

## Analisis Status Pembayaran Group Order NooBlue Menggunakan Algoritma XGBoost

Lyvia Nur Islami Surala<sup>1</sup>, Tukino<sup>2</sup>, April Lia Hananto<sup>3</sup>, Elfina Novalia<sup>4</sup>, Fitria Nurapriani<sup>5</sup>

<sup>1,2,3,4</sup>Program Studi Sistem Informasi, Universitas Buana Perjuangan Karawang

Email: <sup>1</sup>si22.lyviasurala@mhs.ubpkarawang.ac.id, <sup>2</sup>tukino@ubpkarawang.ac.id, <sup>3</sup>aprilialia@ubpkarawang.ac.id,

<sup>4</sup>elfinanovalia@rocketmail.com, <sup>5</sup>Fitria.apriani@ubpkarawang.ac.id

Email Penulis Korespondensi: si22.lyviasurala@mhs.ubpkarawang.ac.id

### Abstrak

Pertumbuhan pesat ekonomi digital telah mendorong meningkatnya penggunaan layanan transaksi daring, termasuk sistem pemesanan kelompok. Namun, tantangan baru turut muncul, khususnya dalam hal keandalan informasi terkait status pembayaran pelanggan. Permasalahan seperti keterlambatan atau kegagalan pembayaran dapat memengaruhi kestabilan keuangan bisnis. Oleh karena itu, penelitian ini bertujuan untuk membangun model prediksi status pembayaran pelanggan menggunakan algoritma *Extreme Gradient Boosting* (XGBoost). Data yang digunakan berasal dari transaksi NooBlue Shop pada Maret 2025, yang terdiri dari 403 entri. Proses analisis mencakup tahapan pra-pemrosesan data, pembagian data menjadi data latih dan uji dengan rasio 60:40, serta pelatihan model XGBoost. Model dievaluasi menggunakan metrik klasifikasi seperti akurasi, *precision*, *recall*, *F1-score*, dan *confusion matrix*. Hasil pengujian menunjukkan bahwa model mampu mengklasifikasikan status pembayaran dengan tingkat akurasi mencapai 90%. Nilai *precision* dan *recall* masing-masing berada pada kisaran 0.89–0.93, sedangkan *F1-score* menunjukkan performa yang seimbang untuk kedua kelas. Analisis lebih lanjut menunjukkan bahwa fitur total pembayaran dan uang muka merupakan kontributor utama dalam proses prediksi. Temuan ini menunjukkan bahwa penerapan XGBoost dapat memberikan solusi yang efektif dalam membantu perusahaan memantau status pembayaran pelanggan secara otomatis dan responsif, serta mendukung pengambilan keputusan berbasis data secara lebih tepat dan efisien di tengah dinamika ekonomi digital.

**Kata Kunci:** Status Pembayaran, Prediksi, Data Penjualan, XGBoost, Machine Learning

### Abstract

The rapid growth of the digital economy has driven the increasing use of online transaction services, including group booking systems. However, new challenges have also emerged, especially in terms of the reliability of information related to customer payment status. Problems such as late or failed payments can affect the financial stability of a business. Therefore, this study aims to build a customer payment status prediction model using the *Extreme Gradient Boosting* (XGBoost) algorithm. The data used comes from NooBlue Shop transactions in March 2025, consisting of 403 entries. The analysis process includes data pre-processing stages, dividing data into training and testing data with a ratio of 60:40, and training the XGBoost model. The model is evaluated using classification metrics such as accuracy, precision, recall, F1-score, and confusion matrix. The test results show that the model is able to classify payment status with an accuracy level of 90%. The precision and recall values are in the range of 0.89–0.93, respectively, while the F1-score shows balanced performance for both classes. Further analysis shows that the total payment and down payment features are the main contributors to the prediction process. These findings indicate that the implementation of XGBoost can provide an effective solution in helping companies monitor customer payment status automatically and responsively, as well as support data-based decision making more accurately and efficiently amidst the dynamics of the digital economy.

**Keywords:** Payment Status, Prediction, Sales Data, XGBoost, Machine Learning

## 1. PENDAHULUAN

Dalam era pertumbuhan ekonomi digital yang begitu cepat, berbagai platform transaksi daring mengalami kemajuan signifikan, termasuk layanan pemesanan kelompok (*group order*). Namun, seiring meningkatnya penggunaan sistem ini, muncul tantangan besar dalam menjamin keakuratan serta keandalan informasi terkait status pembayaran pelanggan. Masalah seperti keterlambatan pembayaran, ketidaksesuaian nominal, atau bahkan gagal bayar sepenuhnya, bisa berdampak serius pada kestabilan keuangan dan operasional bisnis [1]. Ketidakpastian semacam ini bukan hanya mengganggu arus kas, tetapi juga mempersulit pengambilan keputusan manajerial yang bergantung pada data pembayaran yang valid. Oleh karena itu, dibutuhkan suatu pendekatan berbasis teknologi yang dapat memprediksi potensi ketidaksesuaian pembayaran sedini mungkin, agar risiko finansial dapat ditekan dan efisiensi operasional tetap terjaga [2].

Salah satu solusi yang saat ini banyak diandalkan dalam menangani permasalahan prediksi berbasis data adalah penerapan algoritma pembelajaran mesin (*machine learning*) [3][4]. Di antara berbagai algoritma yang tersedia, *Extreme Gradient Boosting* (XGBoost) menonjol sebagai salah satu metode yang paling efektif dan efisien, terutama dalam menangani data tabular dalam skala besar. XGBoost merupakan algoritma boosting yang dikembangkan untuk meningkatkan akurasi model prediksi dengan membentuk ansambel dari pohon keputusan (*decision tree*) secara bertahap. Setiap iterasi pembelajaran difokuskan untuk memperbaiki kesalahan dari model sebelumnya, sehingga hasil akhirnya menjadi jauh lebih akurat dibandingkan pendekatan prediktif lainnya [5].

