BAB III
METODE PENELITIAN

3.1 Objek Penelitian

Objek penelitian dalam penelitian ini adalah *dataset Loan Default Dataset* yang diperoleh dari platform *Kaggle*, dengan format file Loan_default.csv. *Dataset* ini berisi data historis pinjaman yang digunakan untuk menganalisis pola pembayaran serta memprediksi kemungkinan terjadinya keterlambatan pembayaran atau gagal bayar pada pinjaman *online*.

Secara keseluruhan, *dataset* yang digunakan dalam penelitian ini memiliki beberapa karakteristik tipe data, sebagai berikut:

a. Fitur numerik bertipe float64

   Fitur tersebut berisi data numerik kontinu, seperti nilai rasio, persentase, atau besaran terukur lainnya yang merepresentasikan karakteristik peminjam dan pinjaman.

b. Fitur numerik bertipe int64

   Fitur tersebut berisi data numerik diskrit atau nilai kategori dalam bentuk angka yang digunakan untuk membedakan kondisi tertentu pada data pinjaman.

c. 1 fitur target

   Yang menunjukkan status pembayaran pinjaman, yaitu apakah peminjam melakukan pembayaran tepat waktu atau mengalami keterlambatan pembayaran.

d. 1 fitur waktu

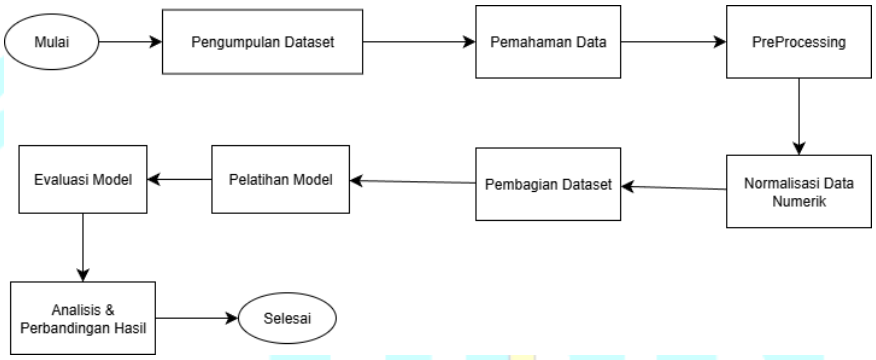
   Menunjukkan periode data dalam format tanggal (YYYY-MM-DD) dan digunakan untuk merepresentasikan waktu terjadinya transaksi atau status pinjaman.

Seluruh data telah melalui proses penyamaran identitas, sehingga tidak memuat informasi pribadi peminjam dan aman digunakan untuk kepentingan penelitian akademik. Meskipun variabel dalam *dataset* bersifat anonim, pola dan hubungan antar fitur tetap dapat dianalisis menggunakan teknik statistik dan *machine learning*. Penelitian ini bertujuan untuk menguji pola keterlambatan

pembayaran pinjaman berdasarkan data historis, bukan untuk menginterpretasikan makna asli dari setiap fitur.

3.2 Prosedur Penelitian

Prosedur penelitian menggambarkan tahapan-tahapan yang dilakukan mulai dari pengumpulan data hingga implementasi model. Setiap tahap dilakukan secara sistematis agar hasil yang diperoleh akurat dan dapat di pertanggung jawabkan. Alur penelitian secara umum dapat dilihat di bawah ini:



Gambar 3. 1 *flowchart* penelitian

Penjelasan *flowchart 3*.1:

a. Mulai

Proses penelitian diawali dengan langkah persiapan awal sebelum pengolahan data dilakukan.

b. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini adalah *Loan Default dataset* yang diperoleh dari platform *Kaggle* dengan format file loan.csv. *Dataset* ini menjadi sumber data utama dalam penelitian.

c. Pemahaman Data

Pada tahap ini dilakukan pemahaman terhadap struktur *dataset*, termasuk identifikasi jenis fitur, karakteristik data, serta penentuan variabel target yang menunjukkan status pembayaran pinjaman.

