BAB II
Kajian teori

2.1 Penelitian Relevan

Untuk keperluan kajian komparatif terhadap penelitian yang akan dilaksanakan, bagian ini menguraikan sejumlah studi terdahulu yang menitikberatkan pada penerapan algoritma machine learning, khususnya XGBoost dalam proses klasifikasi data pada sektor perbankan maupun bidang lainnya. Selain itu, turut disajikan pula penelitian yang mengintegrasikan metode interpretabilitas model, seperti SHAP, untuk mengidentifikasi variabel-variabel dominan yang memengaruhi proses pengambilan keputusan. Seluruh referensi yang digunakan diperoleh dari jurnal-jurnal bereputasi yang terindeks MDPI, IEEE, ScienceDirect, serta berbagai publikasi akademik yang tersedia melalui Google Scholar, dan selanjutnya dirangkum dalam tabel berikut.

2.1.1 Matriks Penilaian

Tabel 2. 1 Metrik Penilaian

1.	Nama Peneliti	Aldo Andika Saputra, Betha Nurina Sari, Chaerur Rozikin
	Judul Jurnal	Penerapan Algoritma Extreme Gradient Boosting (XGBoost) Untuk Analisis Risiko Kredit
	Jenis Data	Data sekunder (numerik dan kategorikal).
	Sumber Data	Dataset diambil dari UCI Machine Learning Repository
	Metode	Knowledge Discovery in Database (KDD)
	Permasalahan	Meningkatnya kredit macet (non-performing loan) akibat kurang akuratnya proses screening calon nasabah.
	Hasil	Berdasarkan hasil pengujian, algoritma Extreme Gradient Boosting (XGBoost) menunjukkan kinerja yang sangat baik dalam mengklasifikasikan risiko kredit. Penerapan metode SMOTE terbukti meningkatkan performa model secara signifikan. Pada skenario terbaik (90% data latih dan 10% data uji), model XGBoost dengan SMOTE mencapai akurasi 83% dan AUC 0,918, lebih tinggi dibanding model tanpa SMOTE yang hanya memperoleh akurasi 74% dan AUC

		0,822. Hasil ini menunjukkan bahwa penyeimbangan data berperan penting dalam meningkatkan kemampuan model untuk mengidentifikasi kelayakan calon penerima kredit secara lebih akurat.
	Model Evaluasi	Confusion Matrix, AUC-ROC Curve, Akurasi
	Tahun	2024
2.	Nama Peneliti	Kui Wang, Meixuan Li, Jingyi Cheng, Xiaomeng Zhou, dan Gang Li
	Judul Jurnal	Research on Personal Credit Risk Evaluation Based on XGBoost
	Jenis Data	Data sekunder kuantitatif
	Sumber Data	Data riil dari X Bank (salah satu bank komersial di Tiongkok).
	Metode	Metodologi penelitian ini dibagi ke dalam dua tahap utama, yaitu seleksi fitur (feature selection) dan klasifikasi risiko kredit (classification).
	Permasalahan	Penelitian ini dilatarbelakangi oleh permasalahan yang umum dihadapi lembaga keuangan, yaitu kesulitan dalam mengidentifikasi calon debitur yang berpotensi gagal bayar (default) secara akurat. Model penilaian risiko kredit tradisional dianggap masih terbatas karena: a. Tidak mampu menangkap hubungan non-linear antara variabel karakteristik debitur dan status pembayaran, b. Tidak memiliki mekanisme seleksi fitur yang efisien, dan c. Sering kali kurang stabil dalam menghadapi data riil yang bervariasi.

