

## BAB II

### LANDASAN TEORI

#### 2.1 Penelitian Terdahulu

Dalam lima tahun terakhir, penelitian mengenai penentuan dan penyesuaian limit kredit pada sektor perbankan menunjukkan perkembangan signifikan melalui pemanfaatan *machine learning* dan metode optimisasi untuk meningkatkan akurasi prediksi serta efektivitas pengelolaan risiko. Sejumlah studi juga mengimplementasikan *reinforcement learning* guna menghasilkan kebijakan penyesuaian limit yang adaptif terhadap dinamika perilaku nasabah dan perubahan kondisi ekonomi. Ringkasan penelitian terdahulu yang relevan disajikan pada Tabel 2.1 berikut.

Tabel 2.1 Penelitian Terdahulu

No	Penulis	Judul, Tahun	Masalah	Metode	Kontribusi	Hasil (Akurasi/ Performa Model)
1	Chang et al. [11]	<i>Credit Risk Prediction Using ML and DL: A Study on Credit Card Customers</i> , 2024	Klasifikasi <i>default</i> nasabah kartu kredit.	NN, LR, AdaBoost, XGBoost, LightGBM	Perbandingan menyeluruh berbagai model ML & DL.	XGBoost unggul dengan akurasi 99,4 %.
2	Akinjole et al. [10]	<i>Ensemble-Based ML Algorithm for Loan Default Risk Prediction</i> , 2024	Prediksi risiko <i>loan default</i> .	RF, XGBoost, MLP, SVM + Stacking	Evaluasi model individual & stacking pada dataset LendingClub.	Stacking akurasi 93,7 %, AUC 97,8 %.
3	Gao et al. [12]	<i>Research on Default Prediction for Credit Card Users Based on XGBoost-LSTM</i> , 2021	Prediksi gagal bayar berbasis perilaku transaksi.	XGBoost, LSTM, Hybrid XGBoost-LSTM	Kombinasi model sekuensial & tree-based untuk data tidak seimbang.	Hybrid LSTM > XGBoost pada data imbalanced.

No	Penulis	Judul, Tahun	Masalah	Metode	Kontribusi	Hasil (Akurasi/ Performa Model)
4	Xu et al. [13]	<i>Novel Embedding Model Predicting Credit Card Default...</i> , 2024	Peningkatan NN dengan optimasi evolusioner.	MLP + HS/VS/MVO/WOA	Optimasi hiper-parameter menggunakan metaheuristik.	AUC terbaik $\pm 0,73$ .
5	Li et al. [14]	<i>Research on Credit Card Default Repayment Prediction Model</i> , 2024	Prediksi kemampuan repayment pasca-default.	Tree Models, BPNN, ELM, XGBoost + EMD	Kerangka pemodelan berbasis frekuensi data.	ELM & XGBoost terbaik ( $R^2 \approx 0,75$ ).
6	Fan, Shen, Peng. [15]	<i>Improved ML-Based Technique for Credit Card Scoring in Internet Financial Risk Control</i> , 2020	Penilaian risiko kredit kartu menggunakan model ML	LR, RF, GBM, XGBoost, LightGBM	Mengembangkan pendekatan credit card scoring berbasis ML untuk kontrol risiko internet finance	LightGBM memberikan performa terbaik dalam klasifikasi risiko kredit
7	Xiao et al. [16]	<i>A Novel Deep Ensemble Model for Imbalanced Credit Scoring in Internet Finance</i> , 2023	Credit scoring pada data imbalanced	Deep Neural Networks, Ensemble Learning, Cost-Sensitive Learning	Mengusulkan deep ensemble model untuk menangani ketidakseimbangan data kredit	Model ensemble meningkatkan performa prediksi pada data kredit tidak seimbang
8	Tian et al. [17]	<i>Credit Risk Assessment based on Gradient Boosting Decision Trees</i> , 2020	Model kredit tradisional kurang akurat untuk pola non-linier.	Gradient Boosting Decision Trees.	Menunjukkan GBDT lebih unggul dari model klasik untuk credit scoring.	GBDT memberi akurasi dan AUC tertinggi dibanding model pembanding.

No	Penulis	Judul, Tahun	Masalah	Metode	Kontribusi	Hasil (Akurasi/ Performa Model)
9	Chang et al. [18]	<i>Machine learning and artificial neural networks to construct P2P lending credit-scoring model</i> , 2022	Skor kredit P2P bawaan platform kurang memisahkan default/non-default.	Berbagai ML (LR, RF, boosting, ANN).	Membangun model credit scoring P2P berbasis ML/ANN.	Model ensemble /ANN melampaui skor kredit konvensional (AUC dan akurasi lebih tinggi).
10	Nguyen & Ngo [19]	<i>Comparative analysis of boosting algorithms for predicting personal default</i> , 2025	Belum jelas algoritma boosting terbaik untuk default personal.	LightGBM, XGBoost, CatBoost, AdaBoost.	Analisis komparatif berbagai boosting untuk default.	LightGBM memperoleh accuracy dan AUC tertinggi.
11	Prahastiwati et al. [20]	<i>The Ensemble Supervised Machine Learning for Credit Scoring at Digital Banking using SHAP-based Interpretability</i> , 2025	Butuh credit scoring digital banking yang akurat dan interpretabile.	RF, AdaBoost, LightGBM, CatBoost, XGBoost + SHAP.	Menggabungkan ensemble dan SHAP untuk credit scoring bank digital.	XGBoost memberi ROC-AUC tertinggi dan berpotensi menurunkan NPL.