d. PreProcessing

Tahap *processing* meliputi pembersihan data dan penanganan nilai kosong (*missing value*) untuk memastikan kualitas data sebelum digunakan pada tahap pemodelan.

e. Normalisasi Data Numerik

Normalisasi dilakukan pada fitur numerik untuk menyamakan skala nilai antar fitur sehingga tidak terjadi perbedaan bobot yang berlebihan dalam proses pelatihan model.

f. Pembagian *Dataset*

Dataset dibagi menjadi data latih, data valid dan data uji. Pembagian ini bertujuan untuk melatih model dan mengevaluasi performa model secara objektif menggunakan data yang belum pernah dilatih sebelumnya.

g. Pelatihan Model

Pada tahap ini dilakukan pelatihan model menggunakan algoritma *Random Forest* dan *XGBoost* untuk membangun model klasifikasi prediksi keterlambatan pembayaran pinjaman.

h. Evaluasi Model

Model yang telah dilatih dievaluasi menggunakan metrik evaluasi seperti *accuracy*, *precision*, *recall* dan *F1-score* untuk mengukur kinerja masing-masing algoritma.

i. Analisis dan Perbandingan Hasil Model

Hasil evaluasi dianalisis dan dibandingkan untuk mengetahui perbedaan performa antara algoritma *Random Forest* dan XGBoost dalam memprediksi keterlambatan pembayaran.

j. Selesai

Tahap ini berisi rangkuman hasil penelitian berdasarkan analisis dan evaluasi model yang telah dilakukan.

3.3 Variabel Penelitian

Variabel penelitian merupakan komponen penting yang digunakan dalam membangun serta mengevaluasi model prediksi. Dalam penelitian ini, variabel dibagi menjadi dua kelompok utama, yaitu variabel dependen dan variabel independen. Pembagian ini menentukan peran masing-masing variabel dalam proses pemodelan *machine learning* yang digunakan.

Di bawah ini merupakan tabel variabel penelitian yang di gunakan pada penelitian ini:

Tabel 3. 1 Variabel Penelitian

No	Nama Fitur	Tipe Data	Digunakan	Keterangan
1	id	int64	Tidak	Identitas unik peminjam
2	member_id	int64	Tidak	ID anggota (dihapus)
3	emp_title	object	Tidak	Jabatan pekerjaan (terlalu variatif)
4	emp_length	object	Ya	Lama bekerja (kategorikal)
5	home_ownership	object	Ya	Status kepemilikan rumah
6	addr_state	object	Tidak	Kode negara bagian (terlalu variatif)
7	loan_amnt	float64	Ya	Jumlah pinjaman yang diajukan
8	term	object	Ya	Jangka waktu pinjaman (36/60 bulan)
9	int_rate	float64	Ya	Tingkat bunga pinjaman (%)
10	installment	float64	Ya	Besaran cicilan bulanan

No	Nama Fitur	Tipe Data	Digunakan	Keterangan
11	purpose	object	Ya	Tujuan penggunaan pinjaman
12	annual_inc	float64	Ya	Pendapatan tahunan peminjam
13	dti	float64	Ya	Rasio utang terhadap pendapatan
14	delinq_2yrs	float64	Ya	Jumlah keterlambatan dalam 2 tahun
15	inq_last_6mths	float64	Ya	Jumlah inquiry kredit 6 bulan terakhir
16	open_acc	float64	Ya	Jumlah akun kredit aktif
17	revol_bal	float64	Ya	Saldo kredit bergulir
18	revol_util	float64	Ya	Utilisasi kredit bergulir (%)
19	Total_acc	Float64	Ya	Total akun kredit
20	Loan status	Object	Ya	Status pinjaman (target)

Berdasarkan tabel 3.1 dari 20 kolom yang tersedia pada *dataset*, sebanyak 16 fitur digunakan sebagai variabel *independent* (x) dan 1 fitur yaitu loan_status digunakan sebagai variabel dependen (y) atau target prediksi. Fitur id, member_id, emp_title, dan addr_state tidak diikutsertakan dalam pemodelan karena bersifat sebagai pengenal unik atau memiliki tingkat keberagaman data yang terlalu tinggi sehingga tidak informatif secara prediktif.

