

Prediksi Risiko Gagal Bayar Pinjaman Menggunakan Xgboost dengan Pendekatan CRISP-DM Berbasis Web

Rendareista Nurillahi Reyzzah¹, Asmunin²

Manajemen Informatika, Universitas Negeri Surabaya
Jl. Ketintang, Ketintang, Kec. Gayungan, Kota Surabaya, Jawa Timur 60231

renda.12.22108@mhs.unesa.ac.id

asmunin@unesa.ac.id

Abstrak - Pertumbuhan layanan kredit berbasis teknologi di Indonesia telah memperluas jangkauan pembiayaan, namun sekaligus memunculkan tantangan serius berupa peningkatan risiko gagal bayar. Metode penilaian kredit konvensional dinilai kurang mampu mendeteksi pola tersembunyi dalam data peminjam yang bersifat kompleks dan nonlinier. Penelitian ini bertujuan mengembangkan sistem prediksi risiko gagal bayar menggunakan algoritma *eXtreme Gradient boosting* (Xgboost) dalam kerangka metodologi CRISP-DM, yang kemudian diintegrasikan ke dalam aplikasi berbasis web. Dataset *Give Me Some Credit* (GMSC) dari Kaggle digunakan sebagai data uji dengan 150.000 observasi. Tahap persiapan data mencakup penanganan nilai hilang menggunakan *Iterative Imputer*, rekayasa fitur, dan penanganan ketidakseimbangan kelas melalui pendekatan *cost-sensitive learning*. Evaluasi model dilakukan menggunakan *confusion matrix* serta metrik kinerja meliputi akurasi, presisi, recall, F1-score, dan AUC. Model Xgboost yang dihasilkan mencatatkan akurasi 80%, presisi 21%, recall 78%, F1-score 33%, dan AUC 86%, mengungguli *Logistic regression* sebagai model pembanding. Sistem berhasil diimplementasikan sebagai aplikasi web fungsional yang mampu menampilkan hasil prediksi secara interaktif disertai laporan analisis risiko kredit.

Kata kunci: Risiko Gagal Bayar, Machine Learning, Xgboost, CRISP-DM, Kredit Digital

Abstrak - The rapid expansion of technology-driven credit services in Indonesia has broadened access to financing while simultaneously intensifying the risk of borrower default. Traditional credit assessment approaches have demonstrated limited capacity to detect intricate, nonlinear patterns within borrower data. This study presents the development of a default risk prediction system employing the *eXtreme Gradient boosting* (Xgboost) algorithm within a CRISP-DM methodological framework, subsequently deployed as a web-based application. The *Give Me Some Credit* (GMSC) dataset from Kaggle, comprising 150,000 observations, was used for model training and evaluation. Data preparation involved handling missing values using *Iterative Imputer*, feature engineering, and addressing class imbalance through *cost-sensitive learning*. Model performance was assessed using a *confusion matrix* alongside accuracy, precision, recall, F1-score, and AUC metrics. The Xgboost model achieved accuracy of 80%, precision of 21%, recall of 78%, F1-score of 33%, and AUC of 86%, outperforming *Logistic regression* as the baseline comparator. The system was successfully integrated into a functional web application capable of interactively presenting prediction outcomes and credit risk analysis reports.

Keywords: Default Risk, Machine Learning, Xgboost, CRISP-DM, Digital Credit

I. PENDAHULUAN

Sektor keuangan digital Indonesia mengalami perkembangan pesat, terutama pada segmen kredit berbasis platform teknologi seperti *peer-to-peer* (P2P) lending dan lembaga multifinansial. Perluasan akses pembiayaan ini membawa dampak ganda: di satu sisi memperluas inklusi keuangan, namun di sisi lain memperbesar eksposur terhadap risiko gagal bayar. Otoritas Jasa Keuangan (OJK) mencatat bahwa total penyaluran pinjaman P2P lending mencapai Rp 81,17 triliun pada Februari 2025, dengan laju pertumbuhan tahunan sebesar 31,06% dan tingkat wanprestasi 90 hari (TWP90) sebesar 2,78% [1]. Tren ini mengindikasikan bahwa ekspansi kredit yang berkelanjutan harus diimbangi dengan mekanisme penilaian risiko yang lebih andal.

Pendekatan konvensional dalam analisis risiko kredit umumnya mengandalkan variabel sederhana seperti tingkat pendapatan, jenis pekerjaan, dan usia peminjam. Variabel-variabel tersebut tidak cukup memadai untuk menangkap relasi nonlinier dan interaksi multivariabel yang kerap terdapat dalam data keuangan [2]. Bukti empiris menunjukkan bahwa kenaikan rasio kredit bermasalah berpengaruh negatif terhadap profitabilitas perbankan, khususnya pada indikator *Return on Assets* (ROA), dan berhubungan dengan memburuknya kondisi makroekonomi pada periode tertentu [2], [3]. Temuan-temuan ini menegaskan urgensi adopsi metode analitik yang lebih adaptif dan berbasis data.

Kemajuan *machine learning* telah membuka peluang bagi pengembangan sistem prediksi risiko kredit yang lebih akurat. Di antara berbagai algoritma yang tersedia, *eXtreme Gradient boosting* (Xgboost) menonjol sebagai salah satu pendekatan yang paling andal untuk permasalahan klasifikasi kredit. Xgboost bekerja dengan membangun ansambel model pohon keputusan secara bertahap, di mana setiap iterasi difokuskan untuk memperbaiki kesalahan prediksi dari iterasi sebelumnya [4]. Mekanisme regularisasi internal yang dimiliki Xgboost memberikan keuntungan tambahan berupa pengurangan risiko *overfitting*, sehingga performa model tetap stabil pada data yang belum pernah dilihat sebelumnya.

