JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. 14 No. 3, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v14i3.9842

ANALISIS EXPLAINABLE AI PADA PERBANDINGAN MODEL XGBOOST DAN LOGISTIC REGRESSION UNTUK CREDIT SCORING

Ari Fullah[1*], Fathir Januarta[2], Vashih Al Farizi[3], Muhammad Nashrul Fakhri[4], Anindita Septiarini[5], Joan Angelina Widians[6], Akhmad Irsyad[7]

[1,2,3,4,5,6,7]Informatika, Universitas Mulawarman; Jalan Sambaliung No.9, Kota Samarinda 75119 Kalimantan Timur; telp. +(62) 811 555 1962, +(0541) 541 749 159

Keywords:

Credit Scoring;
SHAP;
Logistic Regression;
XGBoost;
Explainable Artificial Intelligence.

Corespondent Email:
arifullahtgr@gmail.com

Abstrak. Perkembangan machine learning dalam industri keuangan mendorong penggunaan model prediktif yang semakin kompleks untuk penilaian risiko kredit. Meskipun model seperti XGBoost menawarkan akurasi tinggi, tantangan utama yang muncul adalah kurangnya transparansi dalam pengambilan keputusan. Penelitian ini bertujuan membandingkan performa model XGBoost dan Logistic Regression dalam prediksi credit scoring, serta menganalisis interpretabilitas kedua model menggunakan metode SHAP (SHapley Additive exPlanations). Data yang digunakan adalah Credit Risk Dataset dari platform Kaggle yang terdiri atas 32.581 observasi dengan 11 variabel independen dan 1 variabel target. Tahapan penelitian meliputi preprocessing data, pemodelan, evaluasi, dan analisis SHAP. Hasil evaluasi menunjukkan XGBoost mengungguli Logistic Regression dengan Accuracy 0,9122 berbanding 0,8104, F1-Score 0,7956 berbanding 0,6356, dan ROC AUC 0,9458 berbanding 0,8635. Analisis SHAP mengungkap bahwa loan_percent_income dan loan_int_rate merupakan fitur paling berpengaruh pada kedua model, konsisten dengan konsep Debt-to-Income Ratio dalam analisis kredit konvensional. Meskipun XGBoost unggul secara prediktif, SHAP terbukti mampu menjelaskan mekanisme keputusan kedua model secara transparan. Penelitian ini menunjukkan bahwa kombinasi XGBoost dan SHAP merupakan pendekatan yang direkomendasikan untuk pengembangan sistem credit scoring yang akurat sekaligus dapat dipertanggungjawabkan.

Abstract. *The advancement of machine learning in the financial industry has driven the adoption of increasingly complex predictive models for credit risk assessment. While models such as XGBoost offer high predictive accuracy, a key challenge remains the lack of transparency in decision-making. This study aims to compare the performance of XGBoost and Logistic Regression in credit scoring prediction, and to analyze the interpretability of both models using the SHAP (SHapley Additive exPlanations) method. The dataset used is the Credit Risk Dataset from the Kaggle platform, consisting of 32,581 observations with 11 independent variables and 1 target variable. Research stages include data preprocessing, modeling, evaluation, and SHAP analysis. Evaluation results show that XGBoost outperforms Logistic Regression with an Accuracy of 0.9122 versus 0.8104, F1-Score of 0.7956 versus 0.6356, and ROC AUC of 0.9458 versus 0.8635. SHAP analysis reveals that loan_percent_income and loan_int_rate are the most influential features in both models, consistent with the Debt-to-Income Ratio concept in conventional credit analysis. Although XGBoost is superior in predictive performance, SHAP effectively explains the decision mechanisms of both models in a transparent manner. This study demonstrates that the combination*

of XGBoost and SHAP is a recommended approach for developing credit scoring systems that are both accurate and accountable.

1. PENDAHULUAN

Perkembangan teknologi machine learning dalam industri keuangan telah mengubah cara lembaga keuangan melakukan penilaian risiko kredit (credit scoring). Credit Scoring merupakan salah satu metode yang digunakan untuk menilai kelayakan dan memprediksi lebih awal adanya potensi kredit macet dari calon nasabah kredit [1].

Secara tradisional, model statistik seperti Logistic Regression yang kemampuannya dalam memberikan interpretasi langsung terhadap pengaruh variabel independen terhadap probabilitas keluaran [2]. Namun, meningkatnya kompleksitas data dan kebutuhan akan akurasi prediksi yang lebih tinggi mendorong penggunaan model berbasis ensemble learning seperti XGBoost, telah diakui karena kinerja prediktifnya yang unggul, efektivitasnya dalam menangani kasus klasifikasi berskala besar, dan kemampuannya mencapai tingkat akurasi klasifikasi yang tinggi [3].

Meskipun model seperti XGBoost menawarkan akurasi yang lebih tinggi dibandingkan model linier tradisional, tantangan utama yang muncul adalah kurangnya transparansi atau interpretabilitas model. Dalam konteks keuangan, transparansi menjadi aspek krusial karena keputusan kredit berdampak langsung pada individu dan harus dapat dipertanggungjawabkan secara regulatif maupun etis. Model yang bersifat black-box berpotensi menimbulkan bias serta menyulitkan proses audit dan penjelasan kepada nasabah maupun regulator [4]. Oleh karena itu, konsep Explainable Artificial Intelligence (XAI) berkembang untuk menjembatani kebutuhan antara akurasi model dan transparansi keputusan [5].

Salah satu pendekatan XAI yang banyak digunakan adalah SHAP (SHapley Additive exPlanations), yang didasarkan pada teori nilai Shapley dalam teori permainan. SHAP mampu memberikan interpretasi lokal maupun global terhadap kontribusi setiap fitur dalam menghasilkan prediksi model. Berbeda dengan metode interpretasi tradisional seperti koefisien regresi pada Logistic Regression, SHAP dapat diterapkan pada berbagai jenis model, termasuk model non-linier seperti XGBoost, sehingga memungkinkan perbandingan interpretasi antar model secara lebih komprehensif.

Berbagai penelitian sebelumnya menunjukkan bahwa XGBoost umumnya menghasilkan performa klasifikasi yang lebih baik dibandingkan Logistic Regression dalam kasus credit scoring. Namun, sebagian besar studi hanya berfokus pada perbandingan metrik performa seperti akurasi, precision, recall, atau AUC, tanpa melakukan analisis mendalam terhadap aspek interpretabilitas model [6][7]. Di sisi lain, penelitian yang membahas explain ability seringkali hanya diterapkan pada satu model saja, tanpa membandingkan bagaimana pola interpretasi dapat berbeda ketika dua model dengan karakteristik berbeda digunakan pada dataset yang sama.