No	Penulis	Judul, Tahun	Masalah	Metode	Kontribusi	Hasil (Akurasi/ Performa Model)
12	Odunlami [21]	<i>Credit Risk Prediction using Ensemble and Linear Machine Learning Models</i> , 2025	Risiko kredit dengan data imbalance dan non-linier.	LR, NB, SVM, RF, XGBoost, LightGBM.	Membandingkan ensemble vs model linear untuk credit risk.	LightGBM/XGBoost dengan resampling memberi metrik terbaik.
13	Pertiwi et al. [22]	<i>A Performance Comparison of Data Balancing Model to Improve Credit Risk Prediction in Peer-to-Peer Lending</i> , 2024	Imbalance kelas tinggi pada data P2P menurunkan performa.	Balancing (ROS, RUS, SMOTE) + LightGBM/XGBoost.	Menilai pengaruh teknik balancing pada ensemble credit risk.	Kombinasi SMOTE + LightGBM/XGBoost mencapai akurasi tertinggi.
14	S. R. Tambunan, et al. [23]	<i>Enhancing Credit Risk Classification Using LightGBM with Deep Feature Synthesis</i> , 2021	Credit risk classification pada P2P lending dengan banyak atribut yang kompleks dan klasifikasi akurat sulit dicapai	LightGBM dengan dan tanpa Deep Feature Synthesis (DFS) untuk seleksi dan rekayasa fitur	Menunjukkan penggunaan DFS dengan LightGBM meningkatkan performa model serta kualitas fitur secara otomatis untuk prediksi	Akurasi 0.99 dengan DFS vs 0.97 tanpa DFS; recall dan F1-score meningkat signifikan dengan DFS (recall 0.94 vs 0.68, F1-score

No	Penulis	Judul, Tahun	Masalah	Metode	Kontribusi	Hasil (Akurasi/ Performa Model)
					risiko kredit	0.96 vs 0.81)
15	Zhang & Liu [24]	<i>Personal Loan Default Prediction Based on LightGBM Model and Zhima Credit</i> , 2023	Prediksi default pinjaman personal pada lembaga keuangan.	LightGBM (klasifikasi default/non-default) dengan fitur Zhima Credit.	Mengembangkan model credit risk berbasis LightGBM untuk memprediksi default pinjaman personal dan menunjukkan kegunaannya bagi keputusan risiko kredit.	Model LightGBM dilaporkan memberikan performa klasifikasi yang baik dan efektif untuk memperkirakan probabilitas default nasabah.

Berdasarkan Tabel 2.1, terlihat bahwa Pendekatan *machine learning* dalam kajian risiko kredit pada tahap awal berkembang melalui perbandingan antara model linear dan non-linear pada data kartu kredit. Hasil penelitian menunjukkan bahwa algoritma *tree-based ensemble* seperti XGBoost mampu mengungguli *neural network* dan *logistic regression* dalam memodelkan hubungan yang kompleks antar variabel risiko kredit [11]. Seiring dengan meningkatnya kompleksitas data pinjaman, penelitian selanjutnya mulai mengadopsi teknik *ensemble* lanjutan, seperti *stacking*, untuk menggabungkan keunggulan beberapa model sekaligus, yang terbukti meningkatkan nilai akurasi dan *AUC* dalam prediksi *loan default* dibandingkan model individual [10]. Pada konteks data transaksi yang bersifat temporal dan tidak seimbang, pendekatan hibrida yang menggabungkan XGBoost dengan *Long Short-Term Memory* (LSTM) dikembangkan untuk

menangkap pola sekuensial, dan dilaporkan menghasilkan performa yang lebih baik dibandingkan model boosting tunggal [12].

Perkembangan penelitian berikutnya menyoroti keterbatasan model tunggal dalam menghadapi variasi karakteristik data kredit, sehingga fokus penelitian mulai bergeser ke arah optimasi parameter dan struktur model. Beberapa studi menerapkan algoritma metaheuristik untuk mengoptimalkan *neural network* dalam prediksi default kartu kredit, meskipun hasil evaluasi menunjukkan bahwa peningkatan performa sangat bergantung pada teknik optimasi yang digunakan [13]. Selain itu, ruang lingkup penelitian juga diperluas dari sekadar klasifikasi default menuju prediksi kemampuan pembayaran setelah terjadinya default, di mana model berbasis *boosting* dan *tree-based learning* menunjukkan kinerja yang relatif stabil dalam memodelkan risiko kredit lanjutan [14]. Dalam konteks *credit scoring*, algoritma LightGBM dan XGBoost dilaporkan mampu memberikan hasil klasifikasi yang lebih baik dibandingkan pendekatan tradisional, khususnya pada data kredit berskala besar dan *heterogen* [15].