	Hasil	Hasil penelitian menunjukkan bahwa model XGBoost memberikan performa paling unggul dibandingkan dengan metode pembanding. a. XGBoost menghasilkan AUC sebesar 0,943, akurasi 0,871, dan Type II Error (False Negative Rate) sebesar 0,199. b. Sebagai perbandingan, model Decision Tree hanya memperoleh AUC 0,932 dengan akurasi 0,867, sedangkan KNN mencatat AUC 0,912 dan akurasi 0,851. Temuan ini menunjukkan bahwa algoritma XGBoost memiliki kemampuan yang lebih baik dalam membedakan nasabah berisiko tinggi dan rendah, sekaligus lebih tahan terhadap noise dan overfitting. Dengan demikian, metode ini dapat digunakan sebagai alat bantu yang efektif dalam pengambilan keputusan kredit dan penentuan kelayakan penerima pinjaman.
	Model Evaluasi	AUC (Area Under Curve), Akurasi (Accuracy), Type II Error (False Negative Rate)
	Tahun	2022
3.	Nama Peneliti	Hilmi Suftandar, Meditya Wasesa, Utomo Sarjono Putro
	Judul Jurnal	XGBoost Model for Default Prediction in Credit Scoring of Conventional Bank
	Jenis Data	Data sekunder kuantitatif berupa data historis pinjaman konsumtif (consumer loans).
	Sumber Data	Data riil diperoleh dari salah satu bank konvensional di Indonesia
	Metode	Menggunakan pendekatan kuantitatif berbasis machine learning, khususnya algoritma Extreme Gradient Boosting

		(XGBoost). Untuk mengatasi ketidakseimbangan data, digunakan Random OverSampling Examples (ROSE).
	Permasalahan	Sistem credit scoring tradisional sering gagal mendeteksi debitur berisiko tinggi karena: (1) Ketidakseimbangan kelas antara nasabah default dan non-default, (2) Ketergantungan pada kriteria tetap yang tidak adaptif terhadap data besar, dan (3) Kurangnya model yang mampu menangkap pola non-linear perilaku debitur di bank konvensional Indonesia.
	Hasil	Hasil penelitian menunjukkan bahwa algoritma XGBoost mampu memberikan performa prediksi yang sangat baik pada data pelatihan yang telah diseimbangkan menggunakan ROSE, dengan balanced accuracy mencapai 83,9%, recall kelas default sebesar 87,69%, dan AUC 0,9246. Namun, ketika diuji pada data nyata yang tidak seimbang, performa model mengalami penurunan, dengan recall kelas default turun menjadi 47,84%, balanced accuracy 67,74%, dan AUC menurun ke 0,8208. Hal ini menunjukkan bahwa meskipun XGBoost efektif dalam mempelajari pola risiko kredit pada data yang seimbang, model masih rentan terhadap masalah ketidakseimbangan kelas pada kondisi dunia nyata. Dengan demikian, diperlukan pendekatan lanjutan seperti threshold tuning atau cost-sensitive learning agar model lebih stabil dan mampu mengidentifikasi nasabah berisiko tinggi secara lebih akurat.
	Model Evaluasi	Model dievaluasi menggunakan metrik Precision, Recall, F1-Score, ROC-AUC, AUC-PR, dan Kappa Statistic.
	Tahun	2025
4.	Nama Peneliti	Wei Wang, Christopher Lesner, Alexander Ran, Marko Rukonic, Jason Xue, Eric Shiu

Judul Jurnal	Using Small Business Banking Data for Explainable Credit Risk Scoring
Jenis Data	Data sekunder kuantitatif
Sumber Data	Data diperoleh dari Intuit QuickBooks Capital
Metode	Pendekatan machine learning berbasis XGBoost (Extreme Gradient Boosting) dengan monotonic constraints
Permasalahan	Penelitian ini berangkat dari masalah bahwa model credit scoring tradisional (seperti scorecard atau regresi logistik) memiliki keterbatasan dalam menangkap hubungan kompleks antarvariabel dan sulit menjelaskan keputusan berbasis model pembelajaran mesin kepada regulator. Selain itu, banyak data finansial yang tidak terstruktur dan bersuara tinggi, sehingga menantang untuk digunakan dalam model prediksi risiko kredit yang akurat dan transparan.
Hasil	Hasil penelitian menunjukkan bahwa algoritma XGBoost dengan monotonic constraints memberikan performa terbaik dibandingkan model lainnya. Berdasarkan pengujian dengan 75% data latih dan 25% data uji, model XGBoost mencapai K-S (Kolmogorov–Smirnov) statistic sebesar 0,437, yang berarti 7% lebih tinggi dibandingkan model tradisional terbaik (scorecard model dengan K-S = 0,410). Selain itu, penerapan monotonic constraints meningkatkan K-S performance rata-rata sebesar 8% dan AUC sebesar 3%. Model ini juga menunjukkan peningkatan kemampuan diskriminatif dalam membedakan debitur good loan dan bad loan berdasarkan arus kas dan perilaku transaksi.
Model Evaluasi	K-S Statistic (Kolmogorov–Smirnov), ROC–AUC (Receiver Operating Characteristic), Cross-validation (5-fold)
Tahun	2020