Berdasarkan kondisi tersebut, terdapat kesenjangan penelitian (research gap) pada aspek integrasi antara evaluasi performa dan evaluasi interpretabilitas secara komparatif dalam konteks credit scoring. Diperlukan analisis yang tidak hanya menilai model mana yang lebih akurat, tetapi juga bagaimana masing-masing model menjelaskan keputusan prediksinya, serta sejauh mana hasil interpretasi tersebut konsisten dan relevan terhadap variabel risiko kredit. Kebaruan penelitian ini terletak pada pendekatan komparatif yang menggabungkan evaluasi kinerja model dan analisis explainability menggunakan SHAP pada dua pendekatan berbeda, yaitu model linier (Logistic Regression) dan model berbasis gradient boosting (XGBoost).

Adapun tujuan penelitian ini adalah: (1) membandingkan performa model XGBoost dan Logistic Regression dalam memprediksi kelayakan kredit; (2) menganalisis kontribusi fitur menggunakan metode SHAP pada masing-masing model; serta (3) mengevaluasi perbedaan pola interpretasi yang dihasilkan kedua model dalam konteks pengambilan keputusan kredit. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan model credit scoring

yang tidak hanya akurat, tetapi juga transparan dan dapat dipertanggungjawabkan.

2. TINJAUAN PUSTAKA

2.1 Credit Scoring

Credit scoring atau penilaian kredit adalah proses penilaian kelayakan kredit yang dilakukan dengan menggunakan model machine learning untuk memprediksi risiko gagal bayar, yang dalam penelitian ini dianalisis lebih lanjut menggunakan pendekatan Explainable AI guna meningkatkan transparansi dan interpretabilitas model [8].

2.2 Matchine Learning dalam Credit Scoring

Machine learning merupakan pendekatan kecerdasan buatan yang memungkinkan sistem mempelajari pola dari data serta menghasilkan keputusan atau prediksi secara otomatis [9]. Salah satu penerapannya adalah pada credit scoring, dimana kemampuan klasifikasi digunakan untuk menilai kelayakan kredit nasabah yang umumnya terbagi menjadi dua kelas, yaitu layak dan tidak layak menerima kredit, sehingga pendekatan ini dapat dimanfaatkan untuk meningkatkan akurasi dalam prediksi risiko kredit .

2.3 Logistic Regression

Logistic Regression merupakan metode pengklasifikasi yang sangat populer dan digunakan dalam berbagai penelitian [10]. Perhitungan probilitas menggunakan logistic regression :

$$P(y = 1 \mid X) = \frac{1}{1 + e^{(}\beta a + \beta 1x1 + \cdots + \beta \pi x \pi)}$$

Keterangan :

- Y = 1 kredit macet
- e = bilangan eksponensial atau euler 2,71828
- β_0 = konstanta
- $\beta 1$= Koefisien varialbel X_1
- β_2 = Koefisien variabel X_2
- β_3 = Koefisien variabel X_3
- X_1 = pendapatan
- X_2 = lama pinjaman
- X_3 = riwayat kredit

Variabel metode ini memanfaatkan fungsi sigmoid menghasilkan probabilitas antara 0 hingga 1 untuk menentukan apakah nasabah termasuk kategori *default* atau *non-default* pada credit scoring. Umumnya digunakan ambang batas 0,5, di mana nilai di atas 0,5 dianggap berisiko gagal bayar. Selain itu, Explainable AI (XAI) membantu menjelaskan faktor-faktor yang memengaruhi hasil prediksi pada model Logistic Regression dan XGBoost secara lebih transparan [11].

2.4 XGBoost

XGBoost atau Extreme Gradient Boosting merupakan algoritma machine learning berbasis pohon keputusan yang dikembangkan untuk meningkatkan akurasi dari metode Gradient Boosting [12]. Perhitungan probilitas menggunakan xgboost :

$$obj = \sum i = 1nl(yi, y\hat{}i(t)) + \sum k = 1t\Omega(fk)$$

Keterangan :

- l(yi, ŷi(t)) = fungsi loss/error antara nilai asli dan prediksi
- Ω(fk) = regularisasi untuk mengontrol kompleksitas model
- fk = decision tree ke-k
- n = jumlah data
- t = jumlah tree
- Obj = objective function pada XGBoost
- yi = nilai aktual/data asli
- ŷi(t) = hasil prediksi pada iterasi ke-t
- Σ = penjumlahan seluruh data/tree

XGBoost digunakan pada *credit scoring* karena mampu menghasilkan prediksi risiko gagal bayar dengan akurasi tinggi. Pada model ini, $l(yi, y\hat{}i(t))$ digunakan untuk mengukur kesalahan prediksi, sedangkan $\Omega(fk)$ berfungsi mengontrol kompleksitas model agar tidak terjadi *overfitting*. Namun, XGBoost memerlukan Explainable AI (XAI) agar hasil prediksi dapat dijelaskan secara lebih transparan [13].

2.5 Explainable Artificial Intelligence (XAI)

Explainable Artificial Intelligence (XAI) merupakan pendekatan yang bertujuan untuk membuat keputusan yang dihasilkan oleh sistem kecerdasan buatan menjadi lebih transparan dan mudah dipahami. Konsep ini muncul sebagai respons terhadap kebutuhan untuk menjelaskan proses yang mendasari pengambilan keputusan oleh model kecerdasan

buatan [14]. Pendekatan ini menjadi penting dalam penerapan credit scoring, khususnya pada model seperti XGBoost dan Logistic Regression, untuk meningkatkan interpretabilitas hasil prediksi.

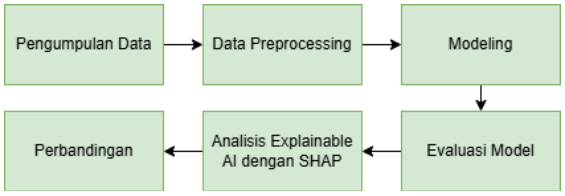
2.6 SHAP (SHapley Additive Explanations)

SHapley Additive Explanations (SHAP) merupakan metode interpretasi dalam Explainable AI yang berfungsi untuk menjelaskan kontribusi masing-masing fitur terhadap hasil prediksi model, baik secara global (keseluruhan model) maupun lokal (setiap observasi) [15].

3. METODE PENELITIAN

Penelitian ini bertujuan untuk melakukan analisis prediksi kelayakan kredit (credit scoring) dengan membandingkan algoritma Logistic Regression dan XGBoost sebagai metode klasifikasi. Data penelitian diperoleh dari dataset Credit Risk Dataset yang tersedia pada platform Kaggle, yang berisi informasi karakteristik individu dan pinjaman.

Tahapan penelitian ini meliputi pengumpulan data, data preprocessing, pembagian data (splitting data), pembangunan model (modeling), evaluasi model, analisis Explainable Artificial Intelligence (SHAP), serta perbandingan hasil model. Alur tahapan penelitian tersebut dapat dilihat pada Gambar 1.