XGBoost memiliki keunggulan struktural yang menjadikannya sangat disukai dalam dunia industri dan penelitian. Dari sisi performa, algoritma ini dirancang untuk menghasilkan model yang tidak hanya cepat dalam pelatihan (training), tetapi juga efisien dalam prediksi [6]. XGBoost mendukung pemrosesan paralel, yang memungkinkan pelatihan dilakukan lebih cepat pada perangkat keras modern. Tak hanya itu, fitur regularisasi ganda (L1 dan L2) yang disematkan dalam algoritma ini berperan penting dalam mencegah overfitting, yaitu kondisi di mana model terlalu menyesuaikan diri terhadap data pelatihan hingga kehilangan kemampuan generalisasi terhadap data baru. Selain itu, kemampuan pruning atau pemangkasan pohon secara cerdas membantu menyederhanakan struktur model, menjaga keseimbangan antara kompleksitas dan akurasi.

Popularitas XGBoost tidak hanya terlihat dari penggunaannya yang luas di industri, tetapi juga dari dominasinya dalam berbagai kompetisi data science seperti Kaggle. Algoritma ini dikenal sebagai salah satu alat utama bagi para praktisi data dalam membangun solusi prediktif yang akurat dan tahan uji. Menurut Chen dan Guestrin (2016), XGBoost mampu mengungguli banyak algoritma tradisional dalam berbagai kasus prediksi, mulai dari pengenalan pola pelanggan hingga prediksi pasar keuangan [7]. Keunggulan ini menjadikan XGBoost sebagai pilihan utama dalam pengembangan sistem intelijen bisnis berbasis data.

Berbagai studi telah menunjukkan keberhasilan XGBoost dalam konteks prediksi perilaku pelanggan, termasuk dalam hal status pembayaran. Sebagai contoh, penelitian yang dilakukan oleh Wang et al. (2023) membangun model prediksi perilaku pembelian pelanggan dengan pendekatan XGBoost, dan hasilnya memperlihatkan kinerja yang jauh lebih unggul dibandingkan algoritma lain seperti *K-Nearest Neighbors* (KNN), *Support Vector Machine* (SVM), *Random Forest*, maupun jaringan saraf tiruan (ANN). Model yang dikembangkan berhasil mencapai nilai F1 score di atas 0,95 secara konsisten, membuktikan akurasi tinggi serta stabilitas performa prediktifnya. Studi ini juga menekankan pentingnya pemilihan fitur (*feature selection*) yang tepat serta pengolahan data awal (*data preprocessing*) yang cermat untuk meningkatkan kualitas prediksi [8].

Sementara itu, Zheng et al. (2024) menggabungkan XGBoost dengan *Synthetic Minority Oversampling Technique* (SMOTE) untuk menangani ketidakseimbangan kelas dalam data penipuan pembayaran. Integrasi ini meningkatkan akurasi model secara signifikan, dengan peningkatan performa hampir 6% dibandingkan metode konvensional [9]. Selain fokus pada aspek pembayaran, XGBoost juga banyak diterapkan untuk menganalisis loyalitas pelanggan, terutama dalam konteks prediksi churn atau kehilangan pelanggan. Studi Hermawan et al. (2023) mengembangkan model prediksi churn pelanggan pada platform e-commerce dengan XGBoost yang dikombinasikan dengan SMOTE untuk mengatasi masalah ketidakseimbangan data. Model ini berhasil mencapai nilai F2 score sebesar 0,849, yang menunjukkan efektivitasnya dalam mengidentifikasi pelanggan berisiko tinggi. Dengan informasi tersebut, perusahaan dapat merancang strategi retensi pelanggan yang lebih proaktif dan tepat sasaran [10].

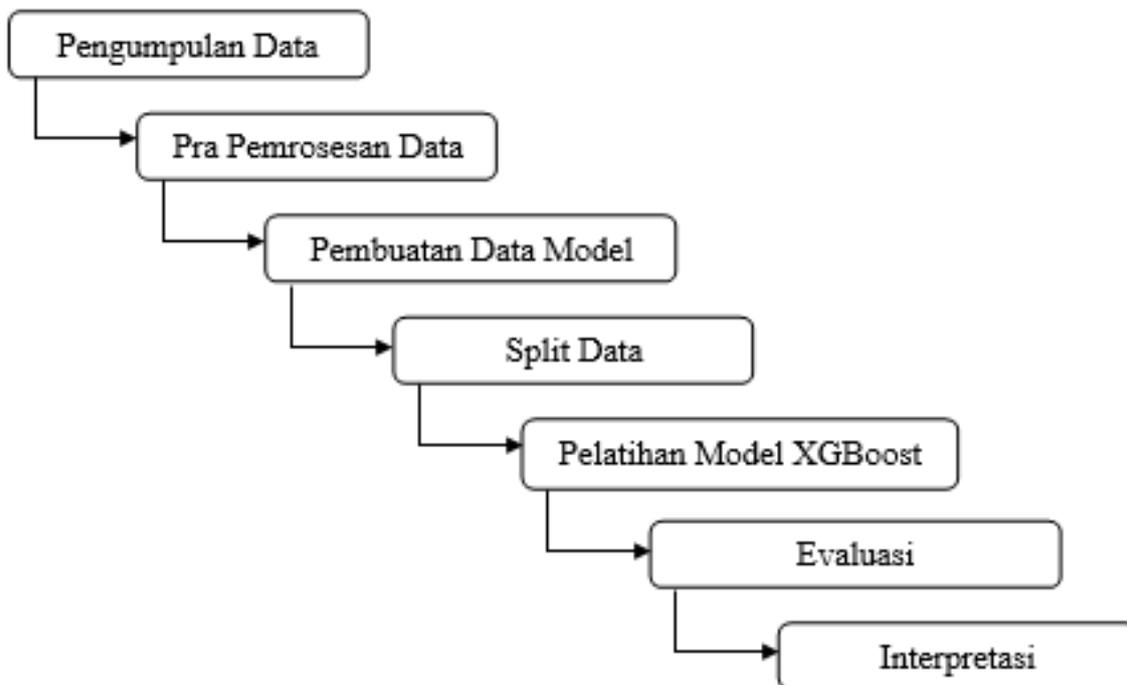
Tak hanya terbatas pada perilaku pengguna, algoritma ini juga terbukti efektif dalam prediksi penjualan. Dankorpo (2024) menerapkan XGBoost untuk memprediksi penjualan di sektor ritel, dan hasilnya menunjukkan penurunan rata-rata kesalahan prediksi (Mean Absolute Error dan Root Mean Square Error) lebih dari 29%. Hal ini mencerminkan kemampuan XGBoost dalam menangkap pola data yang kompleks serta fluktuasi musiman [11]. Studi serupa oleh Li (2023) juga membuktikan bahwa XGBoost mampu menangani data terbatas dan ketidakpastian musiman pada jaringan toko besar, sehingga sangat relevan untuk aplikasi skala besar dalam bisnis ritel [12].