5.	Nama Peneliti	Jian Li, Haibin Liu, Zhijun Yang, & Lei Han
	Judul Jurnal	A Credit Risk Model with Small Sample Data Based on G-XGBoost
	Jenis Data	Data kuantitatif sekunder berupa data pinjaman usaha kecil dan menengah (SME)
	Sumber Data	Data riil dari lembaga keuangan di Tiongkok
	Metode	Kombinasi Generative Adversarial Network (GAN) dan Extreme Gradient Boosting (XGBoost) yang disebut G-XGBoost.
	Permasalahan	Jumlah data kredit yang terbatas dan tidak seimbang menyebabkan model prediksi risiko kredit tradisional (seperti Logistic Regression dan XGBoost) tidak akurat dan sulit membedakan risiko dengan baik. Diperlukan pendekatan baru untuk memperbaiki performa model dengan dataset kecil.
	Hasil	Hasil penelitian menunjukkan bahwa model G-XGBoost memberikan peningkatan performa signifikan dibandingkan XGBoost biasa. Nilai KS meningkat dari 0,3643 menjadi 0,3894, dan AUC meningkat dari 0,7453 menjadi 0,7477. Artinya, kemampuan model dalam membedakan kredit baik dan buruk menjadi lebih akurat. Amplifikasi data menggunakan GAN juga membantu mengatasi ketidakseimbangan kelas dan keterbatasan data sampel kecil tanpa menurunkan stabilitas model.
	Model Evaluasi	Kolmogorov–Smirnov (KS) Value, AUC (Area Under the ROC Curve), Cross-validation (4-fold)
	Tahun	2021
6.	Nama Peneliti	Rastra Ardiansyah Pora, Dina Maulina, Ninik Trihartanti
	Judul Jurnal	Komparasi XGBoost dan C4.5 dalam Klasifikasi Risiko Kredit untuk Kelayakan Pinjaman di India

	Jenis Data	Data sekunder kuantitatif
	Sumber Data	Dataset publik “Applicant Details for Loan Approval in India” milik Yamin Hossain, diunduh dari platform Kaggle.
	Metode	Pendekatan Knowledge Discovery in Database (KDD)
	Permasalahan	Risiko kredit masih menjadi tantangan besar di era digitalisasi keuangan. Model tradisional sering gagal menangkap kompleksitas perilaku nasabah dan tidak akurat dalam memprediksi kemungkinan gagal bayar. Oleh karena itu, penelitian ini bertujuan membandingkan dua algoritma populer, XGBoost dan C4.5, untuk menentukan model terbaik dalam mengklasifikasikan risiko kredit dan meningkatkan akurasi penentuan kelayakan pinjaman.
	Hasil	Hasil penelitian menunjukkan bahwa algoritma C4.5 memberikan performa terbaik dalam mengklasifikasikan risiko kredit, baik sebelum maupun sesudah proses tuning. Pada pembagian data 80:20, model C4.5 mencapai akurasi 96,20% dan AUC 0,9730, sedangkan XGBoost memiliki performa sedikit lebih rendah sebelum tuning tetapi meningkat signifikan setelah optimasi parameter. C4.5 menunjukkan hasil yang lebih stabil di semua rasio data (90:10 hingga 60:40) dengan konsistensi tinggi pada metrik Presisi, Recall, dan F1-Score. Sementara XGBoost, meskipun awalnya tertinggal, menjadi kompetitif setelah dilakukan tuning GridSearchCV.
	Model Evaluasi	Accuracy, Precision, Recall, F1-Score, AUC (Area Under Curve), 10-Fold Cross Validation
	Tahun	2025
7.	Nama Peneliti	Zhouyi Gu, Jiayan Lv, Bingya Wu, Zhihui Hu, dan Xinwei Yu