3.2.1 Variabel Target

Variabel target dalam penelitian ini adalah loan_status, yang merepresentasikan kondisi akhir pinjaman. Variabel ini terdiri dari beberapa kategori, berikut merupakan tabel penjelasannya:

Tabel 3. 2 Distribusi Variabel Target (Loan_status)

Status pinjaman	Jumlah	Persentase (%)
Current	603.220	57,53
Fully Paid	331.494	31,62
Charged Off	94.281	8,89
Late (31-120 days)	12.154	1,16
In Grace Period	5.151	0,49
Late (16-30 days)	2.162	0,21
Default	21	0,002

Pada dataset awal, variabel *loan_status* terdiri dari tujuh kategori. Namun untuk keperluan pemodelan, data disaring hanya pada dua kategori utama yang memiliki label paling jelas, yaitu *fully paid* dan *Charger Off*. Kategori lain seperti *Current, Late, In Grace Period,* dan *Default* tidak diikutsertakan karena bersifat ambigu atau belum memiliki status akhir yang pasti. Setelah penyaringan, variabel loan_status dikonversi menjadi variabel biner, di mana nilai 0 menunjukkan pinjaman segar (*Fully Paid*) dan nilai 1 menunjukkan pinjaman gagal bayar (*Charger Off*).

3.2.2 Variabel Karakteristik Pinjaman

Tabel 3.3 di bawah ini menyajikan variabel-variabel yang berkaitan dengan karakteristik pinjaman yang diajukan oleh peminjam.

Tabel 3. 3 Variabel Karakteristik Pinjaman

Variabel	Tipe Data	Deskripsi
loan_amnt	Numerik	Jumlah pinjaman yang diajukan (USD) berkisar $1.000-40.000

Variabel	Tipe Data	Deskripsi
term	Kategorikal	Jangka waktu pinjaman: 36 bulan atau 60 bulan
int_rate	Numerik	Tingkat bunga pinjaman (%), rata-rata 12,80
installment	Numerik	Angsuran bulanan yang dibayarkan peminjam (USD)
purpose	Kategorikal	Tujuan penggunaan pinjaman (debt_consolidation, credit_card, dll)

3.2.3 Variabel Profil Peminjam

Di bawah ini merupakan variabel yang mencerminkan profil demografis dan finansial peminjam.

Tabel 3. 4 Variabel Profil Peminjam

Variabel	Tipe Data	Deskripsi
emp_title	Kategorikal	Pekerjaan peminjam (teks bebas)
empt_length	Ordinal	Lama bekerja: < 1 tahun hingga 10+ tahun.
home_ownership	Kategorikal	Status kepemilikan rumah: MORTGAGE, RENT, OWN, ANY.
annual_inc	Numerik	Pendapatan tahunan peminjam (USD), median $66.000
addr_state	Kategorikal	Negara bagian domisili peminjam di Amerika Serikat.

3.2.4 Variabel Riwayat Kredit

Di bawah ini tabel 3.5 menyajikan variabel yang berkaitan dengan rekam jejak kredit peminjam sebagai indikator kelayakan kredit.

