. Analisis ROC Curve

Kurva ROC (*Out-of-Fold*, semua 10 *fold*) pada Gambar 12 membandingkan kemampuan diskriminasi kedua model secara visual.

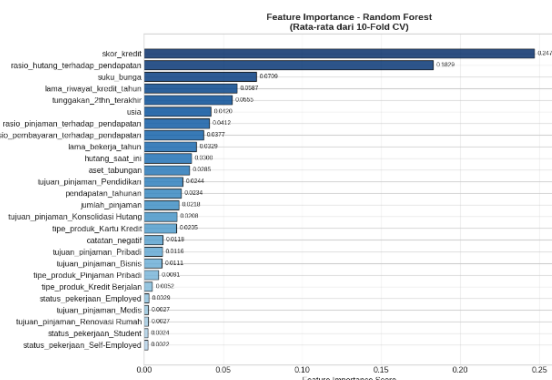


Gambar 12. ROC curve RF vs XGBoost

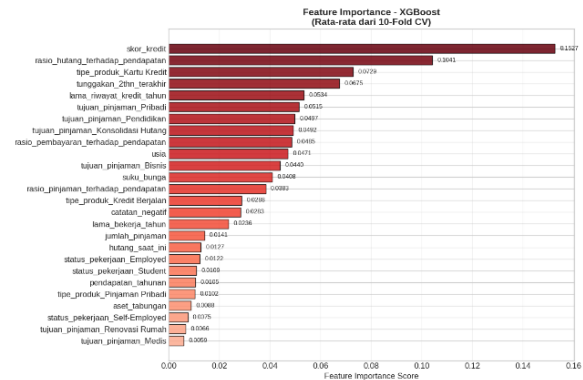
XGBoost memperoleh AUC 0,9745 dibandingkan Random Forest 0,9649 (selisih 0,0096). Kurva XGBoost secara konsisten berada di atas kurva Random Forest pada seluruh rentang *False Positive Rate*, mengkonfirmasi kemampuan diskriminasi yang lebih baik dalam membedakan peminjam Lunas dan Gagal Bayar. Kedua model menunjukkan performa jauh di atas *random classifier* (AUC = 0,5000).

4.4. Analisis Feature Importance

Analisis *feature importance* dihitung sebagai rata-rata nilai *importance* dari 10 *fold* untuk masing-masing model. Dengan penerapan *One-Hot Encoding* per-*fold*, fitur kategorikal (tujuan_pinjaman, tipe_produk, status_pekerjaan) kini pecah menjadi kolom biner tersendiri yang dapat muncul dalam *ranking*. *Feature importance* Random Forest ditampilkan pada Gambar 13 dan XGBoost pada Gambar 14.



Gambar 13. Feature importance random forest



Gambar 14. Feature importance xgboost

Kedua model secara konsisten mengidentifikasi skor_kredit sebagai fitur paling dominan (RF: 0,247; XGB: 0,153), diikuti oleh rasio_hutang_terhadap_pendapatan (RF: 0,183; XGB: 0,104). Konsistensi ini, meskipun berasal dari dua algoritma dengan mekanisme berbeda, memperkuat validitas temuan bahwa skor kredit dan beban hutang merupakan faktor utama indikasi dalam prediksi risiko gagal bayar.

Perbedaan yang menonjol terlihat pada fitur tipe_produk_Kartu Kredit. XGBoost menempatkannya pada peringkat ketiga (0,073), sementara Random Forest menempatkannya jauh lebih rendah, yaitu pada peringkat ke-16 (0,021). Hal ini menunjukkan bahwa XGBoost lebih sensitif dalam menangkap risiko spesifik yang berkaitan dengan jenis produk kartu kredit, yang direpresentasikan melalui proses *One-Hot Encoding*. Sebaliknya, Random Forest memberikan bobot yang lebih tinggi pada suku_bunga (peringkat ketiga, 0,071), yang lebih mencerminkan karakteristik umum dari kondisi pinjaman.