Judul Jurnal	Credit Risk Assessment of Small and Micro Enterprises Based on Machine Learning
Jenis Data	Data kuantitatif sekunder, terdiri atas informasi keuangan dan non-keuangan dari usaha kecil dan mikro.
Sumber Data	Data diperoleh dari perusahaan pihak ketiga penyedia layanan penilaian kredit
Metode	Penelitian menggabungkan algoritma pembelajaran mesin (machine learning) dengan teknik penanganan ketidakseimbangan data (imbalanced processing)
Permasalahan	Usaha kecil dan mikro berperan penting dalam ekonomi, namun mengalami kesulitan pembiayaan akibat penilaian risiko kredit yang tidak akurat dan data yang tidak seimbang (imbalanced) antara debitur lancar dan bermasalah. Selain itu, banyak model konvensional tidak mampu menangani data multidimensi (keuangan dan perilaku) secara efektif. Penelitian ini bertujuan mengembangkan model penilaian risiko kredit berbasis machine learning yang dapat meningkatkan akurasi, mengatasi ketidakseimbangan data, serta memberikan hasil yang mudah diinterpretasi.
Hasil	Model XGBoost dengan SMOTE menunjukkan hasil terbaik dibanding model lain. Berdasarkan pengujian: a. Akurasi = 99,55% b. F1-Score = 99,11% c. AUC = 99,08% Performa ini lebih tinggi dibanding SVM (Akurasi 96,06%), Random Forest (95,57%), dan Decision Tree (94,09%). Model juga berhasil mengidentifikasi dengan benar 99% perusahaan berisiko rendah dan 99% perusahaan berisiko tinggi. Penggunaan SMOTE meningkatkan kemampuan model menangani ketidakseimbangan kelas secara

		signifikan, dan scorecard XGBoost menghasilkan sistem penilaian kredit yang efisien dan mudah diinterpretasikan.
	Model Evaluasi	Accuracy, Precision & Recall, F1-Score, AUC (Area Under Curve), Sensitivity & Specificity, ROC Curve Analysis
	Tahun	2024
8.	Nama Peneliti	Fransiska Aloysia Putri Prasetya dan Paulina H. Prima Rosa
	Judul Jurnal	Klasifikasi Kegagalan Pembayaran Kredit Nasabah Bank dengan Algoritma XGBoost
	Jenis Data	Data sekunder kuantitatif, berupa data historis pinjaman nasabah bank.
	Sumber Data	Dataset publik dari platform Kaggle
	Metode	Penelitian menggunakan pendekatan machine learning dengan algoritma Extreme Gradient Boosting (XGBoost).
	Permasalahan	Data pinjaman bank cenderung tidak seimbang (lebih banyak nasabah lancar dibandingkan yang gagal bayar), sehingga model klasifikasi sering bias terhadap kelas mayoritas. Selain itu, belum ada pendekatan yang secara optimal menggabungkan seleksi fitur, penyeimbangan data, dan penyetelan hiperparameter XGBoost untuk meningkatkan akurasi klasifikasi kegagalan pembayaran kredit.
	Hasil	a. Model awal terbaik diperoleh dengan ambang Information Gain = 0.002 dan K = 10, menghasilkan akurasi 92,94%. b. Setelah dilakukan hyperparameter tuning, model meningkat menjadi akurasi 93,10% pada data latih dan 88,46% pada data uji. c. 14 fitur paling relevan antara lain: Age, HasCoSigner, InterestRate, HasMortgage,