Berdasarkan berbagai bukti empiris dan keunggulan teknis yang dimiliki, studi ini bertujuan untuk menerapkan algoritma XGBoost dalam menganalisis status pembayaran pelanggan pada platform layanan pemesanan kelompok NooBlue. Dengan memanfaatkan kemampuan prediktif yang dimiliki XGBoost, diharapkan perusahaan dapat memperoleh sistem deteksi dini terhadap risiko keterlambatan atau kegagalan pembayaran, sekaligus mendukung pengambilan keputusan yang lebih akurat dan cepat dalam merespons potensi gangguan terhadap stabilitas keuangan. Pendekatan ini diharapkan menjadi solusi praktis dalam meningkatkan efisiensi operasional sekaligus memperkuat daya saing perusahaan di era ekonomi digital yang kompetitif.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Proses pembangunan model prediksi dengan algoritma XGBoost dalam penelitian ini dilakukan melalui beberapa tahapan yang sistematis. Gambar 1 berikut menunjukkan alur kerja yang dimulai dari pengumpulan data hingga pembahasan hasil evaluasi model.



Gambar 1. Tahapan Penelitian

Berikut ini merupakan penjelasan dari tahapan penelitian

- a. Pengumpulan Data  
Data yang digunakan dalam penelitian ini adalah transaksi dari NooBlue Shop pada bulan Maret 2025. Data terdiri dari 403 transaksi.
- b. Pra Pemrosesan Data  
Proses pra pemrosesan data adalah melakukan pemilihan features yang akan digunakan dalam penelitian ini yaitu jumlah pesanan (Qty), uang muka (Dp), pelunasan (Pelunasan), dan total pembayaran (Total). Selain itu, terdapat kolom target klasifikasi yaitu Status yang berisi dua kelas, yaitu Lunas dan Belum Lunas.
- c. Split Data  
Split data yang digunakan di dalam penelitian ini adalah 60:40, 60% data training dan 40% data testing. [13]
- d. Pelatihan Model XGBoost  
Extreme Gradient Boosting (XGBoost) adalah metode boosting yang dikembangkan oleh Chen & Guestrin pada tahun 2016. XGBoost bekerja dengan membentuk serangkaian pohon keputusan di mana setiap pohon selanjutnya dibangun berdasarkan hasil pohon sebelumnya. Pohon pertama biasanya masih lemah dalam melakukan klasifikasi karena hanya menggunakan inisialisasi probabilitas yang telah ditentukan. Selanjutnya, bobot pada setiap pohon akan diperbarui secara bertahap sehingga terbentuk kumpulan pohon klasifikasi yang lebih kuat [14].

Adapun tahapan utama dalam membangun model XGBoost adalah sebagai berikut [15]:

1. Inisialisasi Awal, model dimulai dengan prediksi sederhana seperti rata-rata dari nilai target pada data pelatihan.
 
$$\hat{y}^{(0)} = \frac{1}{n} \sum_{i=1}^n y_i \tag{1}$$

$\hat{y}^{(0)}$  = Prediksi awal untuk semua data  
 $y_i$  = Nilai target aktual ke i  
 $n$  = Jumlah total data
2. Menghitung Residual, selisih antara prediksi awal dan nilai aktual dihitung sebagai residual untuk mencerminkan kesalahan yang harus diperbaiki.
 
$$r_i^{(m)} = y_i - \hat{y}_i^{(m-1)} \tag{2}$$

$r_i^{(m)}$  = Residual data  
 $y_i$  = Nilai target aktual ke i  
 $\hat{y}_i^{(m-1)}$  = Prediksi model sebelum iterasi m
3. Membangun Pohon Baru, pohon keputusan dibuat untuk memperkirakan nilai residual, berfungsi sebagai koreksi terhadap kesalahan model sebelumnya.

- Memperbarui Model, hasil dari pohon baru ditambahkan ke model dengan dikalikan faktor pembelajaran (*learning rate*) untuk mengatur kontribusinya.

$$\hat{y}_i^{(m)} = \hat{y}_i^{(m-1)} + \eta \cdot fm(xi) \tag{3}$$

$\hat{y}_i^{(m)}$  = Prediksi akhir pada iterasi ke-m

$\eta$  = Learning rate ( $0 < \eta \leq 1$ ), untuk mengontrol besarnya koreksi

$fm(xi)$  = Output dari pohon ke-m untuk data ke-i

- Iterasi, proses dari perhitungan residual, pembangunan pohon, dan pembaruan model diulang hingga jumlah iterasi tertentu atau akurasi yang diharapkan tercapai.

- Regularisasi, digunakan untuk menghindari overfitting, memastikan model tetap akurat saat dihadapkan pada data baru.