Tabel 3. 5 Variabel Riwayat Kredit Peminjam

Variabel	Tipe Data	Deskripsi
dti	Numerik	Debt-to-income ratio: perbandingan total cicilan utang terhadap pendapatan, rata-rata 19,29%
deling_2yrs	Numerik	Jumlah keterlambatan pembayaran lebih dari 30 hari dalam 2 tahun terakhir.
inq_last_6mths	Numerik	Jumlah permintaan kredit dalam 6 bulan terakhir.
open_acc	Numerik	Jumlah akun kredit yang masih aktif
revol_bal	Numerik	Total saldo kredit bergulir (revolving credit) saat ini rata-rata $16.725
revol_util	Numerik	Rasio penggunaan kredit bergulir terhadap limit yang tersedia. Rata-rata 47.69
total_acc	Numerik	Total jumlah akun kredit yang pernah dimiliki

Variabel	Tipe Data	Deskripsi
		peminjam, rata-rata 23,7 akun.

3.3 Tahapan Penelitian

Penelitian ini dilaksanakan mengikuti kerangka standar data *science* yang terdiri dari enam tahapan utama, sebagaimana di uraikan di bawah ini:

3.3.1 Pengumpulan Data

Dataset Lending Club diperoleh dari repositori publik *Kaggle*. *Dataset* mencakup seluruh transaksi pinjaman yang telah selesai maupun yang masih berjalan, sehingga mewakili kondisi aktual pasar pinjaman P2P. *Dataset* dengan total lebih dari 1 juta observasi dipilih guna memastikan keterwakilan ststistik yang memadai.

3.4 Pengolahan Data

Pengolahan data merupakan tahapan paling penting sebelum masuk ke pemodelan. Data yang tidak di persiapkan dengan baik dapat menyebabkan model untuk menghasilkan prediksi yang keliru atau *bias*. *Di bawah* ini merupakan tabel Data *Cleaning*:

Tabel 3. 6 Data *Cleaning*

No	Proses	Variabel	Keterangan
1	Cek duplikasi	Semua	Tidak ditemukan duplikasi
2	Cek tipe data	Semua	Tipe data sesuai
4	Hapus identifier	LoanID	Tidak digunakan sebagai input
5	Cek anomali	Numerik	Ditangani oleh model

Oleh karena itu, tahapan pengolahan data pada penelitian ini meliputi beberapa langkah, sebagai berikut:

3.4.1 Pembersihan Data (*Data Cleaning*)

Pada tahap ini dilakukan untuk beberapa pengecekan, tahapan ini meliputi:

a. Mengidentifikasi kolom dengan *missing value* sangat tinggi
b. Menghapus fitur yang tidak informatif (misalnya, fitur dengan satu nilai unik).
c. Menghapus baris duplikat.
d. Menyesuaikan format kolom waktu agar terbaca sebagai *datetime*.

Pembersihan data dilakukan agar *dataset* berada dalam kondisi stabil dan konsisten.

3.4.2 Penanganan *Missing Value*

Dataset menunjukkan jumlah kelas *default* lebih sedikit dibanding tidak *default*. Kondisi ini dapat membuat model terlihat "akurasi tinggi:, tetapi sebenarnya kurang dalam mendeteksi kelas *default*. Pada penelitian ini, tidak menggunakan teknik *oversampling* seperti SMOTE. Untuk menjaga evaluasi tetap objektif, penanganan dilakukan melalui:

1. Pembagian data secara stratified agar proporsi kelas tetap terjaga pada data latih dan data uji.
2. Fokus evaluasi pada metrik yang lebih relevan untuk data tidak seimbang, seperti *Precision*, *Recall*, *F1-Score*, dan *ROC-AUC* serta analisis *Confusion Matrix*.

Di bawah ini merupakan tabel Penanganan *Missing Value*:

Tabel 3. 7 Penanganan *Missing Value*

No	Jenis Data	Variabel	Metode	Alasan
1	Numerik	Income, LoanAmount, InterestRate, LoanTerm, DTIRatio, CreditScore, MonthsEmployed, NumCreditLines, Age	Median	Tidak mudah terpengaruh oleh data ekstrem

No	Jenis Data	Variabel	Metode	Alasan
2	Kategorikal	EmplomentType, Education, MaritalStatus, LoanPurpose	Mode	Menjaga Distribusi
3	Biner	HasMoetgage, HasDependents, HasCoSigner	Nilai Dominan	Konsistensi Kelas