		HasDependents, Income, MaritalStatus, MonthsEmployed, EmploymentType, NumCreditLines, LoanPurpose, Education, LoanAmount, dan LoanTerm. d. Proses balancing dengan SMOTE terbukti efektif dalam meningkatkan kinerja klasifikasi terhadap kelas minoritas.
	Model Evaluasi	Accuracy, Precision, Recall, Specificity, F1-score, K-Fold Cross-validation
	Tahun	2024
9.	Nama Peneliti	Abdul Khahar, Yuni Yamasari
	Judul Jurnal	Eksplorasi Teknik Pre-Processing Berbasis eXtreme Gradient Boosting (XGBoost) pada Klasifikasi Kredit Default
	Jenis Data	Data sekunder kuantitatif, berupa data kredit kartu debitur.
	Sumber Data	Dataset publik Credit Card Default Risk yang diunggah di Kaggle
	Metode	Penelitian menggunakan pendekatan machine learning berbasis XGBoost yang dikombinasikan dengan berbagai teknik pra-pemrosesan data (pre-processing) untuk meningkatkan performa klasifikasi.
	Permasalahan	Penelitian ini dilatarbelakangi oleh tingginya angka gagal bayar kredit konsumtif akibat perilaku konsumtif masyarakat dan meningkatnya aktivitas pinjaman digital. Permasalahan utama yang diidentifikasi adalah ketidakseimbangan data (imbalanced data) yang menyebabkan model cenderung bias terhadap kelas mayoritas (nasabah lancar).

		Selain itu, waktu pemrosesan yang lama pada dataset besar menjadi tantangan yang perlu diatasi dengan teknik feature selection dan sampling yang efisien.
	Hasil	Penelitian ini menunjukkan bahwa kombinasi Random UnderSampler dan feature selection berbasis korelasi (threshold 0,001) memberikan hasil terbaik dalam klasifikasi risiko kredit menggunakan algoritma XGBoost, dengan rasio data 80% pelatihan dan 20% pengujian. Model awal menghasilkan AUC 0,9730, F1-Score 0,7807, Recall 0,9946, Precision 0,6426, dan Akurasi 95,50%. Setelah re-training, performa meningkat menjadi AUC 0,9910, F1-Score 0,9806, Recall 0,9940, Precision 0,9676, dan Akurasi 98,96%. Kombinasi teknik tersebut efektif mengatasi ketidakseimbangan data dan meningkatkan stabilitas model, menjadikan XGBoost andal dalam memprediksi risiko gagal bayar serta mendukung pengambilan keputusan kredit secara akurat.
	Model Evaluasi	Accuracy, Precision, Recall, AUC, F1-Score
	Tahun	2024
10.	Nama Peneliti	Rinno Yunanto, Utomo Budiyanto
	Judul Jurnal	Implementasi XGBoost dan SMOTE untuk Meningkatkan Deteksi Transaksi Fraud di Industri Jasa Keuangan
	Jenis Data	Data sekunder kuantitatif, berupa data transaksi keuangan digital.
	Sumber Data	Dataset publik dari Kaggle
	Metode	Penelitian menggunakan pendekatan Knowledge Discovery in Database (KDD)
	Permasalahan	Salah satu kendala utama dalam deteksi transaksi fraud adalah ketidakseimbangan data, di mana jumlah transaksi