Temuan ini mengindikasikan bahwa, selain skor kredit dan rasio hutang, jenis produk pinjaman juga merupakan faktor yang relevan dan perlu diperhatikan secara khusus dalam proses penilaian risiko oleh lembaga keuangan.

4.5. Estimasi Kerugian Finansial

Untuk mengkuantifikasi dampak praktis dari perbedaan performa kedua model, dilakukan estimasi kerugian finansial berdasarkan beberapa asumsi. Asumsi-asumsi ini bersifat ilustratif dan ditetapkan dalam penelitian ini semata-mata untuk membandingkan dampak finansial relatif antara kedua model, bukan sebagai estimasi kerugian absolut, yaitu rata-rata nilai pinjaman sebesar 33.042, *recovery rate* sebesar 30% sehingga kerugian per *False Positive* adalah 23.129, serta profit margin sebesar 5% sehingga potensi keuntungan yang hilang (*missed profit*) per *False Negative* adalah 1.652. Dalam penelitian ini, definisi *False Positive* (FP) dan *False Negative* (FN) mengikuti standar *machine learning*, di mana *False Positive* merepresentasikan kasus ketika model memprediksi Lunas namun sebenarnya Gagal Bayar, sedangkan *False Negative* merepresentasikan kasus ketika model memprediksi Gagal Bayar namun

sebenarnya Lunas. Dalam konteks risiko kredit, *False Positive* merupakan jenis kesalahan yang paling merugikan karena berkaitan dengan persetujuan

pinjaman kepada peminjam yang berpotensi gagal bayar.

Tabel 3. Perbandingan estimasi kerugian finansial

Komponen	Random Forest	XGBoost	Keterangan
False positive (FP)	2.593	2.300	Prediksi lunas, aslinya gagal bayar
False negative (FN)	2.341	1.864	Prediksi gagal bayar, aslinya lunas
Kerugian FP	59.974.534	53.197.620	23.129
Kerugian FN	3.867.566	3.079.514	1.652 (profit margin 5%)
Total kerugian	63.842.100	56.277.134	XGBoost hemat 7.564.966 (11,8%)

Berdasarkan hasil estimasi, XGBoost berpotensi menghemat total kerugian sebesar 7.564.966 atau sekitar 11,8% dibandingkan Random Forest. Penghematan ini terutama berasal dari penurunan jumlah *False Positive* sebesar 6.776.914, yang menunjukkan berkurangnya jumlah pinjaman bermasalah yang disetujui. Di sisi lain, pengurangan *False Negative* memberikan tambahan penghematan sebesar 788.052, yang mencerminkan lebih sedikitnya peminjam layak yang ditolak sehingga peluang keuntungan tetap terjaga. Dalam skala portofolio kredit yang lebih besar, perbedaan ini dapat

memberikan dampak finansial yang signifikan bagi lembaga keuangan.

4.6. Uji Statistik Paired T-Test

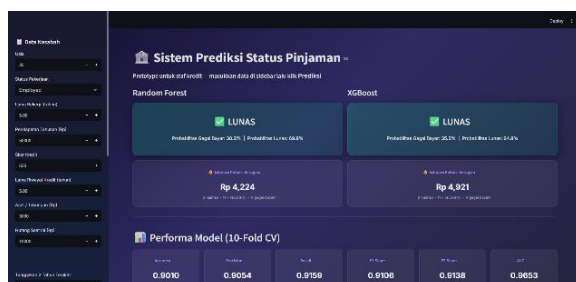
Untuk memverifikasi bahwa perbedaan kinerja antara kedua model tidak disebabkan oleh variasi acak antar *fold*, dilakukan *Paired T-Test* pada 10 pasang nilai metrik. *Paired T-Test* dipilih karena kedua model dievaluasi pada *fold* yang identik, sehingga nilai metrik tiap *fold* bersifat berpasangan dan memungkinkan pengontrolan variasi antar *fold*. Hasil pengujian disajikan pada Tabel 4.