Seiring meningkatnya adopsi kredit digital dan platform *peer-to-peer* (P2P) *lending*, permasalahan ketidakseimbangan kelas (*class imbalance*) menjadi perhatian utama dalam penelitian risiko kredit. Pendekatan *deep ensemble* dan *cost-sensitive learning* dikembangkan untuk mengurangi bias terhadap kelas mayoritas dan meningkatkan kemampuan generalisasi model pada data kredit yang tidak seimbang [16]. Penelitian lain menunjukkan bahwa *Gradient Boosting Decision Trees* (GBDT) mampu menangkap pola non-linier yang tidak dapat dimodelkan secara efektif oleh metode kredit konvensional [17]. Studi pada platform P2P *lending* juga menunjukkan bahwa model *machine learning* dan *artificial neural network* (ANN) mampu melampaui skor kredit bawaan platform dalam membedakan nasabah *default* dan *non-default* [18]. Analisis komparatif algoritma *boosting* terbaru semakin menguatkan temuan bahwa LightGBM cenderung memberikan nilai akurasi dan *AUC* tertinggi dalam prediksi *personal default* [19].



Dalam perkembangan mutakhir, fokus penelitian tidak hanya diarahkan pada peningkatan performa model, tetapi juga pada aspek interpretabilitas dan kualitas fitur. Integrasi metode interpretabilitas seperti SHAP pada model *ensemble* memungkinkan analisis kontribusi variabel risiko yang lebih transparan, khususnya dalam konteks *credit scoring* perbankan digital [20]. Penelitian lain menunjukkan bahwa model *ensemble* tetap unggul dibandingkan model linear ketika diterapkan pada data kredit yang bersifat non-linier dan tidak seimbang [21]. Selain itu, teknik data balancing seperti *Synthetic Minority Over-sampling Technique* (SMOTE) yang dikombinasikan dengan LightGBM dan XGBoost dilaporkan mampu meningkatkan metrik evaluasi secara signifikan pada data P2P *lending* [22]. Pendekatan *feature engineering* otomatis melalui *Deep Feature Synthesis* juga terbukti meningkatkan kemampuan LightGBM dalam mengklasifikasikan risiko kredit secara lebih konsisten [23]. Studi pada pinjaman personal menunjukkan bahwa LightGBM efektif dalam memperkirakan probabilitas *default* ketika dikombinasikan dengan fitur kredit alternatif [24].

Meskipun berbagai penelitian telah mengkaji penerapan *ensemble boosting* untuk prediksi *default*, *credit scoring*, dan penilaian risiko kredit pada konteks kartu kredit, P2P *lending*, maupun pinjaman personal, fokus kajiannya masih terbatas pada pemeringkatan risiko atau klasifikasi gagal bayar secara umum [11], [14], [18]. Upaya peningkatan performa melalui teknik *ensemble*, data *balancing*, optimasi parameter, serta *feature engineering* telah banyak dilakukan, namun evaluasi yang dilaporkan umumnya belum diarahkan pada pengelompokan limit kredit yang bersifat operasional di lingkungan perbankan [13], [16], [22], [23]. Selain itu, meskipun pendekatan interpretabilitas seperti SHAP mulai digunakan, pemanfaatannya masih difokuskan pada analisis risiko kredit secara umum dan belum secara spesifik mendukung proses klasifikasi kategori limit kredit [20], [21]. Berdasarkan kondisi tersebut, penelitian ini mengkaji penerapan *ensemble gradient boosting* untuk klasifikasi kategori limit kredit nasabah (*Low*, *Medium*, dan *High*) sebagai alat bantu analisis dalam proses penetapan limit kredit di perbankan.

## 2.2 Teori yang Berkaitan

### 2.2.1 Kredit

Menurut Undang-Undang Republik Indonesia Nomor 4 Tahun 2023 tentang Pengembangan dan Penguatan Sektor Keuangan (PPSK), kredit merupakan penyediaan dana atau tagihan yang dipersamakan dengan itu, yang diberikan berdasarkan perjanjian atau kesepakatan pinjam-meminjam antara bank atau lembaga pembiayaan dengan pihak lain, di mana pihak peminjam berkewajiban untuk melunasi utang tersebut dalam jangka waktu tertentu dengan imbal hasil berupa bunga atau bentuk imbalan lainnya. Kredit memiliki fungsi strategis dalam perekonomian sebagai instrumen pembiayaan, mendorong pertumbuhan konsumsi, dan memperkuat likuiditas sektor usaha maupun rumah tangga [25].