Sejumlah studi terdahulu mendukung keunggulan Xgboost dalam domain prediksi risiko keuangan. Odegua [5] dan

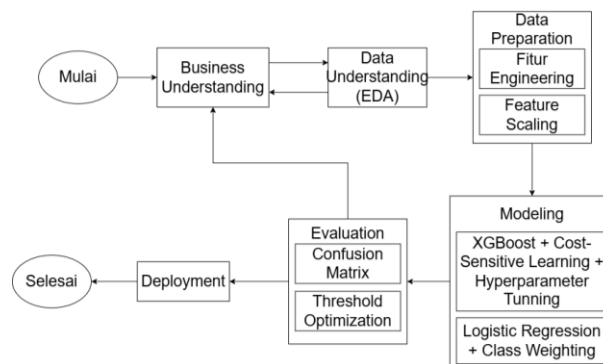
Omogbhemhe & Momodu [6] melaporkan bahwa *Xgboost* mampu mengklasifikasikan risiko gagal bayar pinjaman perbankan dengan akurasi 97%, presisi 87%, *recall* 79%, dan F1-score 79%. Sementara itu, Zhang et al. [7] membandingkan empat algoritma *LightGBM*, *Xgboost*, *Random forest*, dan *Gradient boosting* dan menemukan bahwa *LightGBM* memperoleh nilai *ROC-AUC* tertinggi sebesar 0,9714, meskipun *Xgboost* tetap menunjukkan keseimbangan yang kompetitif antara presisi dan *recall*. Munkhdalai et al. [8] juga mengonfirmasi bahwa model berbasis *gradient boosting* secara konsisten melampaui sejumlah algoritma *machine learning* lainnya pada tugas *credit scoring*.

Selain pemilihan algoritma, keberhasilan suatu sistem prediktif sangat ditentukan oleh metodologi pengembangan yang diterapkan. Penelitian ini mengadopsi *CRISP-DM* (*Cross-Industry Standard Process for Data Mining*) sebagai kerangka kerja, mengingat kemampuannya menyediakan alur proses yang sistematis dan terstruktur mulai dari pemahaman bisnis, eksplorasi data, persiapan data, pemodelan, hingga evaluasi dan implementasi [9]. Beberapa penelitian sebelumnya telah membuktikan efektivitas *CRISP-DM* dalam konteks prediksi risiko kredit, antara lain Saputra et al. [9] yang menggunakan *Random forest* dan Sugiyarto et al. [10] yang menerapkan *Decision tree* dengan akurasi 86,7%. Akan tetapi, kajian-kajian tersebut belum mengintegrasikan *CRISP-DM* dengan *Xgboost*, dan hanya sedikit yang mengimplementasikan hasil model ke dalam antarmuka web interaktif suatu fitur yang sangat relevan untuk mendukung pengambilan keputusan secara langsung di lingkungan lembaga keuangan.

Berdasarkan identifikasi kesenjangan tersebut, penelitian ini mengusulkan suatu sistem prediksi risiko gagal bayar yang menggabungkan kekuatan algoritma *Xgboost* dengan pendekatan metodologis *CRISP-DM*, sekaligus mengimplementasikannya dalam bentuk aplikasi berbasis web. Dataset Give Me Some Credit (GMSC) dari platform Kaggle dipilih sebagai sumber data, mengingat representasinya terhadap histori kredit nyata dengan jumlah observasi yang substansial. Penelitian ini diharapkan memberikan kontribusi ganda: secara teoritis memperkaya literatur tentang prediksi kredit berbasis *machine learning* di Indonesia, dan secara praktis menghadirkan prototipe alat bantu pengambilan keputusan yang dapat diadopsi oleh institusi keuangan digital.

II. METODE PENELITIAN

Penelitian ini menggunakan kerangka *CRISP-DM* (*Cross-Industry Standard Process for Data Mining*) yang terdiri dari enam tahapan berurutan: *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*. Kerangka ini dipilih karena menyediakan alur pengembangan model yang terstruktur, reproduktibel, dan relevan secara praktis bagi institusi keuangan [11]. Kerangka ini juga memungkinkan proses pengembangan bersifat iteratif, artinya temuan pada tahap tertentu dapat mendorong evaluasi ulang terhadap tahap sebelumnya tanpa harus mengulang seluruh proses dari awal. Gambar 1 mengilustrasikan keseluruhan alur penelitian dari tahap awal hingga implementasi sistem.



Gambar. 1 Alur penelitian

A. Business Understanding

Tahap ini bertujuan memetakan permasalahan bisnis yang menjadi landasan penelitian. Sistem yang dikembangkan dirancang untuk membantu proses evaluasi kredit secara objektif dan berbasis data, menggantikan penilaian subjektif yang selama ini banyak diterapkan. Variabel target yang diprediksi adalah *SeriousDlqin2yrs*, yang merepresentasikan kejadian keterlambatan pembayaran lebih dari 90 hari dalam kurun dua tahun (1 = gagal bayar, 0 = tidak gagal bayar). Output model berupa probabilitas risiko dalam rentang 0–100%, yang diklasifikasikan menggunakan *threshold* 0,5 sebagai ambang standar pada classifier probabilistik [12].

Kebutuhan bisnis yang diidentifikasi mencakup kemampuan sistem dalam memproses data input nasabah secara *real-time*, menghasilkan klasifikasi risiko otomatis, serta menyajikan faktor penyebab risiko secara interpretatif. Pendekatan ini bertujuan mentransformasi proses evaluasi kredit dari yang bersifat manual menuju pengambilan keputusan berbasis data (*data-driven decision making*).

B. Data Understanding

Dataset yang digunakan adalah Give Me Some Credit (GMSC) yang bersumber dari platform Kaggle. Dataset ini terdiri dari 150.000 observasi dengan 11 variabel yang mencakup aspek demografis, kondisi keuangan, dan riwayat keterlambatan pembayaran peminjam. Seluruh variabel bertipe numerik sehingga tidak diperlukan proses encoding kategorikal.

Eksplorasi awal melalui *Exploratory Data Analysis* (EDA) mengungkap sejumlah karakteristik penting data. Distribusi variabel target menunjukkan ketidakseimbangan kelas yang signifikan, yakni sekitar 93,3% data termasuk kategori tidak gagal bayar dan hanya 6,7% berkategori gagal bayar. Analisis univariat menunjukkan bahwa variabel *MonthlyIncome* dan *DebtRatio* memiliki distribusi yang sangat miring ke kanan (*positively skewed*) dengan nilai ekstrem yang besar. Analisis bivariat mengungkap bahwa kelompok gagal bayar cenderung memiliki *DebtRatio* lebih tinggi, pendapatan lebih rendah, serta frekuensi keterlambatan yang jauh lebih tinggi dibandingkan kelompok tidak gagal bayar.