Gambar 1. tahapan penelitian

3.1. Jenis dan Pendekatan Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksperimen komparatif yang bertujuan untuk membandingkan kinerja algoritma Logistic Regression dan Extreme Gradient Boosting (XGBoost) dalam prediksi credit scoring.

Pendekatan eksperimen dilakukan dengan membangun dan mengevaluasi kedua model menggunakan metrik klasifikasi untuk mengetahui performa terbaik dalam menentukan kelayakan kredit. Selain itu, penelitian ini juga menggunakan pendekatan Explainable Artificial Intelligence (XAI) dengan metode SHAP (SHapley Additive exPlanations) untuk menjelaskan kontribusi masing-masing fitur terhadap hasil prediksi model

3.2. Sumber dan Jenis Data

Penelitian ini menggunakan data sekunder yang diperoleh dari Credit Risk Dataset yang tersedia secara publik pada platform Kaggle (https://www.kaggle.com/datasets/laotse/credit-risk-dataset). Dataset ini merepresentasikan data historis permohonan kredit nasabah yang mencakup informasi demografis, keuangan, dan riwayat kredit yang banyak digunakan dalam penelitian machine learning di bidang keuangan.

Dataset terdiri atas 32.581 observasi dengan 12 variabel, yaitu 11 variabel independen dan 1 variabel dependen berupa loan_status yang merepresentasikan status kelayakan kredit (0 = good credit, 1 = default). Dari 11 variabel independen tersebut, 7 variabel bertipe numerik meliputi person_age, person_income, person_emp_length, loan_amnt, loan_int_rate, loan_percent_income, dan cb_person_cred_hist_length, sedangkan 4 variabel bertipe kategorikal meliputi person_home_ownership, loan_intent, loan_grade, dan cb_person_default_on_file. Berdasarkan kombinasi tipe data numerik dan kategorikal tersebut, diperlukan tahapan preprocessing sebelum data dapat digunakan dalam proses pemodelan. Seluruh variabel yang digunakan disajikan pada Tabel 1.

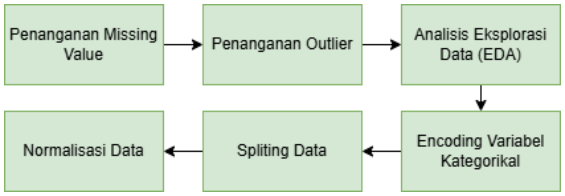
Tabel 1. Variabel Dataset

Variabel	Keterangan	Tipe
person_age	Usia peminjam (tahun)	Numerik
person_income	Pendapatan tahunan	Numerik
person_emp_length	Lama masa kerja (tahun)	Numerik
loan_amnt	Jumlah pinjaman yang diajukan	Numerik

loan_int_rate	Suku bunga pinjaman (%)	Numerik
loan_percent_income	Rasio pinjaman terhadap pendapatan	Numerik
cb_person_cred_hist_length	Panjang riwayat kredit (tahun)	Numerik
person_home_ownership	Status kepemilikan rumah	Kategorikal
loan_intent	Tujuan penggunaan pinjaman	Kategorikal
loan_grade	Peringkat kredit pinjaman	Kategorikal
cb_person_default_on_file	Riwayat gagal bayar sebelumnya	Kategorikal
loan_status (target)	Status pinjaman: 0 = Good Credit, 1 = Default	Biner

3.3. Data Preprocessing

Tahap preprocessing bertujuan mempersiapkan data agar siap digunakan dalam pemodelan. Tahapan yang dilakukan meliputi penanganan missing value, penanganan outlier, analisis eksplorasi data, encoding variabel kategorikal, pembagian data, dan normalisasi fitur. Alur lengkap tahapan preprocessing disajikan pada Gambar 2.



Gambar 2. Tahapan Data Preprocessing

3.3.1 Penanganan Missing Value

Pemeriksaan missing value dilakukan pada seluruh variabel. Variabel yang memiliki nilai hilang diimputasi menggunakan nilai median dari distribusi masing-masing variabel. Metode imputasi median dipilih karena lebih robust terhadap distribusi yang tidak normal (skewed) dan keberadaan outlier dibandingkan imputasi mean, sehingga tidak menggeser ukuran pemusatan data secara berlebihan.

3.3.2 Penanganan Outlier

Identifikasi outlier dilakukan menggunakan visualisasi boxplot pada variabel person_age dan person_emp_length. Penghapusan baris dilakukan terhadap data dengan person_age lebih dari 100 tahun dan person_emp_length lebih dari 60 tahun. Kedua nilai tersebut dianggap tidak realistis secara substantif dalam konteks profil peminjam aktif, sehingga berpotensi mengakibatkan bias pada model jika tidak ditangani.

3.3.3 Analisis Eksplorasi Data (EDA)

Analisis eksplorasi data dilakukan untuk memahami karakteristik dan distribusi data sebelum pemodelan. EDA mencakup: (1) analisis distribusi variabel target untuk mengidentifikasi ketidakseimbangan kelas; (2) visualisasi distribusi tujuh variabel numerik menggunakan histogram; (3) analisis korelasi antar variabel numerik menggunakan heatmap; serta (4) analisis distribusi variabel kategorikal terhadap variabel target menggunakan stacked bar chart. Tahap ini bertujuan untuk memberikan pemahaman awal mengenai fitur-fitur yang berpotensi memiliki daya prediktif tinggi terhadap risiko kredit.

3.3.4 Encoding Variabel Kategorikal

Variabel kategorikal seperti person_home_ownership, loan_intent, loan_grade, dan cb_person_default_on_file dikonversi ke format numerik menggunakan One-Hot Encoding dengan parameter drop_first=True untuk menghindari dummy variable trap. Seluruh hasil encoding bertipe boolean kemudian diubah menjadi integer (0 dan 1) agar kompatibel dengan proses pemodelan.

3.3.5 Spliting Data

Dataset dibagi menjadi data latih sebesar 80% dan data uji sebesar 20% menggunakan fungsi train_test_split dengan parameter stratify=y. Stratifikasi dilakukan untuk memastikan proporsi kelas target pada data latih dan data uji tetap konsisten dengan distribusi pada dataset asal, sehingga evaluasi model tidak terpengaruh oleh distribusi kelas yang tidak representatif pada salah satu subset.

3.3.6 Normalisasi Data

Normalisasi data dilakukan menggunakan StandardScaler untuk mengubah fitur numerik agar memiliki rata-rata 0 dan standar deviasi 1. Fitting hanya dilakukan pada data latih (fit_transform), sedangkan data uji hanya ditransformasi (transform) guna mencegah data leakage. Proses ini penting terutama pada Logistic Regression yang sensitif terhadap perbedaan skala fitur.