### 2.2.2 Limit Kredit

Limit kredit adalah batas maksimum fasilitas dana yang dapat digunakan oleh nasabah sesuai perjanjian kredit yang telah disepakati, baik dalam bentuk kartu kredit, kredit rekening koran, maupun fasilitas pinjaman lainnya. penetapan limit kredit harus mempertimbangkan profil risiko nasabah, pendapatan, riwayat pembayaran, tingkat utilisasi, dan hubungan nasabah dengan bank [26]. Dalam praktik manajemen risiko modern, limit kredit sering kali tidak hanya dihitung sebagai satu nilai tertentu, tetapi juga diklasifikasikan ke dalam beberapa kategori limit (*Low, Medium, High*) untuk menyederhanakan kebijakan produk, penetapan strategi penagihan, serta alokasi sumber daya *marketing* dan *collection*.

### 2.2.3 Resiko Kredit

Risiko kredit adalah potensi kerugian yang dialami bank akibat kegagalan debitur untuk memenuhi kewajiban pembayaran sesuai perjanjian. Risiko kredit merupakan salah satu dari delapan jenis risiko utama yang wajib dikelola secara efektif melalui proses identifikasi, pengukuran, pemantauan, dan pengendalian yang berkesinambungan [26]. Dalam kerangka Basel, pengukuran risiko kredit umumnya diformalkan melalui indikator *Probability of Default* (PD), *Loss Given Default* (LGD), dan *Exposure at Default* (EAD) yang digunakan untuk mengestimasi besaran potensi kerugian finansial pada saat terjadi gagal bayar.



Pengaturan limit kredit kemudian perlu diselaraskan dengan profil risiko debitur, karena limit yang terlalu tinggi pada debitur berisiko akan meningkatkan eksposur kerugian, sedangkan limit yang terlalu rendah pada debitur berisiko rendah dapat membatasi potensi pendapatan bank, sehingga kebijakan limit kredit harus didasarkan pada evaluasi risiko yang komprehensif dan tetap mematuhi prinsip kehati-hatian. [27]

## 2.3 Framework dan Algoritma yang digunakan

### 2.3.1 CRISP-DM



Gambar 2.1 Diagram Tahapan CRISP-DM

Sumber: [28]

CRISP-DM adalah kerangka kerja standar yang digunakan secara luas dalam proses pengembangan model analitik dan *data mining*. Kerangka ini telah diadopsi secara global karena sifatnya yang fleksibel dan industri-agnostik. CRISP-DM membagi proses pengembangan menjadi enam tahap yang saling berinteraksi secara iteratif, memungkinkan penyesuaian pada setiap tahap sesuai kebutuhan proyek [29] Enam tahap dalam CRISP-DM adalah sebagai berikut:

#### a. **Business Understanding**

Memahami tujuan bisnis dan menerjemahkannya menjadi tujuan analitik yang terukur. Pada penelitian ini, tujuan bisnis adalah menghasilkan model machine learning yang dapat

merekomendasikan kategori limit kredit nasabah PT XYZ secara lebih objektif dan konsisten berdasarkan profil risiko dan karakteristik keuangan.

**b. Data Understanding**

Mengumpulkan data awal dan memahami karakteristiknya melalui eksplorasi deskriptif. Tahap ini mencakup pemeriksaan struktur dataset batch berisi sekitar 12.000 baris dan lebih dari 45 variabel yang berkaitan dengan demografi nasabah, riwayat transaksi, perilaku pembayaran, serta informasi rekening dan limit kredit.

**c. Data Preparation**

Menyiapkan data untuk proses pemodelan melalui pembersihan dan transformasi. Langkah ini mencakup penanganan missing value, identifikasi dan penanganan outlier menggunakan IQR capping, encoding variabel kategorikal, scaling variabel numerik, serta pembentukan fitur tambahan (*feature engineering*) yang relevan untuk klasifikasi kategori limit kredit.

**d. Modeling**

Membangun model machine learning berbasis klasifikasi terawasi (*supervised classification*) untuk memprediksi kategori limit kredit (*Low, Medium, High*). Model yang digunakan meliputi Gradient Boosting, XGBoost, dan LightGBM, yang diintegrasikan ke dalam pipeline preprocessing. Setiap model dilatih dan divalidasi menggunakan skema *5-fold cross-validation*, dengan pembagian data 80:20 antara data pelatihan dan data pengujian, untuk memastikan model dapat melakukan generalisasi yang baik pada data yang belum pernah dilihat sebelumnya.

**e. Evaluation**

Mengevaluasi performa model menggunakan metrik evaluasi klasifikasi yang mencakup Accuracy, Precision, Recall, F1-Weighted, dan F1-Macro, serta analisis confusion matrix untuk

menilai ketepatan prediksi pada masing-masing kategori limit kredit. Hasil evaluasi dibandingkan untuk menentukan model yang paling stabil dan konsisten.

f. **Deployment**

Mengimplementasikan model terbaik dalam bentuk prototipe aplikasi berbasis *Streamlit* untuk memberikan prediksi kategori limit kredit secara batch. Tahap ini bersifat demonstratif dan ditujukan sebagai rekomendasi bagi PT XYZ mengenai bagaimana model dapat digunakan dalam proses evaluasi dan penyesuaian limit kredit secara lebih adaptif.