Selain itu, ditemukan nilai-nilai tidak realistis pada variabel keterlambatan, seperti angka 96 dan 98 yang secara logika tidak dapat merepresentasikan frekuensi keterlambatan pembayaran yang wajar. Variabel *MonthlyIncome* memiliki

29.731 nilai hilang (sekitar 19,8%) dan *NumberOfDependents* memiliki 3.924 nilai hilang, yang keduanya perlu ditangani sebelum pemodelan.

C. Data Preparation

Tahap persiapan data dilakukan secara berurutan untuk memastikan kualitas data sebelum masuk ke proses pemodelan. Seluruh langkah preprocessing diterapkan hanya pada *training set*, dan transformasi yang sama kemudian diterapkan pada *validation set* guna mencegah *data leakage*.

Pertama, dataset dibagi dengan proporsi 80% untuk pelatihan dan 20% untuk validasi menggunakan *stratified sampling* agar distribusi kelas target tetap terjaga secara proporsional pada kedua subset. Kedua, penanganan outlier dilakukan dengan mengganti nilai keterlambatan yang melebihi 90 dengan nilai median dari training set, mengingat nilai seperti 96 dan 98 tidak mencerminkan kondisi kredit yang logis.

Ketiga, nilai hilang pada *MonthlyIncome* dan *NumberOfDependents* ditangani menggunakan Iterative Imputer, yang merupakan implementasi dari pendekatan *Multiple Imputation by Chained Equations (MICE)*. Metode ini dipilih karena mampu memperkirakan nilai yang hilang berdasarkan hubungan antarvariabel dalam dataset, sehingga menghasilkan estimasi yang lebih realistis dibandingkan pengisian sederhana menggunakan nilai mean atau median [13].

Keempat, rekayasa fitur dilakukan untuk menghasilkan empat variabel turunan yang memperkaya representasi perilaku kredit nasabah: (1) *DebtIncome*, yaitu hasil perkalian *DebtRatio* dengan *MonthlyIncome* sebagai proksi beban utang absolut; (2) *IncomePerDependent*, rasio pendapatan terhadap jumlah tanggungan sebagai ukuran kapasitas finansial per kepala; (3) *TotalPastDue*, akumulasi seluruh riwayat keterlambatan dari ketiga rentang waktu yang tersedia; dan (4) *LateRatio*, perbandingan *TotalPastDue* terhadap jumlah fasilitas kredit aktif.

Kelima, transformasi logaritma (\log_{1p}) diterapkan pada variabel *MonthlyIncome*, *DebtRatio*, dan *RevolvingUtilizationOfUnsecuredLines* yang memiliki distribusi *heavy-tailed*. Transformasi ini menstabilkan distribusi data dan mereduksi dampak nilai ekstrem terhadap proses pembelajaran model. Khusus untuk model *Logistic regression*, standarisasi tambahan menggunakan *StandardScaler (Z-score normalization)* diterapkan karena algoritma berbasis regresi sensitif terhadap perbedaan skala antarvariabel.

Keenam, ketidakseimbangan kelas ditangani melalui pendekatan *cost-sensitive learning*. Pada *Xgboost*, parameter *scale_pos_weight* ditetapkan sebagai rasio jumlah kelas negatif terhadap kelas positif, sehingga model memberikan penalti lebih besar pada kesalahan prediksi kelas minoritas. Pada *Logistic regression*, parameter *class_weight='balanced'* digunakan untuk tujuan serupa.

D. Modeling

Dua algoritma diimplementasikan dalam penelitian ini. *Xgboost* berfungsi sebagai model utama, sedangkan *Logistic*

regression digunakan sebagai model pembandingan (*baseline*). Pemilihan *Xgboost* didasarkan pada kemampuannya menangani hubungan nonlinier antarvariabel, mekanisme regularisasi internal L1 dan L2 yang mengurangi risiko *overfitting*, serta efisiensi komputasi yang tinggi pada dataset berskala besar [14].

Xgboost membangun model secara iteratif dengan menambahkan pohon keputusan baru pada setiap iterasi, di mana setiap pohon difokuskan untuk memperbaiki residual error dari model sebelumnya. Fungsi objektif yang diminimalkan pada iterasi ke- t adalah kombinasi dari fungsi loss dan fungsi regularisasi:

$$Obj(\theta) = \sum_{i=1}^n l(\hat{y}_i, y_i) + \sum_{k=1}^k \Omega(f_k)$$

di mana $l(\hat{y}_i, y_i)$ merupakan fungsi loss antara nilai prediksi dan label aktual, sedangkan $\Omega(f_k)$ adalah fungsi regularisasi yang mengontrol kompleksitas setiap pohon [14]. Pemilihan split pada tiap node didasarkan pada nilai gradient dan hessian dari fungsi loss, bukan sekadar ukuran impurity seperti pada *Decision tree* konvensional, sehingga model mampu mengenali pola nonlinier yang kompleks.

Logistic regression bekerja dengan mengestimasi probabilitas gagal bayar menggunakan fungsi sigmoid terhadap kombinasi linear variabel input. Model ini menghasilkan koefisien yang mudah diinterpretasikan, sehingga bermanfaat sebagai tolok ukur awal sebelum menerapkan model yang lebih kompleks. Perbedaan mendasar antara keduanya adalah bahwa *Xgboost* membangun model secara bertahap dan mampu menangkap interaksi kompleks antarvariabel, sementara *Logistic regression* mengasumsikan hubungan yang bersifat linear.

Hyperparameter *Xgboost* dioptimalkan menggunakan *GridSearchCV* dengan metrik optimasi *roc_auc*. Parameter yang ditelusuri meliputi *learning_rate*, *max_depth*, *n_estimators*, *subsample*, dan *colsample_bytree*. Validasi silang dengan $k=5$ diterapkan untuk meminimalkan bias estimasi performa. Tabel I merangkum parameter utama yang dikonfigurasi beserta rentang nilainya.