3.4. Modeling

3.4.1 Logistic Regression

Model Logistic Regression dikonfigurasi menggunakan regularisasi L2 dengan parameter C=1.0, solver lbfgs, dan maksimum iterasi 1.000. Parameter class_weight='balanced' digunakan agar kelas minoritas (Default) memperoleh bobot lebih besar selama pelatihan.

Karena sensitif terhadap skala fitur, model dilatih menggunakan data yang telah dinormalisasi (X_train_scaled) dan diuji pada X_test_scaled. Cross-validation juga dilakukan pada data hasil normalisasi untuk menjaga konsistensi evaluasi. Model memprediksi probabilitas default menggunakan fungsi sigmoid yang menghasilkan nilai pada rentang 0–1.

3.4.2 XGBoost

Model XGBoost dikonfigurasi dengan 300 estimator, max_depth=6, learning_rate=0,05, subsample=0,8, dan colsample_bytree=0,8 untuk membantu mengurangi overfitting. Ketidakseimbangan kelas ditangani menggunakan parameter scale_pos_weight yang dihitung dari rasio jumlah kelas negatif dan positif. Berbeda dengan Logistic Regression, XGBoost tidak sensitif terhadap skala fitur sehingga pelatihan dan evaluasi dilakukan menggunakan data tanpa normalisasi (X_train dan X_test). Model menggunakan fungsi objektif binary:logistic untuk menghasilkan probabilitas klasifikasi biner.

3.4.3 Cross-Validation

Evaluasi awal model dilakukan menggunakan Stratified K-Fold Cross-Validation dengan `k=5` dan `shuffle=True` agar distribusi kelas pada setiap fold tetap proporsional. Metrik evaluasi yang digunakan adalah F1 Macro karena mempertimbangkan performa seluruh kelas secara seimbang. Cross-validation Logistic Regression dilakukan pada `X_train_scaled`, sedangkan XGBoost menggunakan `X_train` tanpa normalisasi.

3.5. Evaluasi Model

Evaluasi kinerja model dilakukan pada data uji menggunakan beberapa metrik, yaitu Accuracy, Precision, Recall, F1-Score, ROC AUC, dan Specificity. Penggunaan beberapa metrik diperlukan karena data bersifat tidak seimbang, sehingga evaluasi tidak cukup hanya menggunakan Accuracy.

Evaluasi juga dilengkapi dengan visualisasi Confusion Matrix untuk masing-masing model serta ROC Curve untuk membandingkan performa kedua model. Dalam konteks credit scoring, Recall menjadi metrik yang diprioritaskan karena kesalahan False Negative, yaitu gagal mendeteksi nasabah yang berpotensi default, dapat menimbulkan risiko finansial yang lebih besar bagi lembaga keuangan.

3.6. Analisis Explainable AI dengan SHAP

Interpretabilitas model dianalisis menggunakan metode SHAP (SHapley Additive exPlanations) untuk mengukur kontribusi setiap fitur terhadap hasil prediksi. Pada model XGBoost digunakan shap.TreeExplainer, sedangkan pada Logistic Regression digunakan shap.

Explainer dengan X_train_scaled sebagai background data. Analisis SHAP menghasilkan beberapa visualisasi, yaitu Force Plot, Summary Plot, Bar Plot, dan Decision Plot untuk menjelaskan pengaruh fitur baik secara lokal maupun global. Perbandingan nilai SHAP antar model dilakukan untuk mengidentifikasi fitur yang paling konsisten memengaruhi prediksi kredit.

4. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil eksperimen secara sistematis meliputi preprocessing data, pemodelan, evaluasi perbandingan model, dan analisis interpretabilitas SHAP.

4.1 Preprocessing Data

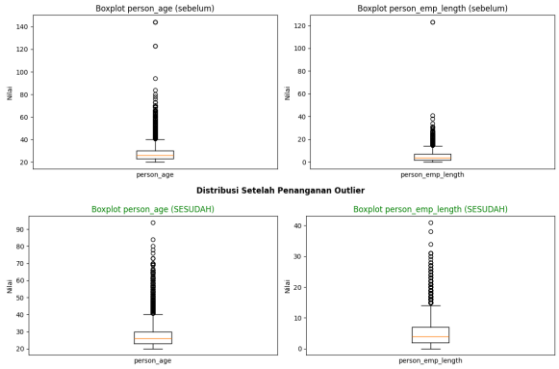
4.1.1 Penanganan Missing Value

Pemeriksaan awal menunjukkan bahwa variabel person_emp_length dan loan_int_rate

memiliki missing value. Kedua variabel diimputasi menggunakan nilai median masing-masing, sehingga tidak terdapat lagi nilai yang hilang pada seluruh kolom dataset.

4.1.2 Penanganan Outlier

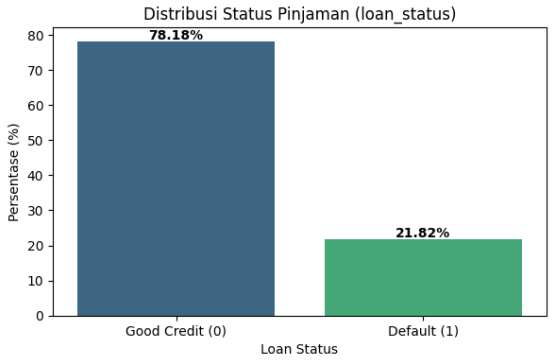
Identifikasi outlier menggunakan boxplot menghasilkan temuan nilai person_age > 100 tahun dan person_emp_length > 60 tahun yang tidak realistis. Baris-baris tersebut dihapus dari dataset. Perbandingan distribusi sebelum dan sesudah penanganan ditunjukkan pada Gambar 3.



Gambar 3. Perbandingan distribusi sebelum dan sesudah penanganan

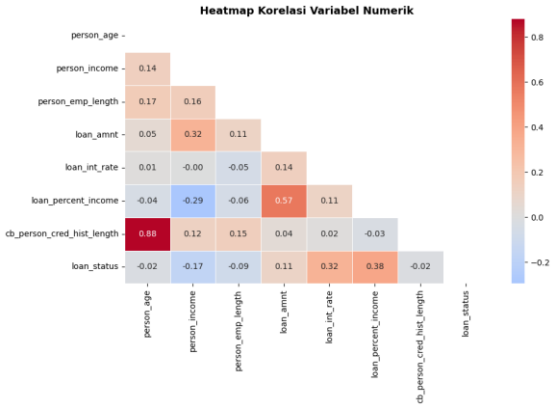
4.1.3 Analisis Eksplorasi Data (EDA)

Distribusi variabel target loan_status menunjukkan ketidakseimbangan kelas yang signifikan antara Good Credit (0) dan Default (1), sebagaimana terlihat pada Gambar 4. Kondisi ini mengonfirmasi perlunya penanganan imbalance pada tahap pemodelan.