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \tag{4}$$

$\Omega(f)$  = Penalti kompleksitas model

$T$  = Jumlah daun pada pohon

$w_j$  = Nilai prediksi pada daun ke-j

$\gamma$  = Penalti untuk jumlah daun (regularisasi struktur)

$\lambda$  = Penalti L2 pada bobot (regularisasi berat prediksi)

- Evaluasi

Model yang telah dilatih dievaluasi untuk mengetahui seberapa baik model dalam melakukan klasifikasi status pembayaran. Metrik evaluasi yang digunakan dalam penelitian ini meliputi akurasi, presisi, *recall*, *F1-score*, dan *confusion matrix*. Evaluasi ini penting untuk memastikan bahwa model tidak hanya tepat secara keseluruhan, tetapi juga dapat mengenali kedua kelas target “Lunas” dan “Belum Lunas” secara seimbang.

- Interpretasi

Tahap akhir adalah melakukan interpretasi dari hasil model yang telah dibangun dan dievaluasi. Berdasarkan metrik performa dan analisis fitur, dapat disimpulkan seberapa baik model bekerja dan fitur apa yang paling relevan. Temuan ini dapat dimanfaatkan untuk mendukung pengambilan keputusan, seperti menentukan strategi penagihan yang lebih efektif atau mengidentifikasi pelanggan yang memiliki potensi keterlambatan pembayaran.

### 3. HASIL DAN PEMBAHASAN

Hasil implementasi algoritma *Extreme Gradient Boosting* (XGBoost) terhadap data yang dianalisis disajikan untuk menunjukkan kinerja model dalam memprediksi variabel target. Proses pemodelan dilakukan secara sistematis, diawali dengan pengumpulan data, pra pemrosesan data, penentuan split data, dan pelatihan model XGBoost. Pemaparan hasil ditujukan untuk menggambarkan efektivitas model dalam mengidentifikasi pola serta hubungan antar variabel dalam dataset. Selanjutnya, evaluasi performa akan dijelaskan berdasarkan metrik yang digunakan, diikuti interpretasi dan analisis terhadap hasil yang diperoleh.

#### 3.1 Pengumpulan Data

Data transaksi yang diperoleh dari NooBlue Shop terdiri dari 403 transaksi dengan 6 kolom utama yaitu no, nama untuk nama pelanggan yang melakukan pre order, Order mewakili jenis barang yang dibeli, Qty mewakili jumlah barang, Dp merupakan uang muka atau pembayaran yang dilakukan oleh pelanggan, Total merupakan jumlah yang harus dibayar oleh pelanggan, dan pelunasan merupakan kekurangan dari pembayaran pelanggan.

Tabel 1. Data Transaksi NooBlue Shop Bulan Maret 2025

No	Nama	Order	Qty	Dp	Total	Pelunasan
1	Fitri	Cc Set Boys	2	150000	260000	110000
2	Indy	Cc Set Boys	1	130000	130000	0
3	Ian	Cc Set Boys	2	150000	260000	110000
4	Novieka	Cc Set Boys,Cc Set Girls	3	400000	400000	0
5	Tania	Cc Set Boys	1	130000	130000	0
...	...	...	...	...	...	...
399	Evelyn	Mie Buldak Ungu	1	60000	60000	0
400	Rani	Mie Hindee Brand	3	30000	60000	30000
401	Kaka	Jele Merah Zeenunew	1	20000	50000	30000
402	Tya	Shampoo Sunsilk 30ml	1	285000	285000	0
403	Lala	Shampoo Sunsilk 30ml	1	170000	285000	115000

#### 3.2 Pra Pemrosesan Data

Pada proses ini akan dilakukan pra pemrosesan data terlebih dahulu dengan menambahkan kolom klasifikasi untuk menentukan pembayaran Lunas atau Belum Lunas dengan melakukan pengecekan pada kolom pelunasan, jika nilai pada

pelunasan 0, maka kolom klasifikasi berisi Lunas dan jika kolom pelunasan bernilai lebih dari 0 maka kolom klasifikasi berisi Belum Lunas.

Tabel 2. Data Klasifikasi

No	Nama	Order	Qty	Dp	Total	Pelunasan	Klasifikasi
0	Fitri	Cc Set Boys	2	150000	260000	110000	Belum Lunas
1	Indy	Cc Set Boys	1	130000	130000	0	Lunas
2	Ian	Cc Set Boys	2	150000	260000	110000	Belum Lunas
3	Novieka	Cc Set Boys,Cc Set Girls	3	400000	400000	0	Lunas
4	Tania	Cc Set Boys	1	130000	130000	0	Lunas
...	...	...	...	...	...	...	...
398	Evelyn	Mie Buldak Ungu	1	60000	60000	0	Lunas
399	Rani	Mie Hindee Brand	3	30000	60000	30000	Belum Lunas
400	Kaka	Jele Merah Zeenunew	1	20000	50000	30000	Belum Lunas
401	Tya	Shampoo Sunsilk 30ml	1	285000	285000	0	Lunas
402	Lala	Shampoo Sunsilk 30ml	1	170000	285000	115000	Belum Lunas