TABEL I
PARAMETER DAN RENTANG NILAI STANDARD

Parameter	Fungsi	Rentang Nilai
<i>learning_rate</i>	Kecepatan pembelajaran model	0,01 – 0,3
<i>max_depth</i>	Kedalaman maksimum pohon keputusan	3 – 7
<i>n_estimators</i>	Jumlah pohon yang dibangun	100 – 500
<i>subsample</i>	Proporsi sampel tiap iterasi	0,7 – 1,0
<i>colsample_bytree</i>	Proporsi fitur tiap pohon	0,7 – 1,0

E. Evaluation

Kinerja model dievaluasi pada *validation set* menggunakan lima metrik: akurasi, presisi, *recall*, *F1-score*, dan *AUC-ROC*. Dalam konteks prediksi risiko kredit dengan data yang tidak seimbang, *recall* dan *AUC-ROC* menjadi indikator utama. *Recall* mengukur kemampuan model mendeteksi seluruh nasabah yang benar-benar gagal bayar, sehingga meminimalkan *false negative* yang secara finansial paling merugikan institusi pemberi kredit. *AUC-ROC* mengukur kemampuan diskriminatif model secara keseluruhan tanpa bergantung pada *threshold* tertentu, sehingga lebih representatif untuk data imbalanced [12].

Threshold klasifikasi ditetapkan pada nilai default 0,5, yakni standar umum pada *classifier probabilistik*. Probabilitas prediksi $\geq 0,5$ dikategorikan sebagai gagal bayar, sedangkan di bawah nilai tersebut dikategorikan tidak gagal bayar. Penggunaan *threshold* standar dipilih untuk menjaga objektivitas evaluasi dan memudahkan reproduksi pada implementasi sistem. Model dinyatakan berhasil apabila memenuhi kriteria: $AUC > 0,7$, *recall* kelas gagal bayar $> 0,7$, dan *Xgboost* mengungguli *Logistic regression* pada keseluruhan metrik seperti pada Tabel II dibawah ini.

TABEL II
EVALUASI METRIK

Metrik	Deskripsi	Fokus Utama
<i>Accuracy</i>	Proporsi prediksi benar secara keseluruhan	Gambaran umum
<i>Precision</i>	Proporsi prediksi gagal bayar yang benar-benar gagal bayar	Ketepatan prediksi positif
<i>Recall</i>	Proporsi nasabah gagal bayar yang berhasil terdeteksi	Minimasi <i>false negative</i>
<i>F1-Score</i>	Rata-rata harmonik <i>precision</i> dan <i>recall</i>	Keseimbangan keduanya
<i>AUC-ROC</i>	Kemampuan diskriminatif antar kelas pada berbagai <i>threshold</i>	Data imbalanced

F. Deployment

Model *Xgboost* terbaik hasil pelatihan disimpan menggunakan library *joblib* dalam format file yang dapat dimuat ulang tanpa perlu proses training ulang. Model kemudian diintegrasikan ke dalam aplikasi berbasis web menggunakan *framework* Flask. Pada sisi backend, data input nasabah yang dikirimkan melalui formulir web diproses oleh model untuk menghasilkan prediksi secara *real-time*. Hasil prediksi berupa probabilitas risiko dan kategori risiko kemudian ditampilkan pada antarmuka frontend yang dibangun dengan HTML dan CSS.

Untuk mendukung interpretabilitas hasil prediksi, sistem mengintegrasikan *SHAP (SHapley Additive exPlanations)* yang menghitung kontribusi masing-masing fitur terhadap probabilitas risiko yang dihasilkan. Informasi ini ditampilkan dalam laporan analisis kredit yang dapat dicetak. Sistem juga dilengkapi dengan penyimpanan riwayat prediksi menggunakan MySQL, sehingga pengguna dapat melacak dan

meninjau hasil analisis sebelumnya. Pengujian sistem mencakup *functional testing*, *performance testing*, dan *validation testing* untuk memastikan konsistensi dan keandalan sistem sebelum dioperasikan.

III. HASIL DAN PEMBAHASAN

Bagian ini menyajikan temuan dari setiap tahapan *CRISP-DM* yang telah dijalankan, mulai dari persiapan data hingga implementasi sistem. Pembahasan difokuskan pada hasil yang secara langsung menjawab pertanyaan penelitian: seberapa efektif *Xgboost* dalam memprediksi risiko gagal bayar, dan bagaimana model tersebut dapat dioperasikan melalui antarmuka web.

A. Hasil Persiapan Data

Proses pembagian dataset menghasilkan 120.000 sampel untuk pelatihan dan 30.000 sampel untuk validasi, dengan distribusi kelas yang terjaga secara proporsional pada kedua subset berkat *stratified sampling*. Proporsi gagal bayar pada *training set* dan *validation set* masing-masing berada di kisaran 6,7%, konsisten dengan distribusi awal dataset sebagaimana diperlihatkan pada Gambar. 2.

```
y = train['SeriousDlqin2yrs']
X = train.drop(columns=['SeriousDlqin2yrs'])

# 2. SPLIT DATA
X_train, X_valid, y_train, y_valid = train_test_split(
    X, y, test_size=0.2, stratify=y, random_state=42
)
```

Gambar. 2 Pemisahan data 80% *training* dan 20% *validasi* sebelum dilakukan *preprocessing*

Penanganan *outlier* pada tiga variabel keterlambatan berhasil menghapus nilai tidak logis seperti 96 dan 98. Sebelum penanganan, nilai-nilai tersebut tercatat pada 264 observasi untuk masing-masing variabel. Setelah substitusi menggunakan median *training set*, distribusi ketiga variabel menjadi lebih stabil dan tidak lagi didominasi oleh nilai ekstrem yang menyimpang dari logika kredit seperti pada Gambar. 3.

```
late_cols = [
    'NumberOfTime30-59DaysPastDueNotWorse',
    'NumberOfTime60-89DaysPastDueNotWorse',
    'NumberOfTimes90DaysLate'
]

for c in late_cols:
    med = X_train[c].median()
    X_train.loc[X_train[c] > 90, c] = med
    X_valid.loc[X_valid[c] > 90, c] = med
```

Gambar. 3 Penanganan *outlier* dengan mengganti ke nilai median

Imputasi menggunakan *Iterative Imputer (MICE)* berhasil mengisi 29.731 nilai hilang pada *MonthlyIncome* dan 3.924 nilai hilang pada *NumberOfDependents*. Metode ini menghasilkan estimasi nilai berdasarkan hubungan antarvariabel, sehingga lebih representatif dibandingkan pengisian sederhana. Fitting imputer dilakukan eksklusif pada *training set*, dan transformasi yang sama diterapkan pada *validation set* untuk menghindari kebocoran data (*data leakage*) seperti yang dilakukan pada Gambar. 4.

```
# Imputer (MonthlyIncome & Dependents) menggunakan MICE
imp = IterativeImputer(max_iter=10, random_state=42)
X_train = pd.DataFrame(imp.fit_transform(X_train), columns=X_train.columns)
X_valid = pd.DataFrame(imp.transform(X_valid), columns=X_valid.columns)
```