### 2.3.2 XGBoost (Extreme Gradient Boosting)

XGBoost adalah pengembangan dari algoritma *gradient boosting* yang dioptimalkan untuk kinerja tinggi dan efisiensi memori. Algoritma ini membangun model secara bertahap dengan menambahkan pohon keputusan yang baru untuk memperbaiki kesalahan prediksi model sebelumnya. Keunggulan XGBoost terletak pada dukungannya terhadap regularisasi (*L1* dan *L2*) untuk menghindari *overfitting*, kemampuan menangani *missing value*, serta skalabilitas pada dataset berukuran besar [30].

Dalam penelitian ini, XGBoost digunakan untuk memprediksi nilai limit kredit ideal (*desired credit limit*) dan kategori penyesuaian limit. Algoritma ini dipilih karena performanya yang konsisten pada berbagai studi terdahulu di bidang risiko kredit, kemampuannya memodelkan hubungan non-linear antar variabel, serta memberikan interpretasi melalui analisis *feature importance*. Fungsi objektif pada XGBoost terdiri dari dua komponen utama, yaitu fungsi *loss* dan fungsi regularisasi:

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

Rumus 2.1 Rumus Fungsi Objektif pada Model XGBoost

Sumber: [30]

Dengan

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

Rumus 2.2 Rumus Regularisasi pada Model XGBoost

Sumber: [30]

Keterangan:

- $l(y_i, y^i)$ ,  $l(y_i, \hat{y}_i)$ ,  $l(y_i, y_i)$  : fungsi *loss* (misalnya *squared error* untuk regresi atau *logistic loss* untuk klasifikasi)
- $\Omega(f_k)$ ,  $\Omega(f_k)$ ,  $\Omega(f_k)$  : fungsi regularisasi untuk mengontrol kompleksitas model
- $T$  : jumlah daun (*leaves*) pada pohon
- $w_j$  : bobot pada daun ke-j
- $\gamma, \lambda$  : parameter regularisasi

### 2.3.3 Light Gradient Boosting Machine (LGBM)

Light Gradient Boosting Machine (LightGBM) adalah algoritma machine learning berbasis *ensemble boosting* yang dikembangkan oleh Microsoft untuk meningkatkan efisiensi dan kecepatan pemodelan data berskala besar. LightGBM menggunakan pendekatan gradient boosting framework, di mana model dibangun secara bertahap dengan menambahkan pohon keputusan baru untuk memperbaiki kesalahan dari pohon sebelumnya. Tidak seperti algoritma boosting konvensional, LightGBM menerapkan teknik leaf-wise growth dengan depth limit, yang memungkinkan model mempartisi data berdasarkan daun dengan gain terbesar, sehingga menghasilkan akurasi yang lebih tinggi dan konvergensi lebih cepat [31].

Dalam penelitian ini, LightGBM digunakan sebagai model utama untuk melakukan klasifikasi kategori limit kredit nasabah (*Low, Medium, High*). Pemilihan LightGBM didasarkan pada kemampuannya dalam menangani data berukuran besar dengan tipe fitur campuran (numerik dan kategorikal), serta

efisiensinya dalam hal waktu pelatihan dan penggunaan memori. Model ini juga mendukung proses hyperparameter tuning seperti pengaturan jumlah pohon ( $n\_estimators$ ), tingkat pembelajaran ( $learning\_rate$ ), dan kedalaman maksimum pohon ( $max\_depth$ ) untuk mencapai kinerja optimal.

Fungsi tujuan (*objective function*) yang digunakan dalam LightGBM untuk klasifikasi dapat dinyatakan sebagai penjumlahan antara loss function dan regularization term berikut:

$$Obj(t) = \sum_{i=1}^n l\left(y_i, \widehat{y_i^{(t-1)}} + f_t(x_i)\right) + \Omega(f_t)$$

Rumus 2.3 Fungsi Objektif XGBoost pada Iterasi ke-t

Sumber: [31]

dengan

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

Rumus 2.4 Rumus Regularisasi pada Pohon ke-t dalam XGBoost

Sumber: [31]

Keterangan:

- $l(y_i, \hat{y}_i)$  : fungsi kerugian antara nilai aktual dan prediksi
- $f_t(x_i)$  : pohon keputusan baru pada iterasi ke-t
- $\Omega(f_t)$  : fungsi regularisasi untuk mengontrol kompleksitas model
- $T$  : jumlah daun pada pohon
- $w_j$  : bobot output pada daun ke-j
- $\gamma, \lambda$  : parameter regulasi untuk menghindari overfitting

#### 2.3.4 Gradient Boosting

Gradient Boosting adalah algoritma ensemble berbasis metode boosting yang membangun model prediktif secara bertahap dengan menambahkan pohon keputusan baru untuk memperbaiki kesalahan dari model sebelumnya. Setiap

pohon dilatih untuk meminimalkan loss function dari prediksi kumulatif model sebelumnya dengan menggunakan pendekatan gradient descent. Dengan demikian, model akhir merupakan kombinasi dari sejumlah pohon lemah (*weak learners*) yang bekerja sama untuk menghasilkan prediksi yang lebih akurat dan stabil [32].