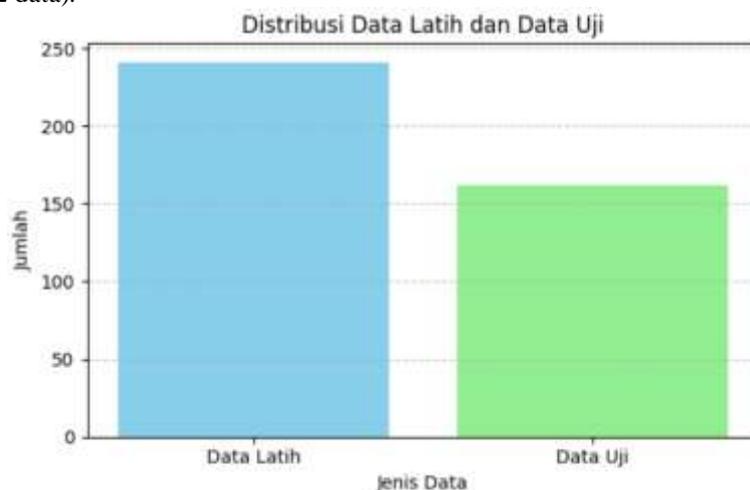
Data klasifikasi akan dirubah ke dalam bentuk numerik (0 dan 1) agar memudahkan dalam proses pelatihan [16]. Nilai 0 melambangkan Belum Lunas dan nilai 1 melambangkan Lunas.

```
array([0, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0,
0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, 1, 1, 0, 1, 1,
1, 1, 1, 0, 0, 1, 0, 1, 1, 1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 0, 1,
0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 1, 0,
1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 1, 1, 1,
1, 1, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0,
0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1,
1, 0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0,
1, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 1, 0, 1, 1, 1, 1,
0, 0, 1, 1, 0, 0, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0,
0, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0,
0, 1, 0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0,
1, 1, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0,
1, 0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1, 1, 1, 0, 1,
0, 0, 1, 0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0,
1, 0, 1, 1, 1, 0, 0, 1, 0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 0, 0,
1, 0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 0, 1, 1, 1, 1, 1, 0, 0, 1,
0, 1, 1, 0, 0, 1, 0])
```

Gambar 2. Hasil Pengkodean Klasifikasi

### 3.3 Split data

Rasio pembagian data (split data) dapat mempengaruhi akurasi model, dengan rasio tertentu memberikan performa yang lebih optimal tergantung pada karakteristik data dan algoritma yang digunakan [17]. Dalam penelitian ini digunakan rasio 60:40 dimana 60% dari data total digunakan untuk data training (241 data) dan 40% dari data total akan digunakan sebagai data testing (162 data).



Gambar 3. Distribusi Data latih dan Data Uji

3.4 Pelatihan Model XGBoost

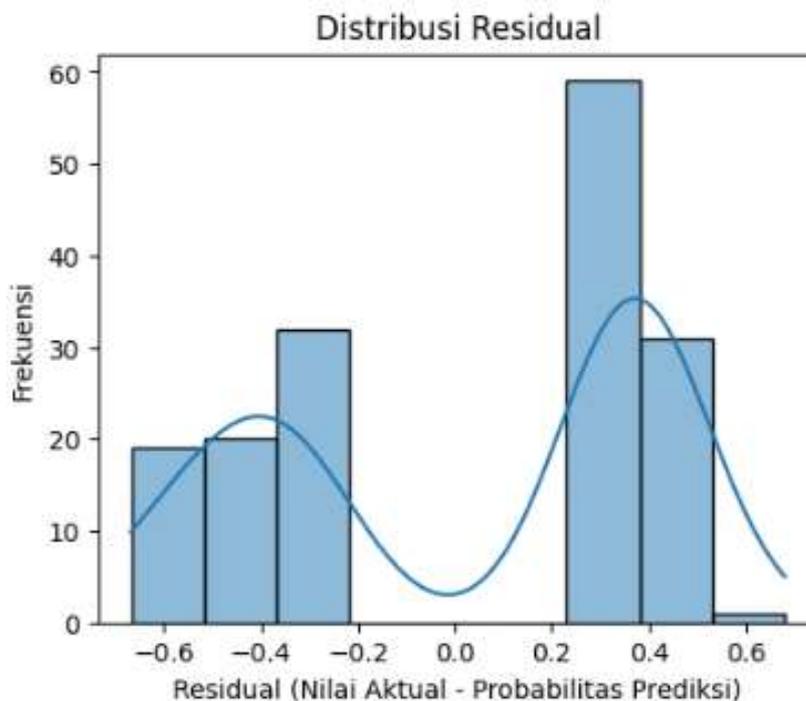
a. Inisialisasi Data

Hasil evaluasi model XGBoost dengan menggunakan rasio 60:40 dalam 10 iterasi pertama, menampilkan nilai *logloss* untuk data pelatihan (*train-logloss*) dan data pengujian (*test-logloss*). Pada iterasi ke-0, terlihat bahwa nilai *train-logloss* sebesar 0.68770 dan *test-logloss* sebesar 0.68704. Ini mencerminkan tahapan inisialisasi awal dari model XGBoost, di mana prediksi masih sangat sederhana, biasanya berupa rata-rata log-odds dari label target. Seiring bertambahnya jumlah iterasi, model mulai membentuk pohon-pohon keputusan tambahan untuk mengoreksi kesalahan dari model sebelumnya. Hal ini ditunjukkan dengan penurunan bertahap pada nilai logloss di setiap iterasi. Hingga iterasi ke-9, nilai *train-logloss* menurun menjadi 0.50238 dan *test-logloss* menjadi 0.52037, yang menunjukkan bahwa model berhasil memperbaiki prediksi secara progresif. Penurunan logloss ini menandakan bahwa proses pembelajaran berjalan efektif dan model belum mengalami overfitting karena selisih antara logloss pelatihan dan pengujian masih tergolong kecil.