Gambar. 4 Penanganan nilai hilang menggunakan Iterative Imputer (MICE)

Rekayasa fitur menghasilkan empat variabel turunan yang memperkaya kapasitas prediktif model. *DebtIncome* dan *IncomePerDependent* merepresentasikan beban dan kapasitas finansial nasabah secara lebih konkret, sementara *TotalPastDue* dan *LateRatio* mengkodekan intensitas perilaku menunggak relatif terhadap jumlah fasilitas kredit aktif. Transformasi log1p pada *MonthlyIncome*, *DebtRatio*, dan *RevolvingUtilizationOfUnsecuredLines* berhasil mereduksi *skewness* yang sebelumnya mencapai nilai di atas 95, menjadikan distribusi jauh lebih simetris dan layak untuk pemodelan seperti yang diperlihatkan pada Gambar. 5.

```
# Feature Engineering
for df in [X_train, X_valid]:
    df.clip(lower=0, inplace=True)

# 1. Debt absolute
df['DebtIncome'] = df['DebtRatio'] * df['MonthlyIncome']
# 2. Income per dependent
df['IncomePerDependent'] = df['MonthlyIncome'] / (df['NumberOfDependents'] + 1)
# 3. Total past due
df['TotalPastDue'] = df['NumberOfTime30-59DaysPastDueNotWorse'] + \
    df['NumberOfTime60-89DaysPastDueNotWorse'] + \
    df['NumberOfTimes90DaysLate']
# 4. Late ratio
df['LateRatio'] = df['TotalPastDue'] / (df['NumberOfOpenCreditLinesAndLoans'] + 1)
```

Gambar. 5 Feature Engineering untuk memperkaya kapasitas prediktif model

B. Hasil Pemodelan

Hyperparameter tuning menggunakan *GridSearchCV* dengan 5-fold *stratified cross-validation* menghasilkan konfigurasi optimal untuk *Xgboost* sebagaimana ditunjukkan pada Tabel III. Kombinasi *learning_rate* = 0,1 dan *max_depth* = 4 memberikan keseimbangan yang baik antara kompleksitas model dan kemampuan generalisasi, sementara nilai *subsample* = 0,8 dan *colsample_bytree* = 0,6 berperan dalam mencegah *overfitting* melalui mekanisme pengacakan data dan fitur pada setiap iterasi pembangunan pohon.

TABEL III
HASIL HYPERPARAMETER TUNING TERBAIK

Parameter	Nilai Terbaik
<i>Learning Rate</i>	0,10
<i>Max Depth</i>	4
<i>N_Estimators</i>	100
<i>Subsample</i>	0,80
<i>Colsample by Tree</i>	0,60
<i>Scale Pos Weight</i>	~13,96

Nilai *scale_pos_weight* sebesar ~13,96 diperoleh dari rasio jumlah observasi kelas negatif (tidak gagal bayar) terhadap kelas positif (gagal bayar) pada *training set*, yang mencerminkan tingkat ketidakseimbangan kelas sekitar 1:14. Parameter ini memastikan model memberikan penalti lebih besar ketika salah mengklasifikasikan nasabah gagal bayar, sehingga sensitivitas terhadap kelas minoritas meningkat secara signifikan. Konfigurasi terbaik ini selanjutnya digunakan sebagai model final yang dievaluasi pada *validation set* dan diintegrasikan ke dalam sistem web.

C. Hasil Evaluasi Model

Evaluasi dilakukan pada *validation set* menggunakan *threshold* klasifikasi default sebesar 0,5. Tabel IV menyajikan perbandingan performa antara *Xgboost* dan *Logistic regression* pada seluruh metrik yang digunakan.

TABEL IV
PERBANDINGAN PERFORMA MODEL

Model	Accuracy	Precision	Recall	F1-Score	AUC
<i>Logistic regression</i>	0.85	0.26	0.62	0.37	0.82
<i>Xgboost</i>	0.80	0.21	0.78	0.33	0.86

Hasil pada Tabel IV mengungkap pola yang menarik: meskipun *Logistic regression* mencatat akurasi dan presisi yang lebih tinggi, *Xgboost* justru unggul pada dua metrik yang paling kritis dalam konteks deteksi risiko kredit, yakni *recall* (0,78 vs 0,62) dan AUC (0,86 vs 0,82). Perbedaan ini tidak bisa dipandang sebelah mata.

Dalam manajemen risiko kredit, *false negative* kondisi ketika nasabah yang sesungguhnya berisiko tinggi lolos dari deteksi model berpotensi menimbulkan kerugian finansial yang jauh lebih besar dibandingkan *false positive*. *Recall Xgboost* yang mencapai 78% berarti model berhasil mengidentifikasi hampir empat dari lima nasabah yang benar-benar gagal bayar, sementara *Logistic regression* hanya mampu mendeteksi sekitar enam dari sepuluh kasus serupa. Peningkatan *recall* sebesar 16 poin persentase ini memiliki implikasi praktis yang signifikan bagi institusi pemberi kredit.

AUC-ROC Xgboost sebesar 0,86 menunjukkan bahwa model memiliki kemampuan diskriminatif yang baik dalam membedakan kelas berisiko dan tidak berisiko di berbagai titik *threshold*. Nilai ini melampaui batas minimum 0,70 yang umumnya dijadikan tolok ukur model yang layak digunakan dalam praktik keuangan [12]. Temuan ini sejalan dengan studi

Zhang dkk. [7] yang melaporkan *Xgboost* sebagai model dengan *AUC-ROC* tertinggi di antara beberapa algoritma yang dibandingkan, serta konsisten dengan hasil Munkhdalai dkk. [8] yang menunjukkan keunggulan metode *gradient boosting* pada permasalahan *credit scoring*.