Tabel 4. Hasil paired t-test random forest vs XGBoost

Metrik	Mean RF	Mean XGB	t-statistic	p-value	Keputusan
Accuracy	0.9013	0.9167	-14.2063	$1,81 \times 10^{-7}$	Tolak H_0
Precision	0.9067	0.9178	-8.0326	$2,14 \times 10^{-5}$	Tolak H_0
Recall	0.9149	0.9323	-15.4385	$8,78 \times 10^{-8}$	Tolak H_0
F1-Score	0.9108	0.9249	-14.7814	$1,28 \times 10^{-7}$	Tolak H_0
F2-Score	0.9133	0.9293	-16.4319	$5,10 \times 10^{-8}$	Tolak H_0
AUC	0.9649	0.9745	-23.1421	$2,50 \times 10^{-9}$	Tolak H_0

4.7. Prototype Streamlit

Model diimplementasikan sebagai prototipe aplikasi web menggunakan *Streamlit* guna memvalidasi kelayakan penerapan model dalam skenario nyata. Antarmuka aplikasi ditampilkan pada Gambar 15.



Gambar 15. Tampilan antarmuka prototipe aplikasi prediksi status pinjaman

Prototipe ini memungkinkan staf kredit memasukkan data peminjam secara interaktif melalui formulir input yang mencakup seluruh 17 fitur relevan, meliputi skor_kredit, rasio_hutang_terhadap_pendapatan, tipe_produk, tujuan_pinjaman, dan atribut lainnya. Sistem kemudian mengeluarkan prediksi status pinjaman (Lunas atau Gagal Bayar) beserta nilai probabilitasnya

sebagai representasi tingkat kepercayaan model terhadap prediksi yang dihasilkan. Prototipe ini bersifat sebagai alat bantu pengambilan keputusan (*decision support tool*) dan tidak dimaksudkan sebagai pengganti penilaian kredit secara menyeluruh oleh analis.

4.8. Diskusi

Keunggulan XGBoost atas Random Forest dalam penelitian ini dapat dijelaskan melalui dua aspek algoritmik. Pertama, mekanisme *boosting* yang secara iteratif memperbaiki *residual error* lebih efektif menangkap pola non-linear kompleks pada data keuangan. Kedua, kombinasi parameter *max_depth*=4 dan *subsample*=0,8 memberikan regularisasi yang mencegah *overfitting* pada 26 fitur pasca-OHE.

Temuan ini sejalan dengan Wardoyo yang juga menemukan XGBoost lebih unggul dari Random Forest, namun berbeda dalam hal signifikansi statistik. Pada penelitian lain menggunakan domain prediksi banjir perbedaan tidak signifikan menurut Uji *McNemar*, sementara pada penelitian ini sangat signifikan ($p < 0,001$). Perbedaan ini kemungkinan disebabkan karakteristik data kredit yang memiliki pola non-linear lebih kompleks dibandingkan data cuaca[18].

Validasi *leakage* melalui Label *Shuffle Test* (RF: AUC 0,5025; XGB: AUC 0,5005) mengkonfirmasi bahwa hasil penelitian ini valid dan tidak diinflasikan oleh kebocoran data metodologis. Perbedaan AUC yang sangat kecil antara *preprocessing global* vs *per-fold* (RF: 0,0004; XGB: 0,0001) juga menunjukkan bahwa dampak *leakage* pada kasus ini minimal, namun penerapan *pipeline per-fold* tetap penting untuk keketatan metodologis.