[0]	train-logloss:0.68770	test-logloss:0.68704
[1]	train-logloss:0.68030	test-logloss:0.68557
[2]	train-logloss:0.63609	test-logloss:0.64467
[3]	train-logloss:0.63372	test-logloss:0.64345
[4]	train-logloss:0.63046	test-logloss:0.64222
[5]	train-logloss:0.58532	test-logloss:0.59957
[6]	train-logloss:0.55262	test-logloss:0.56630
[7]	train-logloss:0.53467	test-logloss:0.54671
[8]	train-logloss:0.53218	test-logloss:0.54657
[9]	train-logloss:0.50238	test-logloss:0.52037

Gambar 4. Hasil Inisialisasi Awal

b. Hasil Residual



Gambar 5. Hasil Distribusi Residual

Histogram menunjukkan sebaran residual (selisih antara nilai aktual dan prediksi probabilitas) yang sebagian besar terkonsentrasi di sekitar nol. Hal ini mengindikasikan bahwa prediksi model secara umum mendekati nilai aktual, mencerminkan tingkat akurasi yang tinggi dalam memperkirakan probabilitas. Kurva distribusi juga menunjukkan pola yang relatif seimbang antara nilai positif dan negatif, yang merupakan indikasi bahwa model bekerja secara konsisten pada berbagai rentang nilai prediksi [18].

c. Membangun Pohon baru

```
Continuing training the existing model for 1 more round...

Model 'bst_continued' now includes one more tree trained based on the gradients/hessians from 'bst'.
Total trees in 'bst': 10
Total trees in 'bst_continued': 11

Continuing training the existing model for 1 more round...

Model 'bst_continued' now includes one more tree trained based on the gradients/hessians from 'bst'.
Total trees in 'bst': 10
Total trees in 'bst_continued': 11
```

Gambar 6. Pembentukan Pohon baru

Model XGBoost yang awalnya memiliki 10 pohon (bst) diperluas menjadi 11 pohon dalam model lanjutan (bst\_continued) setelah satu putaran pelatihan tambahan. Meskipun proses dilanjutkan dua kali, hanya satu pohon yang berhasil ditambahkan. Hal ini bisa disebabkan oleh tidak adanya pembaruan signifikan pada gradien atau parameter pelatihan yang membatasi penambahan pohon. Model dasar tetap tidak berubah, sementara pelatihan tambahan hanya efektif pada iterasi pertama [19].

d. Hasil Prediksi Xgboost

Dengan menggunakan rasio data 60:40 yang berarti bahwa jumlah data training sebanyak 241 dan data uji sebanyak 162 data menghasilkan hasil prediksi sebagai berikut.

Tabel 3. Hasil Prediksi XGBoost

Index	Qty	Dp	Total	Status Aktual	Status Prediksi	Probabilitas Lunas	Probabilitas Belum Lunas
70	1	80000	80000	Lunas	Lunas	0.600	0.400
281	2	200000	315000	Belum Lunas	Belum Lunas	0.367	0.633
283	1	65000	135000	Belum Lunas	Belum Lunas	0.456	0.544
33	1	140000	140000	Lunas	Lunas	0.624	0.376
42	1	50000	140000	Belum Lunas	Belum Lunas	0.433	0.567
...	...	...	...	...	...	...	...
103	1	220000	220000	Lunas	Lunas	0.625	0.375
81	1	210000	210000	Lunas	Lunas	0.661	0.339
208	1	75000	75000	Lunas	Lunas	0.620	0.380
362	3	285000	285000	Lunas	Lunas	0.634	0.366
378	3	60000	60000	Lunas	Lunas	0.696	0.304

Berdasarkan hasil pengujian, diketahui bahwa dari 162 data, XGBoost melakukan prediksi benar sebanyak 146 kali dan prediksi salah sebanyak 16 kali.

**3.5 Evaluasi**

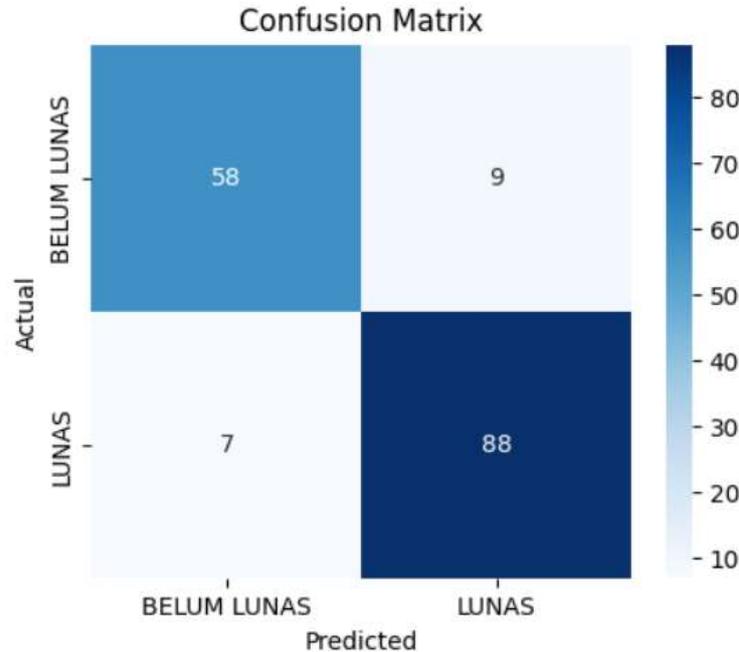
Evaluasi performa menggunakan beberapa metrik klasifikasi: akurasi, *precision*, *recall*, *F1-score*, dan *confusion matrix*.

	precision	recall	f1-score	support
BELUM LUNAS	0.89	0.87	0.88	67
LUNAS	0.91	0.93	0.92	95
accuracy			0.90	162
macro avg	0.90	0.90	0.90	162
weighted avg	0.90	0.90	0.90	162

Gambar 7. Evaluasi Hasil Prediksi XGBoost

Berdasarkan gambar tersebut, model klasifikasi menghasilkan performa yang cukup baik dengan tingkat akurasi sebesar 90%, yang berarti dari 162 data, 90% di antaranya berhasil diklasifikasikan dengan benar. Nilai *precision* untuk kelas BELUM LUNAS adalah 0.89 dan untuk LUNAS adalah 0.91, menunjukkan bahwa model cukup tepat dalam mengidentifikasi kedua kelas tersebut. Nilai *recall* masing-masing adalah 0.87 untuk BELUM LUNAS dan 0.93 untuk LUNAS, yang menunjukkan bahwa model mampu mengenali mayoritas data yang benar-benar termasuk dalam masing-

masing kelas. Lebih lanjut, skor F1 yang merupakan harmonisasi antara *precision* dan *recall* menunjukkan nilai 0.88 untuk BELUM LUNAS dan 0.92 untuk LUNAS, menandakan performa yang seimbang. Rata-rata keseluruhan (*macro dan weighted average*) juga berada pada angka 0.90, memperkuat kesimpulan bahwa model memiliki performa yang stabil pada kedua kelas.