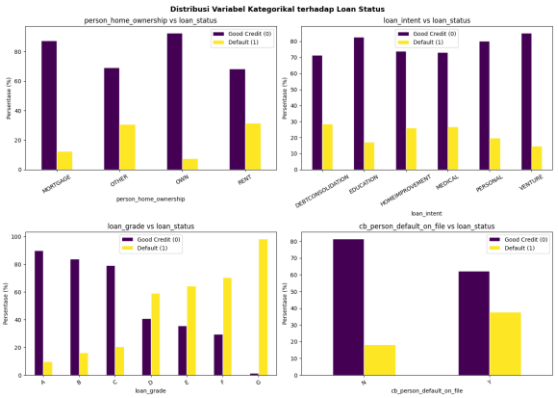
Gambar 4. Distribusi status pinjaman

Heatmap korelasi pada Gambar 5 menunjukkan bahwa loan_percent_income dan loan_int_rate memiliki korelasi positif tertinggi terhadap loan_status dibandingkan variabel numerik lainnya, mengindikasikan keduanya sebagai fitur dengan potensi daya prediktif tertinggi.



Gambar 5. Heatmap korelasi

Analisis variabel kategorikal terhadap loan_status disajikan pada Gambar 6. Variabel loan_grade menunjukkan pola yang jelas: grade D hingga G memiliki proporsi Default jauh lebih tinggi dibandingkan grade A dan B. Variabel cb_person_default_on_file juga menunjukkan asosiasi kuat dengan kelas Default.



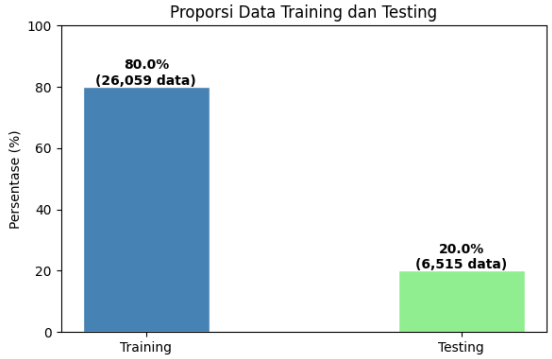
Gambar 6. Distribusi fitur kategorikal terhadap loan_status

4.1.4 Encoding Variabel Kategorikal

One-Hot Encoding dengan drop_first=True diterapkan pada empat variabel kategorikal (person_home_ownership, loan_intent, loan_grade, cb_person_default_on_file). Seluruh kolom boolean hasil encoding dikonversi ke integer (0 dan 1). Jumlah fitur bertambah dari 12 menjadi lebih banyak kolom setelah proses encoding selesai.

4.1.5 Spliting Data

Dataset dibagi menjadi data latih (80%) dan data uji (20%) dengan stratified splitting. Proporsi kelas pada data latih dan data uji tetap konsisten dengan distribusi dataset asal, sebagaimana divisualisasikan pada Gambar 7.



Gambar 7. proporsi split

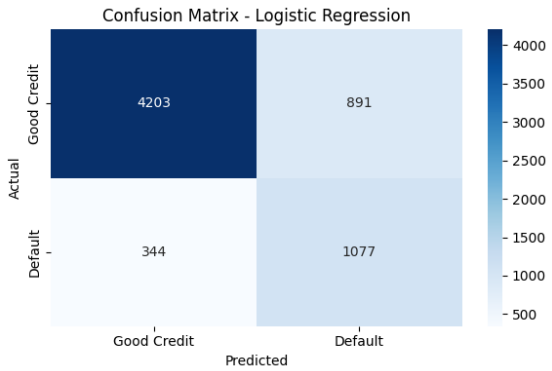
4.1.6 Normalisasi Data

StandardScaler di-fit pada data latih dan diterapkan pada data uji (transform only) untuk mencegah data leakage. Hasil normalisasi (X_train_scaled, X_test_scaled) digunakan khusus untuk Logistic Regression, sedangkan XGBoost menggunakan data tanpa normalisasi (X_train, X_test).

4.2 Modeling

4.2.1 Logistic Regression

Model Logistic Regression dilatih pada X_train_scaled. Cross-validation 5-fold Stratified pada X_train_scaled menghasilkan F1 Macro = 0.7588 ± 0.0063. Confusion Matrix hasil prediksi pada X_test_scaled disajikan pada Gambar 8.

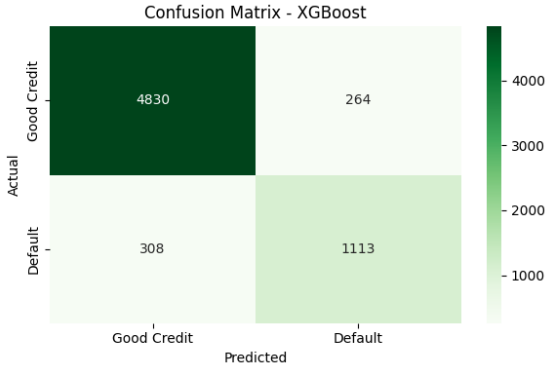


Gambar 8. Confusion Matrix Logistic Regression

4.2.2 XGBoost

Model XGBoost dilatih pada X_train dengan scale_pos_weight dihitung dari rasio kelas negatif terhadap positif pada data latih. Cross-validation 5-fold Stratified pada X_train menghasilkan F1 Macro = 0.8778 ± 0.0046. Confusion Matrix hasil prediksi pada X_test disajikan pada Gambar 9.

Gambar 9. Confusion Matrix Logistic Regression



4.3 Evaluasi Model

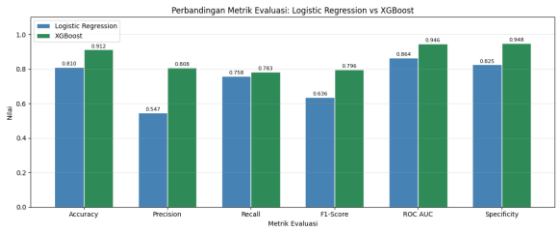
Perbandingan kinerja kedua model pada data uji menggunakan enam metrik klasifikasi disajikan pada Tabel 2.

Tabel 2. perbandingan metrik

Metrik	Logistic Regression	XGBoost
Accuracy	0.8104	0.9122
Precision	0.5473	0.8083
Recall (Sensitivity)	0.7579	0.7833
F1-Score	0.6356	0.7956
ROC AUC	0.8635	0.9458
Specificity	0.8251	0.9482

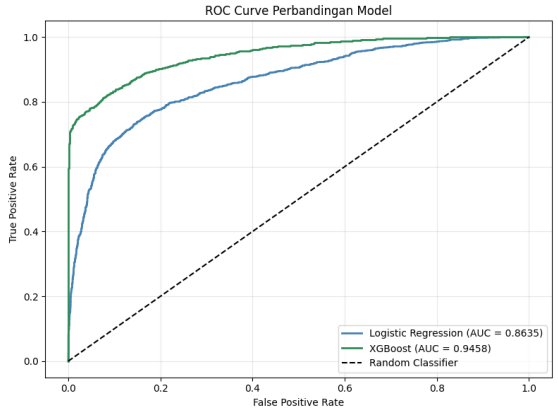
XGBoost unggul pada seluruh metrik dibandingkan Logistic Regression. Nilai ROC AUC dan Recall XGBoost yang lebih tinggi menunjukkan kemampuan diskriminasi dan deteksi default yang lebih baik. Dalam konteks credit scoring, Recall menjadi prioritas karena False Negative — gagal mendeteksi nasabah

yang sesungguhnya default — menimbulkan kerugian finansial lebih besar dibandingkan False Positive. Perbandingan visual keenam metrik disajikan pada Gambar 10.