		sah jauh lebih banyak dibandingkan dengan transaksi fraud. Kondisi ini menyebabkan model cenderung bias terhadap kelas mayoritas dan gagal mendeteksi pola transaksi mencurigakan. Selain itu, sebagian besar model tradisional tidak mampu mengatasi data yang kompleks dan besar secara efisien, sehingga dibutuhkan model berbasis machine learning yang lebih adaptif dan sensitif terhadap pola anomali.
	Hasil	Hasil penelitian menunjukkan bahwa kombinasi XGBoost dan SMOTE memberikan peningkatan performa yang signifikan dibandingkan XGBoost tanpa penyeimbangan data. a. Pada rasio 90:10, akurasi meningkat dari 92% menjadi 96% dan AUC dari 0,92 menjadi 0,96. b. Pada rasio 80:20, akurasi meningkat dari 85% menjadi 93%, dan AUC dari 0,85 menjadi 0,93. c. Rata-rata peningkatan performa sebesar +7% akurasi dan +0,08 AUC. d. Model juga menunjukkan balanced accuracy 0,925, precision 0,975, recall 0,851, F1-score 0,909, dan nilai kappa 0,908, menandakan konsistensi deteksi yang sangat baik.
	Model Evaluasi	Accuracy, Precision, Recall, F1-Score, Balanced Accuracy, AUC (Area Under the Curve)
	Tahun	2024
11	Nama Peneliti	Risky Harahap, M. Irpan, M. Azzuhri Dinata, Lusiana Efrizoni, Rahmaddeni
	Judul Jurnal	Perbandingan Algoritma Random Forest dan XGBoost untuk Klasifikasi Penyakit Paru-Paru Berdasarkan Data Demografi Pasien

	Jenis Data	Data sekunder kuantitatif berupa data demografi pasien yang digunakan untuk klasifikasi penyakit paru-paru.
	Sumber Data	Dataset publik yang diperoleh dari Kaggle, berjumlah 30.000 data dengan 9 atribut dan 1 label.
	Metode	Penelitian menggunakan pendekatan data mining dengan tahapan preprocessing, pembagian data training dan testing (80:20), serta perbandingan algoritma Random Forest dan XGBoost.
	Permasalahan	Klasifikasi penyakit paru-paru secara manual sulit dilakukan secara akurat karena banyaknya atribut pasien dan kompleksitas hubungan antar variabel. Metode tradisional kurang mampu menangani data besar dan kompleks, sehingga diperlukan algoritma machine learning yang mampu menghasilkan prediksi lebih tepat.
	Hasil	Hasil penelitian menunjukkan bahwa algoritma XGBoost memiliki performa lebih baik dibandingkan Random Forest: a. Random Forest memperoleh akurasi 91% dan AUC 0,97. b. XGBoost memperoleh akurasi 94% dan AUC 0,98. c. XGBoost lebih konsisten dalam menangani data kompleks serta lebih baik dalam mengatasi overfitting.
	Model Evaluasi	Confusion Matrix, Accuracy, Precision, Recall, F1-Score, dan ROC Curve (AUC).
	Tahun	2024
12.	Nama Peneliti	Yani Prihantini Hiola, Mega Maulina, Indahwati, Aam Alamudi
	Judul Jurnal	Performance Evaluation of Multinomial Logistic Regression, Random Forest, and XGBoost Methods in Data Classification

Jenis Data	Data sekunder kuantitatif berupa berbagai dataset klasifikasi (multikelas dan biner) dengan karakteristik berbeda seperti jumlah fitur, jumlah kelas, jumlah observasi, serta kondisi data seimbang dan tidak seimbang (imbalanced)
Sumber Data	Dataset publik yang berasal dari: a. UC Irvine Machine Learning Repository (UCI Repository) b. Kaggle (Digunakan 10 dataset, misalnya: student dropout, glass identification, cirrhosis survival, zoo, drug, page blocks, stellar, personality, iris, dan e-coli)
Metode	Pendekatan penelitian kuantitatif komparatif (comparative classification study)
Permasalahan	Belum diketahui metode klasifikasi mana yang paling efektif untuk berbagai karakteristik data (jumlah kelas, fitur, ukuran data, serta kondisi seimbang atau tidak seimbang). Banyak penelitian sebelumnya hanya menguji satu dataset sehingga hasilnya tidak general. Oleh karena itu diperlukan evaluasi komprehensif terhadap beberapa algoritma klasifikasi pada beragam dataset.
Hasil	Hasil penelitian menunjukkan: a. Random Forest unggul pada beberapa dataset terutama data tidak seimbang. b. XGBoost menunjukkan performa sangat baik pada dataset kompleks dan data besar. c. Multinomial Logistic Regression cenderung memiliki performa lebih rendah pada data kompleks tetapi masih baik pada dataset sederhana.
Model Evaluasi	Accuracy, Precision, Recall, dan F1-Score