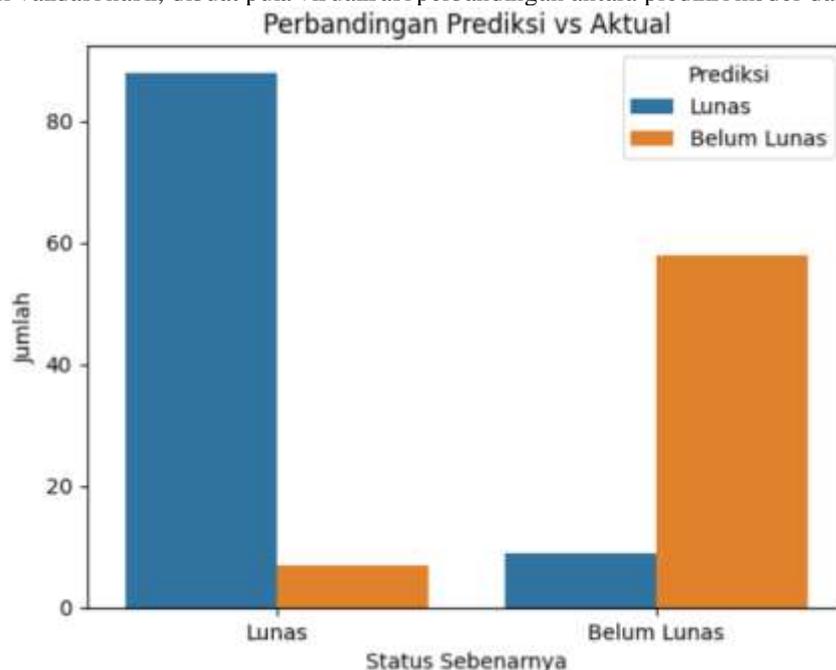


Gambar 8. Confusion matrix Hasil Prediksi XGBoost

Analisis dari *confusion matrix* menunjukkan bahwa dari 67 data aktual BELUM LUNAS, sebanyak 58 diprediksi dengan benar dan 9 salah diklasifikasikan sebagai LUNAS. Sementara dari 95 data aktual LUNAS, sebanyak 88 diprediksi dengan tepat dan hanya 7 yang keliru. Kesalahan klasifikasi masih tergolong kecil dan tidak signifikan terhadap performa keseluruhan model. Dengan demikian, model ini menunjukkan kinerja klasifikasi yang sangat baik dan seimbang dalam mengidentifikasi kedua kelas target.

### 3.6 Interpretasi

Sebagai bagian dari validasi hasil, dibuat pula visualisasi perbandingan antara prediksi model dan data actual.



Gambar 9. Perbandingan Prediksi dan Aktual

Gambar 9 menunjukkan perbandingan antara status sebenarnya dengan hasil prediksi yang dilakukan oleh model XGBoost. Terlihat bahwa sebagian besar data dengan status “Lunas” berhasil diprediksi secara akurat, sementara pada data “Belum Lunas” masih terdapat prediksi yang tidak sepenuhnya tepat. Meski demikian, model tetap menunjukkan kinerja yang baik dan cukup seimbang dalam membedakan kedua kelas. Visualisasi ini memperlihatkan bahwa pendekatan machine learning yang digunakan memiliki potensi besar untuk diimplementasikan dalam sistem monitoring status pembayaran pelanggan, serta membantu perusahaan dalam mengidentifikasi pelanggan yang memerlukan tindak lanjut lebih dini.

#### 4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma Extreme Gradient Boosting (XGBoost) mampu mengklasifikasikan status pembayaran pelanggan dengan performa yang sangat baik. Model dibangun berdasarkan data transaksi yang terdiri dari jumlah pesanan (Qty), uang muka (Dp), dan total pembayaran (Total). Setelah melalui tahapan pra-pemrosesan dan pembagian data menjadi 60% untuk pelatihan dan 40% untuk pengujian, model menunjukkan kinerja yang tinggi. Model menghasilkan akurasi sebesar 90%, artinya 90% dari seluruh prediksi sesuai dengan status pembayaran yang sebenarnya. Nilai *precision* sebesar 0.91 diperoleh karena dari 97 data yang diprediksi sebagai “Lunas”, sebanyak 88 di antaranya memang benar-benar “Lunas”, dan sisanya merupakan kesalahan prediksi.

Sementara itu, nilai *recall* sebesar 0.93 menunjukkan bahwa dari 95 data yang benar-benar “Lunas”, model berhasil mengenali 88 di antaranya. Nilai *F1-score* yang dihasilkan sebesar 0.92, mencerminkan keseimbangan antara *precision* dan *recall* pada prediksi model. Hasil analisis juga menunjukkan bahwa fitur Total dan Dp memberikan kontribusi terbesar terhadap prediksi model, dibandingkan dengan fitur Qty. Hal ini menunjukkan bahwa besarnya pembayaran pelanggan sangat mempengaruhi klasifikasi status pelunasan. Dengan hasil ini, model XGBoost dapat dijadikan sebagai alat bantu dalam sistem informasi untuk memantau status pembayaran pelanggan secara otomatis, serta membantu perusahaan dalam mengambil keputusan lebih cepat terhadap pelanggan yang berpotensi belum melunasi tagihannya.