Gambar 10. perbandingan metrik

ROC Curve pada Gambar 11 menunjukkan kurva XGBoost berada lebih dekat ke sudut kiri atas, mengonfirmasi True Positive Rate yang lebih tinggi pada seluruh threshold klasifikasi.



Gambar 11. ROC Curve Perbandingan

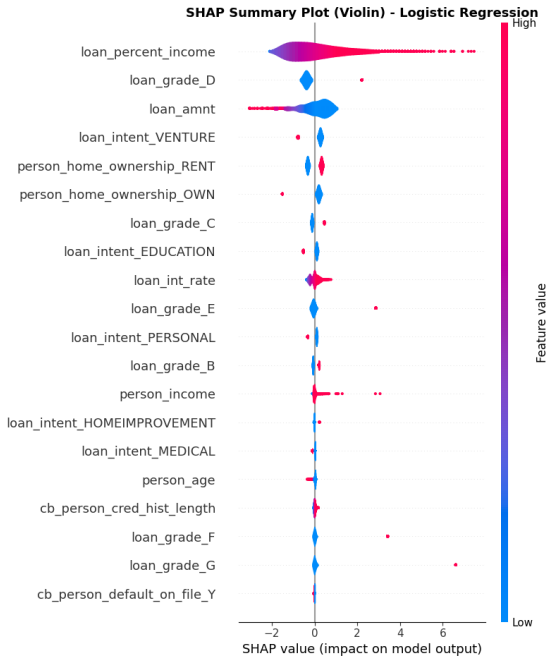
4.4 Analisis SHAP

4.4.1 SHAP Logistic Regression

SHAP dihitung menggunakan shap.Explainer dengan X_train_scaled sebagai background data dan X_test_scaled sebagai input. Force Plot pada Gambar 12 menampilkan kontribusi fitur terhadap prediksi satu observasi pada model Logistic Regression.

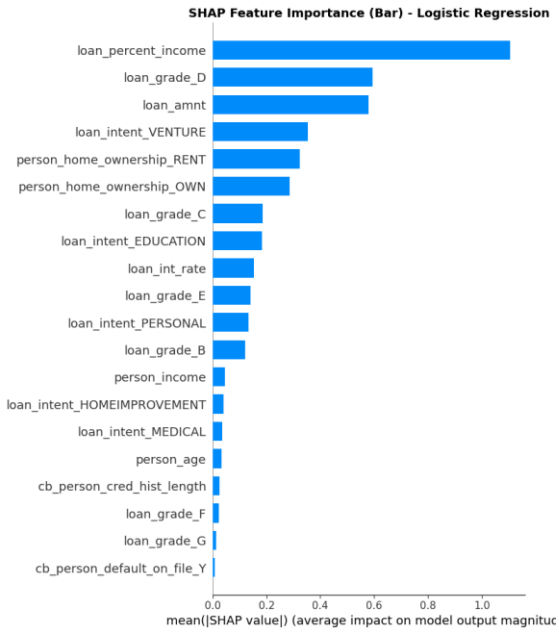
Gambar 12. Force Plot Logistic Regression

Summary Plot violin pada Gambar 13 menunjukkan distribusi nilai SHAP global Logistic Regression. Dibandingkan XGBoost, distribusi nilai SHAP lebih simetris karena kontribusi fitur bersifat aditif murni tanpa interaksi non-linear.



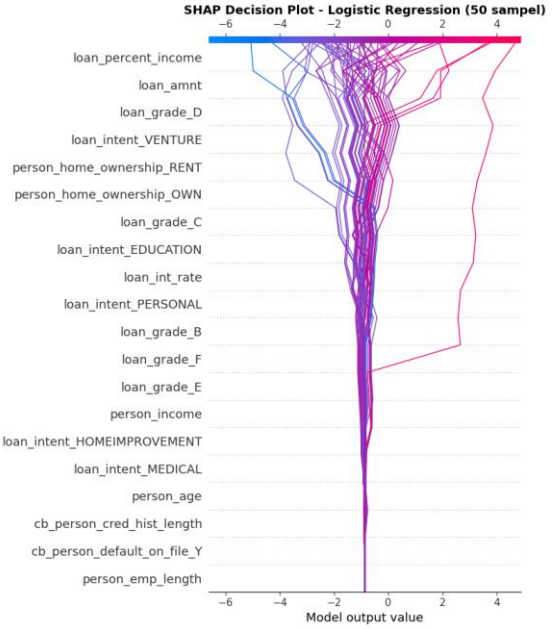
Gambar 13. Summary Plot Logistic Regression

Bar Plot pada Gambar 14 menunjukkan peringkat fitur berdasarkan rata-rata SHAP value| pada Logistic Regression. Pola peringkat serupa dengan XGBoost, dengan loan_percent_income dan loan_int_rate kembali mendominasi.



Gambar 14. Bar Plot Logistic Regression

Decision Plot pada Gambar 15 menampilkan jalur keputusan model Logistic Regression untuk 50 sampel pertama data uji. Setiap garis merepresentasikan jalur prediksi satu observasi dari nilai ekspektasi menuju prediksi akhir. Pola jalur yang serupa pada sebagian besar sampel mengonfirmasi dominasi loan_percent_income dan loan_int_rate secara global.



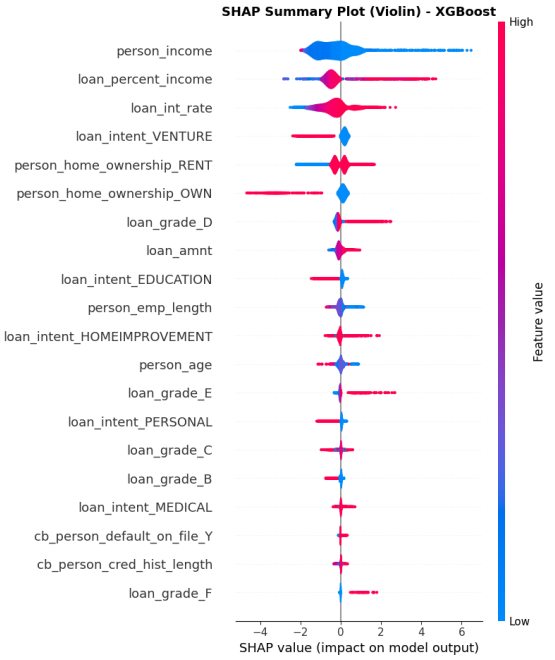
Gambar 15. Decision Plot - Logistic Regression

4.4.2 SHAP XGBoost

SHAP dihitung menggunakan shap.TreeExplainer pada X_test_f (X_test.astype(float)). Force Plot pada Gambar 16 menampilkan kontribusi fitur terhadap prediksi satu observasi secara lokal: fitur berwarna merah mendorong ke arah Default, fitur biru menahan ke arah Good Credit.