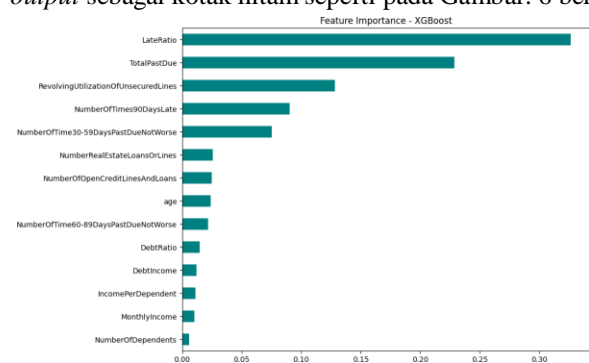
Nilai akurasi *Xgboost* yang sedikit lebih rendah (80% vs 85%) merupakan konsekuensi logis dari peningkatan *recall*. Ketika model dipaksa lebih sensitif terhadap kelas minoritas melalui *cost-sensitive learning*, sebagian prediksi untuk kelas mayoritas akan bergeser ke kelas minoritas, sehingga akurasi keseluruhan sedikit menurun namun kemampuan deteksi risiko meningkat. Pada dataset *imbalanced* seperti GMSC, *trade-off* semacam ini justru mencerminkan model yang berperilaku lebih tepat secara praktis dibandingkan model yang sekadar mengejar akurasi tinggi.

D. Analisis Feature Importance

Analisis *feature importance* pada model *Xgboost* mengungkap bahwa variabel turunan hasil rekayasa fitur mendominasi kontribusi prediksi. *LateRatio* dan *TotalPastDue* menempati posisi dua teratas, diikuti *NumberOfTimes90DaysLate* dan

RevolvingUtilizationOfUnsecuredLines. Temuan ini memvalidasi relevansi rekayasa fitur yang dilakukan: riwayat keterlambatan yang direpresentasikan secara lebih komprehensif ternyata memberikan sinyal prediktif yang lebih kuat dibandingkan variabel aslinya secara individual.

Untuk mendukung interpretabilitas hasil prediksi pada level individual, sistem mengintegrasikan *SHAP (SHapley Additive exPlanations)* yang menghitung kontribusi tiap fitur secara spesifik per observasi. Pada setiap laporan prediksi, sistem menampilkan lima faktor risiko teratas beserta arah kontribusinya: nilai SHAP positif menunjukkan peningkatan probabilitas gagal bayar, sementara nilai negatif menunjukkan penurunan risiko. Pendekatan *explainable AI* ini memungkinkan pengguna memahami mengapa model menghasilkan skor risiko tertentu, bukan sekadar menerima *output* sebagai kotak hitam seperti pada Gambar. 6 berikut.

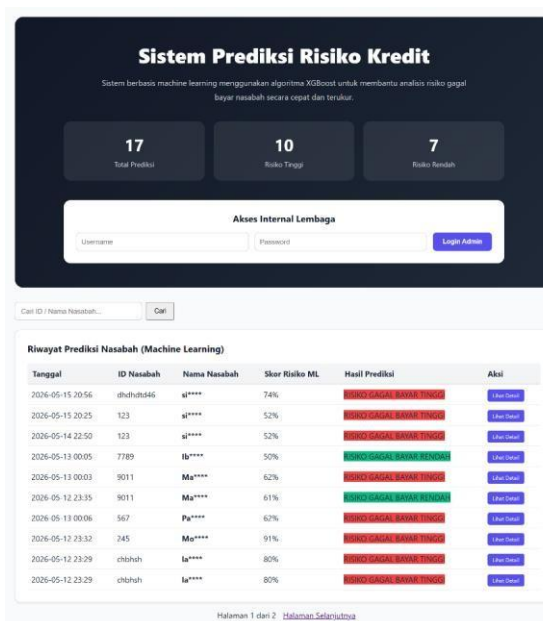


Gambar. 6 Feature Importance pada Xgboost

Temuan ini konsisten dengan literatur yang menekankan pentingnya interpretabilitas model dalam pengambilan keputusan kredit [7]. Nasabah dengan *LateRatio* tinggi dan *TotalPastDue* besar secara konsisten mendapatkan probabilitas risiko yang lebih tinggi, yang selaras dengan logika penilaian kredit konvensional bahwa riwayat pembayaran merupakan indikator risiko yang paling relevan.

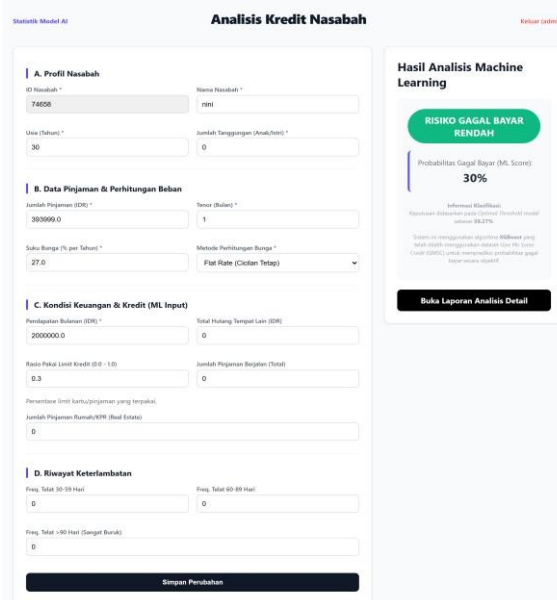
E. Implementasi Sistem Web

Model *Xgboost* terbaik berhasil diintegrasikan ke dalam aplikasi web berbasis Flask yang dapat diakses melalui browser. Sistem terdiri dari beberapa modul fungsional utama yang bekerja secara terintegrasi. Pertama, *landing page* publik menampilkan statistik keseluruhan sistem total prediksi, jumlah risiko tinggi, dan risiko rendah tanpa mengekspos data sensitif nasabah. Identitas nasabah disamarkan untuk pengguna yang belum terautentikasi, sehingga privasi data tetap terjaga. Halaman ini juga memuat formulir login yang terintegrasi langsung pada *landing page*, sehingga pengguna internal tidak perlu berpindah halaman untuk melakukan autentikasi. Desain ini sengaja dirancang agar antarmuka terasa ringkas dan tidak membingungkan, baik bagi pengunjung umum maupun petugas lembaga keuangan yang menggunakannya secara rutin. Tampilan *landing page* beserta formulir login internal ditunjukkan pada Gambar 7. Setelah autentikasi berhasil, sistem mengarahkan pengguna ke halaman analisis utama yang menampilkan panel formulir input di sisi kiri dan ringkasan hasil prediksi di sisi kanan.