#### UCAPAN TERIMAKASIH

Ucapan terima kasih ditujukan kepada semua pihak yang telah memberikan dukungan dalam pelaksanaan penelitian ini.

#### DAFTAR PUSTAKA

- [1] W. H. Susanto and A. F. Chawa, “AKSIGAGAL BAYAR PADA PERUSAHAAN FINTECH,” *JSSH (Jurnal Sains Sos. dan Humaniora)*, vol. 5, no. 1, 2021, doi: 10.30595/jssh.v5i1.9305.
- [2] M. N. R. Fitriani, B. Priyatna, B. Huda, A. L. Hananto, and T. Tukino, “Implementasi Metode K-Means Untuk Memprediksi Status Kredit Macet,” *J. Sist. Komput. dan Inform.*, vol. 4, no. 3, 2023, doi: 10.30865/json.v4i3.5953.
- [3] F. Shinta Dewi and T. Dewayanto, “Peran Big Data Analytics, Machine Learning, Dan Artificial Intelligence Dalam Pendeteksian Financial Fraud: a Systematic Literature Review,” *Diponegoro J. Account.*, vol. 13, no. 3, pp. 1–15, 2024, [Online]. Available: <http://ejournal-s1.undip.ac.id/index.php/accounting>
- [4] B. Priyatna, Z. A. Bakar, N. Zamin, and Y. Yahya, “Machine Learning Trends in Mushroom Agriculture: A Systematic Review Methodology,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2024, doi: 10.1007/978-981-99-7339-2\_47.
- [5] M. Erkamim, A. M. Tanniewa, and I. Ap, “Analisis Perbandingan Metode AdaBoost, Gradient Boosting, dan XGBoost Untuk Kalsifikasi Status Gizi Pada Balita,” *J. Build. Informatics, Technol. Sci.*, vol. 6, no. 3, pp. 1799–1807, 2024, doi: 10.47065/bits.v6i3.5717.
- [6] P. Septiana Rizky, R. Haiban Hirzi, and U. Hidayaturrohmah, “Perbandingan Metode LightGBM dan XGBoost dalam Menangani Data dengan Kelas Tidak Seimbang,” *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 15, no. 2, 2022, doi: 10.36456/jstat.vol15.no2.a5548.
- [7] R. Siringoringo, R. Perangin-angin, and M. J. Purba, “SEGMENTASI DAN PERAMALAN PASAR RETAIL MENGGUNAKAN XGBOOST DAN PRINCIPAL COMPONENT ANALYSIS,” *METHOMIKA J. Manaj. Inform. dan Komputerisasi Akunt.*, vol. 5, no. 1, 2021, doi: 10.46880/jmika.vol5no1.pp42-47.
- [8] W. Wang *et al.*, “A User Purchase Behavior Prediction Method Based on XGBoost,” *Electron.*, vol. 12, no. 9, 2023, doi: 10.3390/electronics12092047.
- [9] Q. Zheng, C. Yu, J. Cao, Y. Xu, Q. Xing, and Y. Jin, “Advanced Payment Security System: XGBoost, CatBoost and SMOTE Integrated,” <https://arxiv.org/>, 2024, doi: 10.1109/MetaCom62920.2024.00063.
- [10] A. Hermawan, A. Saputra, M. D. Rafi, and S. Basmallah, “Implementing XGBoost Model for Predicting Customer Churn in E-Commerce Platforms,” *J. Repeater Publ. Tek. Inform. dan Jar.*, vol. 3, no. 2, 2025.
- [11] P. Dankorpo, “Sales Forecasting for Retail Business using XGBoost Algorithm,” *J. Comput. Sci. Technol. Stud.*, vol. 6, no. 2, pp. 136–141, 2024, doi: 10.32996/jcsts.
- [12] K. Li, “A Sales Prediction Method Based on XGBoost Algorithm Model,” *BCP Bus. Manag.*, vol. 36, 2023, doi: 10.54691/bcpbm.v36i.3487.
- [13] M. A. Y. Linda, T. R. I. Alfianti, and R. Supriyanto, “Perbandingan Kinerja Algoritma Random Forest, AdaBoost, dan XGBoost dalam Memprediksi Risiko Penyakit Osteoporosis Comparison of the Performance of Random Forest, AdaBoost, and XGBoost Algorithms in Predicting the Risk of Osteoporosis Disease,” *J. Ilmu Komput. dan Agri-informatika*, vol. 11, no. 2, pp. 172–184, 2024.

- [14] R. Yoris, "PENGUNAAN METODE XGBOOST UNTUK KLASIFIKASI STATUS OBESITAS DI INDONESIA," *Univ. HASANUDDIN MAKASSAR*, 2021.
- [15] G. A. Shafila, "Implementasi Metode Extreme Gradient Boosting (Xgboost) untuk Klasifikasi pada Data Bioinformatika (Studi Kasus : Penyakit Ebola , GSE 122692)," *Dspace.Uii.Ac.Id*, 2020.
- [16] N. J. Teague, "Feature Encodings for Gradient Boosting with Automunge," *arXiv Prepr. arXiv2209.12309*, no. NeurIPS, 2022.
- [17] R. Oktafiani, A. Hermawan, and D. Avianto, "Pengaruh Komposisi Split data Terhadap Performa Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma Machine Learning," *J. Sains dan Inform.*, 2023, doi: 10.34128/jsi.v9i1.622.
- [18] T. Nagamani and S. Logeswari, "Hybrid classification of XGBoost-based ADAM optimization for coronary artery disease diagnosis," *J. Intell. Fuzzy Syst.*, vol. 46, no. 4, 2024, doi: 10.3233/JIFS-233804.
- [19] N. Faqir, C. Loqman, and J. Boumhidi, "Deep Q-learning Approach based on CNN and XGBoost for Traffic Signal Control," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 9, 2022, doi: 10.14569/IJACSA.2022.0130961.