	Tahun	2025

Penelitian ini ditujukan untuk menganalisis penerapan algoritma Extreme Gradient Boosting (XGBoost) dalam proses klasifikasi risiko kredit konsumtif pada Bank MalukuMalut Cabang Masohi, dengan menerapkan metrik evaluasi Accuracy, Precision, Recall, serta F1-Score. Adapun perbandingan antara penelitian ini dengan penelitian-penelitian terdahulu dapat dijelaskan sebagai berikut:

a. Dataset yang digunakan

   Pada penelitian terdahulu, dataset yang digunakan umumnya berasal dari sumber terbuka seperti Kaggle, UCI Machine Learning Repository, atau dataset publik lainnya yang berisi data pinjaman, transaksi keuangan, maupun data nasabah kartu kredit. Beberapa penelitian juga memanfaatkan data non-keuangan untuk menguji performa algoritma XGBoost dalam klasifikasi. Sementara itu, penelitian ini menggunakan dataset internal dari Bank MalukuMalut Cabang Masohi yang berisi data kredit konsumtif. Penggunaan data riil ini memberikan kontribusi baru karena mencerminkan karakteristik lokal nasabah pada lembaga keuangan daerah, yang selama ini belum banyak dijadikan objek penelitian.

b. Metode yang digunakan

   Penelitian terdahulu banyak menerapkan algoritma XGBoost untuk klasifikasi data tabular dan pengelolaan data tidak seimbang dengan bantuan teknik seperti SMOTE (Synthetic Minority Over-sampling Technique), Random UnderSampler, maupun GridSearchCV untuk optimasi parameter model.

   Penelitian ini mengadopsi pendekatan serupa, namun difokuskan pada penerapan XGBoost terhadap data kredit konsumtif aktual dengan kompleksitas tinggi dan ketidakseimbangan kelas yang signifikan. Pendekatan ini diharapkan mampu menghasilkan model yang lebih representatif dan adaptif terhadap data perbankan daerah. Penelitian-penelitian sebelumnya secara umum menggunakan metrik evaluasi standar seperti accuracy, precision, recall, F1-score, dan AUC, bahkan beberapa menambahkan specificity, sensitivity, atau confusion matrix untuk memperkuat pengukuran performa model.

Penelitian ini akan menggunakan keempat metrik utama tersebut, yaitu:

1. Accuracy, untuk menilai tingkat ketepatan keseluruhan model;
2. Precision, untuk mengukur ketepatan model dalam mengklasifikasikan nasabah layak kredit;
3. Recall, untuk menilai kemampuan model dalam mendeteksi nasabah berisiko tinggi;
4. F1-Score, sebagai ukuran keseimbangan antara precision dan recall.

Dengan menggunakan kombinasi metrik tersebut, hasil penelitian ini diharapkan dapat dibandingkan secara objektif dengan penelitian terdahulu serta memberikan gambaran yang komprehensif mengenai efektivitas model XGBoost dalam memprediksi risiko kredit konsumtif di lingkungan perbankan daerah.

2.2 Landasan Teori

Landasan teori pada penelitian ini memuat uraian mengenai teori-teori yang memiliki keterkaitan dengan topik yang diteliti. Adapun penjelasan landasan teori dalam penulisan ini disajikan sebagai berikut:

2.2.1 Kredit Konsumtif

Kredit konsumtif merupakan fasilitas pembiayaan yang diberikan oleh bank maupun lembaga keuangan non-bank kepada nasabah dengan tujuan pemanfaatan pribadi atau konsumsi. Umumnya, kredit konsumtif digunakan untuk memperoleh barang-barang dengan nilai ekonomi yang relatif tinggi namun tidak termasuk kebutuhan pokok, seperti kendaraan bermotor, peralatan elektronik, gawai, maupun perabot rumah tangga. Kredit konsumtif dipandang sebagai alternatif pembiayaan yang praktis karena memungkinkan masyarakat memperoleh barang yang diinginkan secara langsung tanpa harus menunggu terkumpulnya dana pribadi. Melalui fasilitas ini, kebutuhan konsumsi dapat segera terpenuhi dengan mengajukan pembiayaan baik pada bank konvensional maupun bank syariah.(Adi Caksana & Wulandari, 2021)