Dalam penelitian ini, Gradient Boosting digunakan bersama dengan XGBoost dan LightGBM untuk membandingkan performa klasifikasi dalam menentukan kategori limit kredit nasabah. Algoritma ini dipilih karena kemampuannya dalam menangkap hubungan non-linear dan memberikan keseimbangan antara akurasi serta kemampuan generalisasi, meskipun dengan tuning parameter yang relatif sederhana.

Rumus umum prediksi pada Gradient Boosting dapat dituliskan sebagai berikut:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

Rumus 2.5 Rumus Pembaruan Model pada Proses Gradient Boosting

$$\gamma_m = \arg \min_{\gamma} \sum_{i=1}^n L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i))$$

Rumus 2.6 Penentuan Nilai Optimal  $\gamma$  pada Iterasi ke- $m$

$$F_0(x) = \arg \min_{\gamma} \sum_{i=1}^n L(y_i, \gamma)$$

Rumus 2.7 Model Awal (Baseline) dalam Gradient Boosting

Sumber: [32]

Dengan:

- $F_m(x)$  merupakan model prediksi pada iterasi ke- $m$ .
- $F_{m-1}(x)$  adalah model kumulatif dari iterasi sebelumnya.
- $h_m(x)$  menunjukkan pohon keputusan lemah (*weak learner*) yang dilatih berdasarkan nilai residual dari model sebelumnya.



- $\gamma_m$  merupakan koefisien pembobot yang diperoleh dari proses minimisasi fungsi kerugian.
- $L(y_i, \hat{y}_i)$  adalah fungsi kerugian yang mengukur perbedaan antara nilai aktual dan nilai prediksi.
- $n$  menyatakan jumlah data pelatihan yang digunakan dalam proses pembelajaran model.

### 2.3.5 Confusion Matrix

*Confusion matrix* adalah metode evaluasi yang digunakan untuk mengukur performa model klasifikasi dengan membandingkan hasil prediksi model terhadap nilai aktual. Pada klasifikasi biner, confusion matrix berbentuk 2x2 dan terdiri dari empat komponen utama, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Komponen-komponen ini memberikan gambaran jelas mengenai jenis prediksi yang dilakukan oleh model, sehingga memudahkan dalam menganalisis kelebihan dan kekurangan model [33].

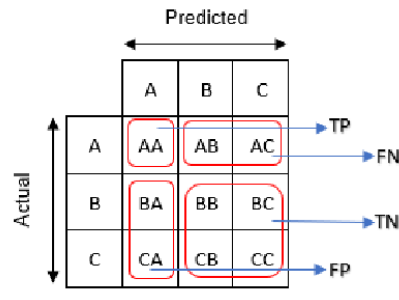
		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Gambar 2.2 *Confusion Matrix*

Sumber: [34]

Pada Gambar 2.1, *True Positive* (TP) merepresentasikan data positif yang diprediksi benar, sedangkan *True Negative* (TN) menunjukkan data negatif yang diprediksi dengan tepat. *False Positive* (FP) menggambarkan data negatif yang salah diklasifikasikan sebagai positif, dan *False Negative* (FN) menunjukkan data positif yang salah diprediksi sebagai negatif.

Selain klasifikasi biner, *confusion matrix* juga digunakan untuk klasifikasi multikelas, misalnya pada kasus tiga kelas seperti rendah, medium, dan tinggi. Pada konteks ini, confusion matrix berbentuk 3x3, di mana setiap baris menunjukkan kelas aktual dan setiap kolom menunjukkan kelas hasil prediksi.



Gambar 2.3 Confusion Matrix (3x3)

Sumber: [35]

Pada confusion matrix 3x3, elemen diagonal (AA, BB, CC) merupakan prediksi benar untuk masing-masing kelas. Elemen-elemen di luar diagonal menunjukkan kesalahan prediksi antar kelas, seperti contoh AC yang menunjukkan kelas A diprediksi sebagai C. Struktur ini memungkinkan analisis lebih rinci terhadap performa model di tiap kelas, terutama pada permasalahan klasifikasi yang memiliki lebih dari dua kategori [35].

Confusion matrix, baik 2x2 maupun 3x3, juga memungkinkan perhitungan berbagai metrik seperti accuracy, precision, recall, dan F1-score yang memberikan pemahaman mendalam mengenai kualitas model, khususnya ketika menangani dataset yang tidak seimbang.