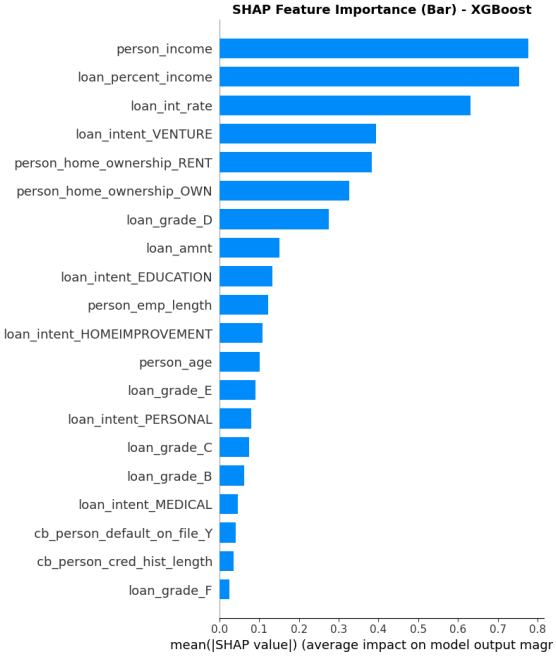
Gambar 16. Force Plot XGBoost

Summary Plot violin pada Gambar 17 menampilkan distribusi nilai SHAP seluruh fitur secara global pada model XGBoost. Fitur dengan sebaran nilai SHAP paling lebar adalah fitur dengan pengaruh terbesar terhadap prediksi.



Gambar 17. Summary Plot XGBoost

Bar Plot pada Gambar 18 menunjukkan peringkat fitur berdasarkan rata-rata |SHAP value|. Fitur loan_percent_income dan loan_int_rate mendominasi peringkat teratas, konsisten dengan temuan korelasi pada tahap EDA.



Gambar 18. Bar Plot XGBoost

4.5 Perbandingan Interpretabilitas SHAP Kedua Model

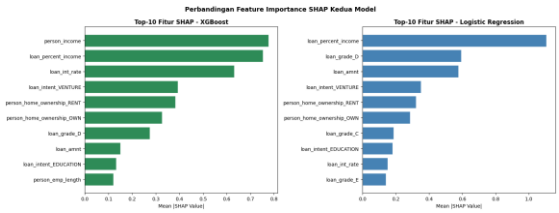
Gambar 19 menyajikan perbandingan peringkat Top-5 fitur berdasarkan rata-rata SHAP value pada kedua model. Fitur yang konsisten masuk Top-5 di kedua model diidentifikasi sebagai fitur kunci prediksi risiko kredit.

=== Perbandingan Ranking Fitur SHAP ===

Feature	Rank_XGBoost	SHAP_XGBoost	Rank_LR	SHAP_LR
person_income	1	0.7777	13	0.0460
loan_percent_income	2	0.7532	1	1.1054
loan_int_rate	3	0.6320	9	0.1532
loan_intent_VENTURE	4	0.3933	4	0.3532
person_home_ownership_RENT	5	0.3837	5	0.3237
person_home_ownership_OWN	6	0.3270	6	0.2863
loan_grade_D	7	0.2757	2	0.5933
loan_amnt	8	0.1504	3	0.5782
loan_intent_EDUCATION	9	0.1325	8	0.1827
person_emp_length	10	0.1218	21	0.0036
loan_intent_HOMEIMPROVEMENT	11	0.1077	14	0.0410
person_age	12	0.1016	16	0.0326
loan_grade_E	13	0.0914	10	0.1418
loan_intent_PERSONAL	14	0.0800	11	0.1320
loan_grade_C	15	0.0743	7	0.1869
loan_grade_B	16	0.0626	12	0.1215
loan_intent_MEDICAL	17	0.0470	15	0.0364
cb_person_default_on_file_Y	18	0.0415	20	0.0077
cb_person_cred_hist_length	19	0.0352	17	0.0242
loan_grade_F	20	0.0260	18	0.0220
loan_grade_G	21	0.0159	19	0.0132
person_home_ownership_OTHER	22	0.0010	22	0.0004

Gambar 19. Perbandingan Ranking Fitur SHAP

Visualisasi perbandingan Top-10 fitur SHAP kedua model disajikan pada Gambar 20. Fitur loan_percent_income menduduki peringkat pertama di kedua model, menegaskan bahwa rasio beban pinjaman terhadap pendapatan merupakan indikator risiko kredit paling informatif, selaras dengan konsep Debt-to-Income Ratio dalam analisis kredit konvensional.



Gambar 20. perbandingan Top-10 fitur SHAP kedua model

XGBoost memberikan nilai SHAP yang lebih besar pada fitur kategorikal seperti loan_grade dan cb_person_default_on_file_Y karena mampu menangkap interaksi non-linear antar fitur. Pada Logistic Regression, kontribusi fitur-fitur tersebut bersifat aditif sehingga magnitude-nya lebih kecil. Meskipun demikian, konsistensi peringkat fitur di kedua model menunjukkan bahwa SHAP berhasil mengungkap pola risiko kredit yang intrinsik, terlepas dari kompleksitas algoritma. Keunggulan prediktif XGBoost tidak mengorbankan interpretabilitas karena SHAP tetap mampu menjelaskan mekanisme keputusannya secara transparan.

5. KESIMPULAN

Berdasarkan hasil penelitian, model XGBoost menunjukkan kinerja yang lebih baik dibandingkan Logistic Regression dalam tugas credit scoring. XGBoost memperoleh nilai Accuracy, Precision, F1-Score, ROC AUC, dan Specificity yang lebih tinggi, serta menunjukkan performa yang lebih stabil berdasarkan hasil cross-validation. Temuan ini mengindikasikan bahwa XGBoost lebih efektif dalam menangkap pola kompleks pada data kredit sehingga mampu menghasilkan prediksi risiko kredit yang lebih akurat dan konsisten dibandingkan model linier tradisional seperti Logistic Regression.