Gambar. 7 Login beserta Landing Page

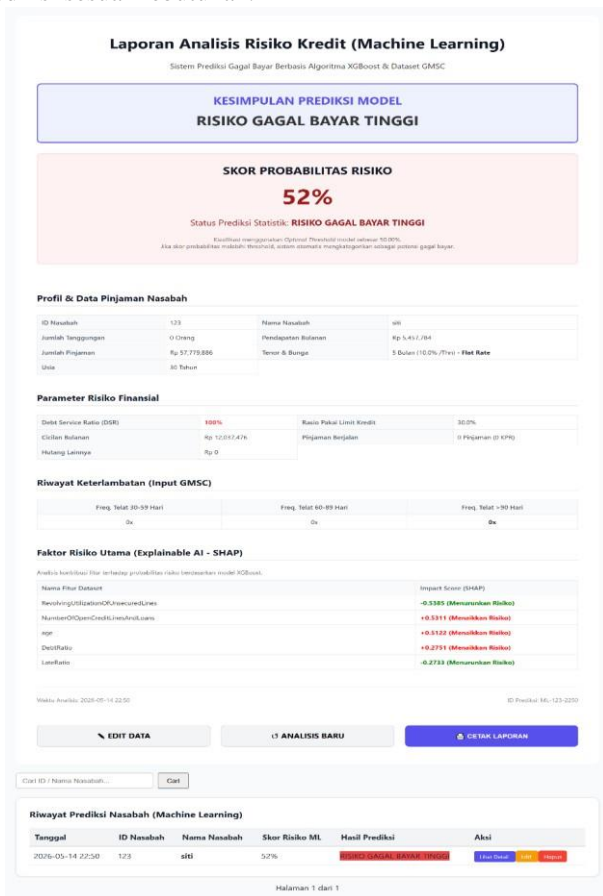
Kedua, setelah melewati proses autentikasi, pengguna internal dapat mengakses formulir input yang terstruktur dalam empat kelompok: profil nasabah, data pinjaman dan perhitungan beban cicilan, kondisi keuangan dan kredit, serta riwayat keterlambatan pembayaran. Data yang dikirimkan melalui formulir diproses secara *real-time* oleh backend Flask, yang menerapkan *pipeline preprocessing* yang identik dengan tahap pelatihan model sebelum memanggil fungsi *predict_proba()* dari model *Xgboost* yang terlihat pada Gambar. 8.



Gambar. 8 Form input pinjaman

Ketiga, hasil prediksi ditampilkan dalam laporan analisis kredit yang memuat probabilitas risiko, klasifikasi risiko berdasarkan *threshold*, ringkasan profil dan parameter keuangan nasabah, serta lima faktor risiko teratas dari analisis SHAP. Laporan ini dapat dicetak langsung melalui browser

seperti pada Gambar.9. Kemudian, sistem menyimpan setiap hasil analisis ke database MySQL secara otomatis, sehingga pengguna dapat meninjau, mengedit, atau menghapus riwayat prediksi sesuai kebutuhan.



Gambar. 9 Laporan analisis kredit

Selanjutnya hasil *black box testing* pada Tabel V menunjukkan bahwa seluruh fitur utama sistem berjalan sesuai spesifikasi yang ditetapkan. Waktu respons prediksi berada di bawah 3 detik untuk setiap skenario pengujian, memenuhi kebutuhan non-fungsional yang telah ditetapkan pada tahap perancangan. Pengujian juga mencakup simulasi berbagai profil nasabah dengan karakteristik risiko yang berbeda, dan sistem secara konsisten menghasilkan klasifikasi yang sesuai dengan evaluasi model pada *validation set*.

TABEL V
PENGUJIAN BLACK BOX TESTING

Fitur	Skenario Pengujian	Hasil yang Diharapkan
Login	Memasukkan kredensial yang valid	Masuk ke dashboard
Input Profil	Mengisi data nasabah lengkap	Data tersimpan di session
Prediksi AI	Klik tombol Analisis Kelayakan	Probabilitas dan badge risiko muncul
Laporan Detail	Klik tombol Lihat Laporan Detail	Halaman laporan lengkap tampil

Simpan Database	Setelah prediksi selesai	Data muncul di tabel riwayat
Pencarian Riwayat	Input ID/nama pada kolom cari	Data yang relevan tampil
Hapus Riwayat	Klik hapus dan konfirmasi	Data terhapus dari DB dan tabel

IV. KESIMPULAN

Penelitian ini berhasil membangun sistem prediksi risiko gagal bayar berbasis *machine learning* yang mengintegrasikan algoritma *Xgboost* dengan kerangka kerja *CRISP-DM*, lalu mengimplementasikannya sebagai aplikasi web yang dapat dioperasikan secara langsung. Dari rangkaian proses yang telah dijalankan, beberapa hal dapat disimpulkan.

Pertama, penerapan *CRISP-DM* terbukti memberikan alur pengembangan yang teratur dan dapat direplikasi. Setiap fase dari pemahaman masalah bisnis, eksplorasi data, pembersihan dan rekayasa fitur, pelatihan model, evaluasi, hingga *deployment* berjalan secara berkesinambungan dan saling mendukung. Kerangka ini sangat membantu dalam memastikan bahwa keputusan teknis yang diambil di setiap tahap tetap berakar pada kebutuhan nyata lembaga keuangan.

Kedua, *Xgboost* menunjukkan performa yang lebih unggul dibandingkan *Logistic regression* sebagai model *baseline*, terutama pada dua metrik yang paling relevan dalam konteks kredit: *recall* sebesar 78% dan *AUC-ROC* sebesar 86%. Keunggulan ini tidak lepas dari sejumlah teknik pendukung yang diterapkan secara sistematis mulai dari imputasi berbasis MICE, rekayasa fitur yang berangkat dari logika keuangan, transformasi distribusi, hingga *cost-sensitive learning* melalui parameter *scale_pos_weight*. Kombinasi ini secara kolektif mendorong model untuk lebih peka terhadap nasabah berisiko tinggi yang jumlahnya minoritas dalam dataset.