### 2.3.6 Precision

*Precision* adalah metrik yang mengevaluasi ketepatan model dalam proses mengenali data pada kelas positif secara tepat dibandingkan dengan seluruh prediksi positif yang dibuat oleh model. *Precision* dihitung dengan membagi jumlah *True Positive* (TP) dengan total prediksi positif, yaitu penjumlahan antara *True Positive* (TP) dan *False Positive* (FP). Metrik ini memberikan gambaran tentang proporsi prediksi positif yang benar-benar sesuai dengan nilai

aktual [33]. Rumus 2.4 berikut adalah rumus yang digunakan untuk menghitung *precision*:

$$Precision = \frac{TP}{TP + FP}$$

Rumus 2.8 Rumus *Precision*

Sumber: [33]

Keterangan:

1. TP (*True Positive*): Jumlah data yang secara benar diprediksi sebagai positif.
2. FP (*False Positive*): Jumlah data negatif yang salah diklasifikasikan sebagai positif.

### 2.3.7 Recall

*Recall* merupakan metrik evaluasi yang berfungsi untuk menilai sejauh mana model mampu mendeteksi kelas positif secara benar dari seluruh data aktual kelas positif. Nilai *recall* dihitung dengan membagi jumlah *True Positive* (TP) dengan total data aktual positif, yang merupakan penjumlahan dari *True Positive* (TP) dan *False Negative* (FN). Metrik ini memberikan informasi mengenai kemampuan model dalam mengidentifikasi data positif secara menyeluruh [33]. Rumus 2.5 berikut adalah rumus yang digunakan untuk menghitung *Recall*:

$$Recall = \frac{TP}{TP + FN}$$

Rumus 2.9 Rumus *Recall*

Sumber: [33]

Keterangan:

1. TP (*True Positive*): Jumlah prediksi positif yang benar oleh model.
2. FN (*False Negative*): Jumlah data aktual positif yang salah diprediksi sebagai negatif oleh model.

### 2.3.8 F-Measure

*F-Measure*, lebih dikenal sebagai *F1-Score*, adalah metrik evaluasi yang evaluasi yang mengintegrasikan *precision* dan *recall* ke dalam satu nilai harmonis. Metrik ini dihitung sebagai rata-rata harmonis dari *precision* dan *recall* [36]. Rumus 2.6 berikut adalah rumus yang digunakan untuk menghitung *F1-Score*:

$$F1\ Score = 2 \times \frac{Recall \times Precision}{Recall + Precision}$$

Rumus 2.10 Rumus *F1-Score*

Sumber: [36]

Keterangan:

1. *Recall*: Mengukur sensitivitas model, yaitu seberapa baik model dapat mendeteksi semua instance positif dalam data aktual.
2. *Precision*: Mengukur akurasi model dalam memprediksi kelas positif.
3. *F1-Score*: Nilai metrik yang menyeimbangkan kedua aspek di atas.

Selain itu, terdapat juga dua varian *F1-Score* yang sering digunakan untuk evaluasi model dalam konteks klasifikasi multikelas:

1. *F1-Weighted*: *F1-Score* berbobot, menghitung rata-rata *F1-Score* dengan memperhitungkan jumlah instance pada setiap kelas, sehingga kelas yang lebih besar memiliki dampak yang lebih besar pada skor akhir.
2. *F1-Macro*: *F1-Score* rata-rata sederhana, menghitung *F1-Score* untuk setiap kelas secara terpisah, kemudian menghitung rata-ratanya tanpa mempertimbangkan jumlah instance setiap kelas, memberikan bobot yang sama pada setiap kelas.

### 2.3.9 Accuracy

*Accuracy* adalah metrik evaluasi yang digunakan untuk menilai proporsi prediksi yang sesuai dari keseluruhan data. Nilai ini dihitung dengan membagi total prediksi yang benar, yaitu *True Positive* (TP) dan *True Negative* (TN), dengan jumlah keseluruhan data. Metrik ini memberikan gambaran menyeluruh mengenai kemampuan model dalam mengklasifikasikan data secara akurat

untuk kedua kelas, baik positif maupun negatif [36]. Rumus 2.7 berikut adalah rumus yang digunakan untuk menghitung Accuracy:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Rumus 2.11 Rumus *Accuracy*

Sumber: [36]

Keterangan :

1. TP (*True Positive*): Jumlah prediksi positif yang benar (*instance* positif yang diprediksi sebagai positif).
2. TN (*True Negative*): Jumlah prediksi negatif yang benar (*instance* negatif yang diprediksi sebagai negatif).
3. FP (*False Positive*): Jumlah prediksi positif yang salah (*instance* negatif yang diprediksi sebagai positif).
4. FN (*False Negative*): Jumlah prediksi negatif yang salah (*instance* positif yang diprediksi sebagai negatif).