5. KESIMPULAN DAN SARAN

Penelitian ini membandingkan kinerja Random Forest dan XGBoost dalam memprediksi status pinjaman menggunakan pendekatan CRISP-DM dengan *Stratified 10-Fold Cross Validation* serta *pipeline preprocessing* yang dirancang untuk meminimalisir risiko data *leakage*. Hasil evaluasi menunjukkan bahwa XGBoost secara konsisten mengungguli Random Forest pada semua metrik yang diuji dalam penelitian ini, yaitu *Accuracy* (0,9167 vs 0,9013), *F1-Score* (0,9249 vs 0,9108), *F2-Score* (0,9293 vs 0,9133), dan *AUC* (0,9745 vs 0,9649). Seluruh perbedaan tersebut terbukti signifikan secara statistik berdasarkan uji *paired t-test* ($p < 0,001$). Dari perspektif finansial, pada simulasi yang dilakukan penggunaan XGBoost berpotensi mengurangi total kerugian sebesar 7.564.966 (11,8%) dibandingkan Random Forest. Analisis *feature importance* menunjukkan bahwa skor_kredit dan rasio_hutang_terhadap_pendapatan merupakan fitur paling dominan pada kedua model. Selain itu, XGBoost mampu mengidentifikasi pentingnya fitur tipe_produk_Kartu Kredit, yang tidak terungkap secara optimal pada pendekatan Label *Encoding*, sehingga memberikan *insight* tambahan dalam penilaian risiko kredit.

Untuk penelitian selanjutnya disarankan melakukan eksplorasi SHAP values untuk interpretabilitas prediksi individual, pengujian *LightGBM* atau *CatBoost* sebagai pembanding dan validasi pada *dataset* kredit nyata dari lembaga keuangan Indonesia untuk meningkatkan generalisasi hasil.

DAFTAR PUSTAKA

- [1] H. Ighni and F. Komariyah, "Peran Ojk Dalam Menangani Kredit Macet Untuk Stabilitas Keuangan Perusahaan: Studi Kasus Pt Tab," *JURNAL ILMIAH EKONOMI, MANAJEMEN, BISNIS DAN AKUNTANSI*, vol. 2, no. 4, pp. 340–350, 2025.
- [2] I. A. Wibowo and M. Kom, *KECERDASAN BISNIS (BI) DIDUKUNG AI: Meningkatkan prediksi dan pembuatan keputusan dengan ML (Machine Learning)*. yayasan penerbit, 2025.
- [3] M. R. A. Pramudya, C. Arafat, M. C. Ramadhan, F. A. Jafar, and E. Ismanto, "Perbandingan Algoritma Random Forest Dan Xgboost Untuk Klasifikasi Penyakit Jantung Berdasarkan Data Medis," *JURNAL FASILKOM*, vol. 15, no. 2, pp. 430–435, 2025.
- [4] S. NURADILLA, K. SADIK, C. SUHAENI, and A. M. SOLEH, "Klasifikasi Halaman SEO Berbasis Machine Learning Melalui Mutual Information dan Random Forest Feature Importance," *MIND (Multimedia Artificial Intelligent Networking Database) Journal*, vol. 10, no. 1, pp. 114–129, 2025.
- [5] T. P. Wahjuningsih, T. A. Setiawan, A. Ilyas, and A. Subagyo, "Optimasi Sistem Credit Scoring Pembiayaan pada Lembaga Keuangan Mikro dengan Neural Network Integrasi Sample Bootstrapping dan Weighted PCA," *Dinamik*, vol. 31, no. 1, pp. 267–275, 2026.
- [6] S. Wibisono, "KOMPARASI DAN IMPLEMENTASI ALGORITMA MACHINE LEARNING UNTUK KLASIFIKASI KREDIT BERMASALAH PADA PT BPR NUSUMMA KLATEN," *ISMETEK*, vol. 19, no. 1, 2025.
- [7] H. Putra and R. Rumini, "Comparative Study of Logistic Regression, Random Forest, and XGBoost for Bank Loan Approval Classification," *Journal of Applied Informatics and Computing*, vol. 9, no. 5, pp. 2822–2835, 2025.
- [8] A. Sah, C. Niesa, R. R. Jafar, and M. Muharrom, "Analisis Model Prediksi Penyakit Jantung Menggunakan Adaptive Boosting, Gradient Boosting, dan Extreme Gradient Boosting," *Jurnal Ilmiah FIFO*, vol. 17, no. 1, pp. 46–56, 2025.
- [9] S. Mujiyono, U. P. Sanjaya, I. S. Wibisono, and H. Setyowati, "Prediksi Fluktuasi Berat Badan Berdasarkan Pola Hidup Menggunakan Model XGBoost dan Deep Learning," *Jurnal Algoritma*, vol. 22, no. 1, pp. 221–233, 2025.
- [10] R. Wahyudi, K. I. Manik, M. Alfin, J. B. Henrydunan, A. Arnita, and F. Ramadhani, "ANALISIS FAKTOR KEPUTUSAN NASABAH BERLANGGANAN TERM DEPOSIT: PERBANDINGAN RANDOM FOREST DAN XGBOOST," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 4, pp. 5600–5607, 2025.
- [11] I. A. Ubed, I. Syarif, and F. A. Saputra, "Analisis Kinerja Algoritma Machine Learning Untuk Klasifikasi Potensi FRAUD Klaim Layanan Kesehatan Rumah Sakit," 2025.
- [12] A. A. Qolbu, N. Fitriyati, and N. Inayah, "Performa Naïve Bayes, SVM, dan IndoBERT pada Analisis Sentimen Twitter IndiHome dengan Strategi Penanganan Data Tidak Seimbang," *Jurnal Fourier*, vol. 14, no. 1, pp. 29–44, 2025.
- [13] A. Ichwani, R. I. Kesuma, A. Setiawan, I. E. Wicaksono, and R. Hanifah, "Preventing Data Leakage in Classification via Integrated Machine Learning Pipelines: Preprocessing, Feature Transformation, and Hyperparameter Tuning,"