3.4.3 *Encoding* Variabel Kategorikal

Karena fitur kategorikal ditampilkan dalam bentuk kode *integer*, diperlukan *encoding* agar model dapat mengenalinya. Teknik yang digunakan adalah *One-Hot Encoding*

Semua kategori diubah menjadi kolom biner sehingga model tidak salah menginterpretasikan nilai kategori sebagai urutan. *Di bawah* ini merupakan proses *encoding* yang dilakukan pada penelitian:

Tabel 3. 8 Proses Encoding Variabel Kategorikal

No	Variabel	Jenis Awal	Metode Encoding	Bentuk Akhir
1	emp_length	Ordinal	Ordinal Encoding	Numerik (0-10)
2	home_ownership	Nominal	Label Encoding	Numerik (0,1,2,..)
3	term	Nominal	Label Encoding	Numerik (0,1)
4	purpose	Nominal	Label Encoding	Numerik (0,1,2,..)
5	loan_status	Target	Label Encoding	Biner (0 atau 1)

3.4.4 Normalisasi Fitur Numerik

Fitur numerik memiliki skala berbeda, jika tidak dinormalisasi model akan dapat memberikan bobot berlebih pada fitur dengan nilai besar. Oleh karena itu digunakan:

StandardScaler

Mengubah nilai menjadi distribusi standar (*mean* = 0, standar deviasi =1).

Normalisasi ini sangat penting terutama untuk algoritma *XGBoost*. *Di bawah* ini merupakan tabel Normalisasi pada *dataset* penelitian:

Tabel 3. 9 Normalisasi

No	Nama Variabel	Jenis Data	Metode Normalisasi
1	loan_amnt	Numerik	*StandarScaler*
2	int_rate	Numerik	*StandarScaler*
3	installment	Numerik	*StandarScaler*
4	annual_inc	Numerik	*StandarScaler*
5	dti	Numerik	*StandarScaler*
6	delinq_2yrs	Numerik	*StandarScaler*
7	inq_last_6mghs	Numerik	*StandarScaler*
8	open_acc	Numerik	*StandarScaler*
9	revol_bal	Numerik	*StandarScaler*
10	revol_util	Numerik	*StandarScaler*
11	total_acc	Numerik	*StandarScaler*

Berdasarkan tabel 3.5 proses normalisasi data difokuskan khusus pada 11 fitur yang memiliki tipe numerik, teknik yang digunakan adalah *StandarScaller* yang berfungsi untuk meratakan dan menyamakan skala seluruh rentang nilai agar lebih seimbang. Langkah normalisasi ini sangat penting untuk dilakukan karena nilai asli pada *dataset* memiliki rentang angka yang sangat jauh berbeda.

3.4.5 Penanganan Ketidakseimbangan Kelas

Dataset yang digunakan menunjukkan bahwa jumlah data pinjaman yang mengalami keterlambatan pembayaran lebih sedikit dibandingkan dengan data pinjaman yang dibayar tepat waktu. Kondisi ini menyebabkan *dataset* bersifat tidak seimbang dan dapat mempengaruhi kinerja model, terutama dalam mendeteksi kasus keterlambatan pembayaran.

Pada penelitian ini, tidak menggunakan teknik *oversampling* seperti SMOTE. Penanganan ketidakseimbangan kelas dilakukan dengan menjaga proporsi data pada saat pembagian *dataset* serta menerapkan strategi evaluasi

yang sesuai. Model dievaluasi tidak hanya berdasarkan nilai akurasi, tetapi juga menggunakan metrik *precision, recall, F1-Score, ROC-AUC,* serta analisis *confusion* metrik untuk melihat bagaimana kemampuan model dalam mendeteksi kelas keterlambatan secara lebih jelas. *Di bawah* ini merupakan tabel penanganan ketidakseimbangan kelas:

Tabel 3. 10 Penanganan Ketidak Seimbangan Kelas

No	Aspek	Strategi	Alasan
1	Distribusi Kelas	*Stratified* Split	Menjaga proporsi kelas tetap sama pada data latih dan uji
2	Teknik Penanganan	SMOTE dan *Class Weight*	Membandingkan performa data buatan (*oversampling*) dengan pembobotan penalti
3	Fokus evaluasi	*Precision, Recall,* F1-*Score, ROC-AUC*	Memberikan penilaian objektif pada kelas minoritas (*default*)
4	Alat Analisis	*Confusion Matrix*	Melihat detail kesalahan prediksi secara mendalam

Penerapan strategi dalam Tabel 3.10 bertujuan agar model tidak hanya condong memprediksi kelas mayoritas (nasabah tepat waktu), tetapi juga memiliki sensitivitas yang baik dalam mendeteksi nasabah yang berisiko mengalami gagal bayar (*default*). Penggunaan *Stratified Split* memastikan bahwa meskipun jumlah data sangat besar, porsi data gagal bayar tetap terwakili secara adil dalam proses pelatihan model.Selain itu, pemilihan metrik seperti F1-*Score* dan *ROC-AUC* memberikan gambaran yang lebih akurat mengenai efektivitas algoritma dibandingkan hanya mengandalkan nilai akurasi semata.

3.4.6 Pembagian *Dataset*

Dataset kemudian dibagi menjadi tiga bagian, yaitu:

a. 80% merupakan data latih
b. 15% merupakan data valid
c. 5% merupakan data uji

Pembagian ini dilakukan menggunakan *stratified split* agar proporsi kelas tetap seimbang pada kedua bagian.

Tabel 3. 11 Pembagian *Dataset*

No	Jenis Data	Persentase	Jumlah Data
1	Data Latih	80%	340.620
2	Data Valid	15%	63.866
3	Data Uji	5%	21.289
4	Total	100%	425.775

3.5 Pelatihan Model

Setelah seluruh data siap, model akan dilatih menggunakan dua algoritma berikut:

3.5.1 *Random Forest*

Random forest merupakan metode *ensemble* yang menggabungkan banyak pohon keputusan (*decision tree*) dengan teknik *bagging* (*boostrap aggregating*). Setiap pohon dibangun dari sampel acak data dan subset fitur acak. Hasil prediksi akhir diperoleh melalui voting mayoritas dari seluruh pohon. Kelebihannya adalah, stabil terhadap *noise* dan tidak rentan terhadap *overfitting*. Pada penelitian ini, *Random Forest* diuji dengan dua skenario penanganan ketidakseimbangan kelas, yaitu menggunakan SMOTE dan menggunakan *class_weight*=’*balanced*’.

3.5.2 XGBoost

XGBoost adalah metode *ensemble learning* dalam kelompok gradient *boosting* yang membangun model secara berurutan, di mana setiap pohon baru memperbaiki kesalahan prediksi dari pohon sebelumnya. *XGBoost* memiliki mekanisme regulasi bawaan (L1dan L2) untuk mencegah *overfitting* serta mendukung pemrosesan paralel. Pada peneliatian ini, *XGBoost* diuji dengan tiga skenario, yaitu menggunakan SMOTE, menggunakan *scale_pos_weight* dan menggunakan *GridSearchCV* untuk optimasi *hiperparameter*.

3.5.3 Optimasi HiperParameter (*GridSearchCV*)

Optimasi hiperparameter dilakukan menggunakan *GridSraechCV* dengan strategi *cross-validation* 5-fold. *GridSearchCV* melakukan dan memilih

kombinasi yang menghasilkan nilai *F1-Score* terbaik untuk kelas 1 (*default*). Pendekatan ini memastikan bahwa model yang dihasilkan memiliki performa yang optimal dan etrhindar dari *overfitting*.