### 2.3.10 Cross-Validation

Cross-validation merupakan teknik evaluasi model machine learning yang digunakan untuk menilai kemampuan generalisasi model terhadap data baru dengan cara membagi data pelatihan menjadi beberapa subset atau lipatan (*folds*), kemudian melatih model pada sebagian lipatan dan mengujinya pada lipatan yang tersisa secara bergantian. Pendekatan ini membantu memastikan bahwa hasil evaluasi model tidak bergantung pada satu pembagian data tertentu dan memberikan estimasi performa yang lebih stabil. Salah satu bentuk cross-validation yang paling umum adalah *k-fold cross-validation*, di mana dataset dibagi menjadi bagian yang berukuran sama, lalu proses pelatihan dan pengujian dilakukan sebanyak *k* kali, dengan setiap lipatan secara bergantian berperan sebagai data uji dan sisanya sebagai data latih, sehingga nilai performa model dapat dirata-ratakan dari seluruh iterasi untuk memperoleh estimasi akurasi yang lebih andal [37].

Dalam penelitian ini, digunakan *5-fold cross-validation*, di mana data pelatihan dibagi menjadi lima lipatan yang digunakan secara bergantian untuk melatih dan menguji model. Pendekatan ini dipilih karena mampu memberikan keseimbangan antara akurasi evaluasi dan efisiensi komputasi. Melalui metode ini, model seperti Gradient Boosting, XGBoost, dan LightGBM dapat dinilai secara objektif berdasarkan metrik performa seperti F1-score, precision, recall, dan accuracy, sehingga hasil yang diperoleh lebih konsisten dan tidak bias terhadap pembagian data tertentu

### 2.3.11 Purposive Sampling

Purposive sampling merupakan teknik pengambilan sampel non-probabilitas di mana pemilihan sampel dilakukan secara sengaja berdasarkan kriteria tertentu yang relevan dengan tujuan penelitian. Teknik ini digunakan ketika peneliti membutuhkan data dari subjek yang memiliki karakteristik spesifik sehingga dianggap mampu memberikan informasi yang paling sesuai dengan konteks penelitian. Dalam penelitian ini, purposive sampling diterapkan untuk memastikan bahwa data yang digunakan berasal dari nasabah yang memenuhi kriteria tertentu sesuai kebutuhan analisis klasifikasi limit kredit, sehingga proses pemodelan dapat dilakukan secara lebih terfokus dan sesuai dengan ruang lingkup penelitian. [44]

## 2.4 Tools yang digunakan

### 2.4.1 Python

Python adalah bahasa pemrograman tingkat tinggi yang bersifat *general-purpose* dan banyak digunakan dalam pengembangan aplikasi, analisis data, serta *machine learning*. Keunggulan Python terletak pada sintaks yang sederhana dan mudah dipahami, sehingga mempercepat proses pengembangan perangkat lunak [38]. Python juga memiliki ekosistem pustaka yang luas untuk *data science*, seperti Pandas, NumPy, Scikit-learn, XGBoost, dan Matplotlib, yang memudahkan proses pengolahan data hingga pemodelan.

Dalam konteks penelitian ini, Python digunakan sebagai bahasa utama untuk melakukan *data preprocessing*, pelatihan model *machine learning*,



evaluasi model, serta implementasi *deployment*. Ketersediaan pustaka yang lengkap dan dokumentasi yang baik menjadikan Python sebagai pilihan ideal untuk membangun prototipe prediksi dan penyesuaian limit kredit.

#### 2.4.2 Google Colab

Google Colab adalah lingkungan pengembangan berbasis web yang disediakan oleh Google, yang memungkinkan pengguna menulis dan mengeksekusi kode Python secara interaktif melalui browser tanpa perlu instalasi lokal. Platform ini terintegrasi dengan Google Drive, mendukung penggunaan GPU/TPU secara gratis, serta sangat populer di kalangan peneliti dan praktisi data science untuk keperluan analisis data, pembelajaran mesin, dan penelitian akademik [39].

Dalam penelitian ini, Google Colab digunakan pada tahap eksplorasi data (*exploratory data analysis*), *feature engineering*, pelatihan model, dan evaluasi. Keunggulannya yang memungkinkan eksekusi kode secara langsung serta integrasi dengan visualisasi data memudahkan analisis distribusi data, hubungan antar variabel, dan interpretasi hasil model sebelum dilanjutkan ke tahap deployment.

#### 2.4.3 Streamlit

Streamlit adalah *framework* Python yang dirancang untuk membangun aplikasi *web* interaktif secara cepat, khususnya untuk keperluan *data science* dan *machine learning*. Streamlit memungkinkan pengembang membuat antarmuka pengguna dengan sedikit baris kode, tanpa memerlukan keahlian khusus dalam pengembangan *front-end* [40].

Pada penelitian ini, Streamlit digunakan untuk tahap *deployment* model prediksi limit kredit. Prototipe aplikasi yang dibangun dapat memproses data batch, menampilkan hasil prediksi dan rekomendasi penyesuaian limit kredit, serta memberikan simulasi interaktif kepada pengguna. Dengan demikian, *Streamlit* berperan penting dalam mewujudkan model yang dapat direkomendasikan kepada PT XYZ sebagai ilustrasi implementasi model dalam sistem yang lebih luas.