Analisis Explainable Artificial Intelligence (XAI) menggunakan SHAP berhasil memberikan interpretasi yang transparan terhadap keputusan kedua model. Hasil analisis menunjukkan bahwa fitur loan_percent_income dan loan_int_rate merupakan faktor yang paling berpengaruh dalam menentukan risiko kredit, sejalan dengan konsep Debt-to-Income Ratio yang umum digunakan dalam penilaian kelayakan kredit. Selain itu, ditemukan bahwa XGBoost lebih mampu menangkap hubungan non-linear dan interaksi antar fitur, sedangkan Logistic Regression menghasilkan pola kontribusi fitur yang lebih sederhana dan linier. Meskipun demikian, kedua model menunjukkan konsistensi dalam mengidentifikasi fitur-fitur utama yang memengaruhi risiko kredit.

Secara keseluruhan, penelitian ini membuktikan bahwa kombinasi XGBoost dan SHAP mampu memberikan keseimbangan antara akurasi prediksi dan transparansi model, sehingga berpotensi diterapkan dalam sistem credit scoring yang membutuhkan keputusan yang akurat sekaligus dapat dijelaskan. Untuk penelitian selanjutnya, disarankan melakukan optimasi hyperparameter, membandingkan SHAP dengan metode XAI lainnya.

UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada Program Studi Informatika Universitas Mulawarman yang telah memberikan fasilitas dan dukungan yang diperlukan selama penelitian ini berlangsung. Terima kasih juga disampaikan kepada penyedia Credit Risk Dataset melalui platform Kaggle yang telah membuat data tersebut tersedia secara publik sehingga dapat digunakan dalam penelitian ini.

DAFTAR PUSTAKA

[1] A. R. Hakim, M. A. Mukid, H. Yasin, and S. Sugito, "Analisis Klasifikasi Credit Scoring Menggunakan Weighted Probabilistic Neural Network (Wpnn)," *J. Stat.*, vol. 7, no. 1, pp. 63–70, 2019, [Online]. Available: https://jurnal.unimus.ac.id/index.php/statistik/article/view/4807

[2] H. D. Fata, G. I. Marthasari, and Y. Azhar, "Sistem Pendukung Keputusan Kelayakan Kredit pada PT. BPR Mitra Catur Mandiri Menggunakan Metode Credit Scoring," *J. Repos.*, vol. 2, no. 5, pp. 649–658, 2020, doi: 10.22219/repositor.v2i5.608.

[3] A. Asro, J. Chaidir, C. Cahairuddin, and J. Friadi, "Evaluasi Kinerja Algoritma Klasifikasi dalam Studi Kasus Prediksi Kelulusan di Universitas XYZ," *Zo. Tek. J. Ilm.*, vol. 19, no. 1, pp. 15–22, 2025, doi: 10.37776/zt.v19i1.1674.

[4] S. Mujiyono, U. P. Sanjaya, I. S. Wibisono, and H. Setyowati, "Prediksi Fluktuasi Berat Badan Berdasarkan Pola Hidup Menggunakan Model XGBoost dan Deep Learning," *J.* Algoritm., vol. 22, no. 1, pp. 221–233, 2025, doi: 10.33364/algoritma/v.22-1.2253.

[5] H. Sy and S. Alam, "Explainable Artificial Intelligence (XAI) Pada Time Series Forecasting Menggunakan Gated Recurrent Unit (GRU)," vol. XV, no. 1, pp. 50–62, doi: 10.36774/sisiti.v15i1.2134.

[6] H. Putra and R. Rumini, "Comparative Study of Logistic Regression, Random Forest, and XGBoost for Bank Loan Approval Classification," *J. Appl. Informatics Comput.*, vol. 9, no. 5, pp. 2822–2835, 2025, doi: 10.30871/jaic.v9i5.10862.

[7] S. E. Herni Yulianti, Oni Soesanto, and Yuana Sukmawaty, "Gani, A. G., Firdaus, I., Amien, J. A., Informatika, T., Komputer, F. I., Riau, U. M., & Matching, S. (2020). Jurnal Computer Science and Information Technology (CoSciTech). Jurnal Sistem Informasi Universitas Suryadarma, 3(2), 1–19.," *J. Math. Theory Appl.*, vol. 4, no. 1, pp. 21–26, 2022.

[8] I. W. Permadani, R. Sulistyo, M. Fadli, and E. R. Susanto, "Prediksi Risiko Kredit Nasabah Menggunakan Algoritma Data Mining: Studi Kasus pada PT Toyota Astra Finance," *Progresif J. Ilm. Komput.*, vol. 21, no. 2, p. 952, 2025, doi: 10.35889/progresif.v21i2.2909.

[9] M. I. M. Nur Jamal, "Implementasi Machine Learning Berbasis Fitur Warna Rgb Dan Hsv Untuk Klasifikasi Kualitas Air," J. Inform. dan Tek. Elektro Terap., vol. 14, no. 1, 2026, doi: 10.23960/jitet.v14i1.8667.

[10] Sutarman, R. Siringoringo, D. Arisandi, E. Kurniawan, and E. B. Nababan, "Model Klasifikasi Dengan Logistic Regression Dan Recursive Feature Elimination Pada Data Tidak Seimbang," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 4, pp. 735–742, 2024, doi: 10.25126/jtiik.1148198.

[11] J. Garcke and R. Roscher, "Explainable Machine Learning," *Mach. Learn. Knowl. Extr.*, vol. 5, no. 1, pp. 169–170, 2023, doi: 10.3390/make5010010.

[12] DHIKA MALITA PUSPITA ARUM, KARTIKA IMAM SANTOSO, ANDRI TRIYONO, EKO SUPRIYADI, AGUS SUSILO NUGROHO, and E. Widodo, "Algoritma Random Forest, Decision Tree, Dan Xgboost Untuk Klasifikasi Stunting Pada Balita," *J. Transform.*, vol. 23, no. 1, pp. 67–76, 2025, doi: 10.26623/transformatika.v23i1.12202.

[13] D. Azura, P. Reksi, B. Kurniawan, and S. Susandri, "Model Prediksi Penjualan Berbasis XGBoost – SHAP untuk Decision Support dalam Arsitektur TOGAF Prosiding Semnas 2025 Sekolah Tinggi Teknologi Dumai," vol. 1, no. 2, pp. 343–357, 2025.

[14] M. Emhandyksa and I. Nur, "Metode Explainable Artificial Intelligence (XAI) Berbasis Feature Importance Untuk Deteksi Dini Penyakit Jantung," vol. 1, no. 1, pp. 68–79, 2026.

[15] M. Kristanaya, Melinda Putri Azzahra, Trimono, and Mohammad Idhom, "Klasifikasi Status Rujukan Pasien Poliklinik Bandara Berbasis Random Forest dan Interpretabilitas Model Menggunakan SHAP," *Data Sci. Indones.*, vol. 5, no. 1, pp. 60–74, 2025, doi: 10.47709/dsi.v5i1.6078.