Ketiga, integrasi SHAP ke dalam sistem memberikan nilai tambah yang signifikan. Laporan prediksi tidak hanya menampilkan angka probabilitas, tetapi juga menjelaskan faktor-faktor spesifik yang mendorong atau menekan risiko setiap nasabah. Fitur *LateRatio* dan *TotalPastDue* muncul sebagai kontributor risiko paling dominan, yang konsisten dengan prinsip penilaian kredit bahwa riwayat pembayaran adalah cerminan paling jujur dari perilaku keuangan seseorang. Keempat, sistem web yang dibangun berhasil menjembatani model *machine learning* dengan kebutuhan operasional pengguna. Fitur input terstruktur, laporan yang dapat dicetak, manajemen riwayat berbasis MySQL, serta dashboard feature importance menjadikan sistem ini bukan sekadar prototipe akademis, melainkan alat bantu pengambilan keputusan yang memiliki potensi nyata untuk diadopsi di lingkungan lembaga keuangan skala kecil hingga menengah.

V. SARAN

Penelitian ini masih menyisakan beberapa ruang untuk dikembangkan lebih lanjut. Dari sisi pemodelan, perbandingan dengan algoritma *gradient boosting* lain seperti *LightGBM* dan *CatBoost* layak untuk dijajaki. Keduanya dikenal memiliki efisiensi komputasi dan performa yang kompetitif

pada dataset besar, sehingga eksplorasi ini dapat memperjelas posisi *Xgboost* secara lebih menyeluruh dalam konteks prediksi kredit. Selain itu, penerapan strategi penanganan *imbalanced data* alternatif seperti *SMOTE* atau kombinasi *oversampling-undersampling* juga menarik untuk dibandingkan secara sistematis dengan pendekatan *cost-sensitive* yang digunakan pada penelitian ini.

Dari sisi data, penggunaan dataset lokal dari lembaga keuangan Indonesia baik koperasi simpan pinjam, BPR, maupun platform *P2P lending* akan menghasilkan model yang lebih relevan secara kontekstual. Dataset GMSC yang digunakan saat ini berasal dari pasar kredit Amerika Serikat, sehingga pola perilaku keuangan yang terekam belum tentu sepenuhnya mencerminkan karakteristik peminjam di Indonesia.

Dari sisi sistem, ada beberapa aspek yang masih terbuka untuk ditingkatkan. Penambahan mekanisme keamanan yang lebih *robust* seperti enkripsi data, manajemen sesi yang lebih ketat, dan kontrol akses berbasis peran akan membuat sistem lebih layak digunakan pada lingkungan produksi. Integrasi dengan database skala enterprise dan pengujian beban dengan jumlah pengguna simultan yang lebih besar juga perlu dilakukan sebelum sistem ini benar-benar siap dioperasikan di luar lingkungan pengembangan.

REFERENSI

- [1] OJK, "Financing Institutions, Venture Capital, Fintech P2P Lending and Micro Finance Industry Update February 2025," *OJK International Information Hub*, 2025. <https://iru.ojk.go.id/iru/dataandstatistics/detaildataandstatistics/13661/financing-institutions-venture-capital-fintech-p2p-lending-and-micro-finance-industry-update-february-2025/>
- [2] W. Wahyuni, I. Badollahi, N. Nurhidayah, dan W. Mardiasuti, "Analyzing the Impact of Non-Performing Loans and Loan-to-Deposit Ratios on Return on Assets: A Study of Conventional Commercial Banks in Indonesia," *Adv. Manag. Financ. Report.*, vol. 1, no. 3, hal. 107–118, 2023, doi: 10.60079/amfr.v1i3.124.
- [3] S. U. Widodo dan P. M. Ramadhania, "Macroeconomic Indicators and Non-performing Loans and Banking Profitability Effect For Gross Domestic Product (Case Study Indonesia Economic and Banking)," *J. Tek. Ind.*, vol. 2, no. 2, hal. 57–69, 2024, doi: 10.30872/jatri.v2i2.1499.
- [4] BIS, "Asia-Pacific fixed income markets: evolving structure, participation and pricing," in *BIS Papers No 102*, 2019, hal. 163.
- [5] R. Odegua, "Predicting Bank Loan Default with Extreme Gradient boosting," no. 2015, 2018.
- [6] Omogbhemhe MI dan Momodu I.B.A., "Model for Predicting Bank Loan Default using *Xgboost*," *Int. J. Comput. Appl.*, vol. 183, no. 32, hal. 975–8887, 2021.
- [7] X. Zhang dkk., "Data-Driven Loan Default Prediction: A Machine Learning Approach for Enhancing Business Process Management," *Systems*, vol. 13, no. 7, hal. 1–26, 2025, doi: 10.3390/systems13070581.
- [8] L. Munkhdalai, T. Munkhdalai, O. E. Namsrai, J. Y. Lee, dan K. H. Ryu, "An empirical comparison of machine-learning methods on bank client credit assessments," *Sustain.*, vol. 11, no. 3, hal. 1–23, 2019, doi: 10.3390/su11030699.
- [9] D. B. Saputra, V. Atina, dan F. E. Nastiti, "Penerapan Model *Crisp-dm* Pada Prediksi Nasabah Kredit Menggunakan Algoritma *Random forest*," *IDEALIS Indones. J. Inf. Syst.*, vol. 7, no. 2, hal. 240–247, 2024, doi: 10.36080/idealis.v7i2.3244.
- [10] I. Sugiyarto, B. Sudarsono, dan U. Faddillah, "Performance Comparison of Data Mining Algorithm to Predict Approval of Credit Card," vol. 4, no. 1, 2019, doi: <https://doi.org/10.33395/sinkron.v4i1.10181>.
- [11] P. Chapman dkk., "Step-by-step data mining guide," *SPSS inc*, vol. 78, hal. 1–78, 2000, [Daring]. Tersedia pada: <http://www.crisp-dm.org/CRISPWP-0800.pdf>
- [12] Scikit-learn, "Tuning the decision threshold for class prediction," *Scikit-learn Threshold Documentation*, 2024. https://scikit-learn.org/stable/modules/classification_threshold.html
- [13] S. van Buuren dan K. Groothuis-oudshoorn, "mice : Multivariate Imputation by Chained," vol. 45, no. 3, 2011.
- [14] T. Chen, T. He, dan M. Benesty, "*Xgboost*: A Scalable Tree Boosting System," *R Packag. version 0.71-2*, hal. 1–4, 2024.