3.5.4 Threshold Tuning

Threshold Tuning dilakukan untuk menemukan ambang batas keputusan (*deccision threshold*) yang optimal bagi setiap model. Secara *default*, model klasifikasi meggunakan *threshold* 0,5 namun pada data yang tidak seimbang nilai tidak ada selalu optimal. *Threshold tuning* dilakukan dengan mengevaluasi F1-*Score* kelas 1 pada berbagai nilai *threshold* menggunakan data validasi, kemudian *threshold* terbaik diterapkan pada prediksi akhir dan uji.

Tabel 3. 12 Pelatihan Model

No	Model	Fungsi	Alasan
1	*Random Forest*	Pembanding	Stabil & tahan *noise*
2	*XGBoost*	Model utama	Akurasi tinggi & efisien

3.6 Evaluasi Model

Evaluasi dilakukan untuk memastikan bahwa model yang dibangun benar-benar mampu memprediksi keterlambatan pembayaran secara akurat. Pada tahap ini, sejumlah metrik digunakan untuk menilai kualitas prediksi model. *Accuracy* digunakan untuk melihat tingkat ketepatan model secara keseluruhan dalam mengklasifikasikan data ke dalam kategori tepat waktu atau terlambat. *Precision* digunakan untuk menilai seberapa tepat model dalam mengidentifikasi data yang diprediksi sebagai terlambat, sehingga dapat diketahui seberapa besar kemungkinan model memberikan prediksi positif yang benar. Sementara itu, *Recall* digunakan untuk mengukur kemampuan model dalam menemukan seluruh kasus yang benar-benar masuk kategori terlambat, sehingga metrik ini sangat penting pada situasi yang menuntut deteksi dini keterlambatan. Karena *Precision* dan *Recall* sering kali memiliki kecenderungan berlawanan, digunakan pula *F1-Score* sebagai ukuran keseimbangan antara keduanya. Selain itu, *Confusion Matrix* dianalisis untuk memahami jenis kesalahan yang dilakukan model, seperti salah memprediksi data

tepat waktu sebagai terlambat maupun sebaliknya. Seluruh matriks ini kemudian dibandingkan pada kedua algoritma, sehingga dapat ditentukan model mana yang memberikan kinerja terbaik dan paling layak untuk digunakan dalam prediksi keterlambatan pembayaran.

Tabel 3. 13 Evaluasi Model

No	Metrik	Fungsi
1	*Accuracy*	Ketepatan prediksi secara keseluruhan
2	*Precision* (kelas 1*)*	Ketepatan model saat memprediksi *default*
3	*Recall (*kelas 1*)*	Kemampuan mendeteksi seluruh kasus *default*
4	*F1-Score (*kelas 1*)*	Keseimbangan antara *precision* dan *recall*
5	*ROC AUC*	Kemampuan diskriminasi model antar kelas
6	*Confusion Matrix*	Analisis detail TP, TN, FP, dan FN

Karena data bersifat tidak seimbang, fokus evaluasi utama diberikan pada *F1-Score* kelas 1 dan *ROC AUC*, bukan hanya pada nilai akurasi keseluruhan. *Confusion Matrix* juga dianalisis untuk memahami detail kesalahan prediksi model (*False Positive* dan *False Negative)*.

3.7 Hasil Yang Diharapkan

Penelitian ini diharapkan mampu menghasilkan model *machine learning* yang akurat dalam memprediksi potensi keterlambatan pembayaran pada layanan pinjaman *online* (*Lending Club*). Dari proses komparasi berbagai skenario pemodelan diharapkan dapat diketahui kombinasi algoritma dan teknik penanganan ketidakseimbangan kelas yang memberikan performa paling optimal. Selain itu, penelitian ini diharapkan dapat menunjukkan bahwa penerapan teknik seperti SMOTE, *class weight*, *GridSearchCV*, dan *threshold tuning* secara signifikan meningkatkan kemampuan model dalam mendeteksi kasus *default* dibandingkan model *baseline*.