- Jurnal Teknik Informatika (Jutif)*, vol. 7, no. 1, pp. 391–410, 2026.
- [14] H. A. Nabila and E. W. Pamungkas, “PERBANDINGAN ALGORITMA MACHINE LEARNING: SVM, RANDOM FOREST, DAN XGBOOST UNTUK PREDIKSI STROKE: COMPARISON OF MACHINE LEARNING ALGORITHMS: SVM, RANDOM FOREST, AND XGBOOST FOR STROKE PREDICTION,” *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 10, no. 2, pp. 1098–1110, 2025.
- [15] R. T. Nugraha, D. Hidayat, and F. Mahardika, “Analisis Perbandingan Algoritma Random Forest, SVM, dan Neural Network untuk Klasifikasi Risiko Keamanan E-Learning,” *Jurnal Teknologi dan Informasi*, vol. 16, no. 1, pp. 1–10, 2026.
- [16] K. P. Ulandari, N. Chamidah, and A. Kurniawan, “Prediksi risiko gagal bayar kredit kepemilikan rumah dengan pendekatan metode random forest,” *SAINSMAT: Jurnal Ilmiah Ilmu Pengetahuan alam*, vol. 13, no. 2, pp. 162–170, 2024.
- [17] M. Syamsuddin Wisnubroto, F. Farid, A. M. Putri, D. A. Khayyuninafsyah, F. Aqila, and S. Achjar, “Jurnal Restikom: Riset Teknik Informatika dan Komputer Penerapan Logistic Regression untuk Prediksi Status Persetujuan Pinjaman Nasabah,” vol. 7, no. 3, pp. 483–497, 2025, [Online]. Available: <https://restikom.nusaputra.ac.id>
- [18] A. P. Wardoyo, “Analisis Perbandingan Model XGBoost dan Random Forest untuk Prediksi Potensi Banjir di DKI Jakarta,” 2026.
- [19] Q. Amin Siregar *et al.*, “PREDIKSI RISIKO GAGAL BAYAR PREMI MENGGUNAKAN ALGORITMA GRADIENT BOOSTING: STUDI TRAVEL INSURANCE PREDICTION INFO ARTIKEL ABSTRAK”, doi: 10.58290/jukomte.
- [20] M. A. Rayadin, M. Musaruddin, R. A. Saputra, and I. Isnawaty, “Implementasi Ensemble Learning Metode XGBoost dan Random Forest untuk Prediksi Waktu Penggantian Baterai Aki,” *BIOS: Jurnal Teknologi Informasi dan Rekayasa Komputer*, vol. 5, no. 2, pp. 111–119, Aug. 2024, doi: 10.37148/bios.v5i2.128