2.2.2 Risiko Kredit

Risiko kredit diartikan sebagai potensi kerugian yang berkaitan dengan kemungkinan debitur gagal memenuhi kewajibannya pada saat jatuh tempo. Dengan kata lain, risiko kredit muncul ketika peminjam tidak mampu melunasi kewajiban utangnya. Risiko kredit merupakan risiko yang timbul akibat kegagalan debitur maupun pihak lawan transaksi (*counterparty*) dalam memenuhi kewajiban yang telah disepakati. Kredit pada dasarnya menjadi sumber utama pendapatan serta keuntungan bagi perbankan. Namun demikian, penyaluran kredit juga memiliki potensi risiko yang tinggi karena apabila tidak dikelola dengan baik dapat menimbulkan kredit bermasalah atau Non Performing Loan (NPL).(Sante et al., 2021)

Risiko kredit merupakan potensi kerugian yang timbul ketika pihak lawan (counterparty) gagal memenuhi kewajiban pembayaran sesuai perjanjian. Risiko ini tidak hanya berasal dari aktivitas pemberian kredit kepada nasabah, tetapi juga dapat muncul dari berbagai kegiatan keuangan lainnya, seperti akseptasi, transaksi antarbank, pembiayaan perdagangan, kewajiban komitmen dan kontinjensi, penerbitan obligasi, serta transaksi nilai tukar dan derivatif.(Rohimah et al., 2023)

Dalam manajemen risiko perbankan, setiap aktivitas tersebut memiliki tingkat potensi gagal bayar yang berbeda, tergantung pada profil debitur, kondisi pasar, dan efektivitas pengendalian risiko bank. Oleh karena itu, diperlukan penilaian risiko kredit yang menyeluruh melalui analisis kelayakan debitur, pemantauan eksposur, serta penerapan strategi mitigasi seperti penggunaan agunan, diversifikasi portofolio, dan pengawasan berkelanjutan. Pendekatan yang sistematis dan terukur ini penting untuk menjaga stabilitas keuangan bank serta mendukung keberlanjutan fungsi intermediasi secara sehat.(Rohimah et al., 2023)

2.2.3 Non-Performing Loan (NPL)

NPL merupakan rasio yang digunakan untuk mengukur kemampuan bank dalam menanggung risiko kegagalan pembayaran kembali kredit oleh debitur. Berdasarkan Peraturan Bank Indonesia Nomor 6/10/PBI/2004 tanggal 12 April 2004 mengenai Sistem Penilaian Tingkat Kesehatan Bank Umum, bank

dikategorikan tidak sehat apabila tingkat NPL melebihi 5%. Tingginya rasio NPL akan berdampak pada penurunan laba yang dapat diperoleh bank. Sejalan dengan hal tersebut, Almilia dan Herdiningtyas (2022) menegaskan bahwa semakin buruk kualitas kredit suatu bank, yang ditandai dengan meningkatnya jumlah kredit bermasalah, maka semakin besar pula kemungkinan bank tersebut berada dalam kondisi bermasalah.(Sante et al., 2021)

Salah satu indikator yang digunakan untuk menilai tingkat risiko kredit adalah Non-Performing Loan (NPL). Rasio NPL mencerminkan kemampuan manajemen bank dalam mengelola kredit bermasalah atau kredit macet yang telah disalurkan. Semakin tinggi nilai NPL, semakin rendah kualitas portofolio kredit bank tersebut, yang menunjukkan meningkatnya jumlah kredit bermasalah. Peningkatan kredit bermasalah berdampak negatif terhadap kinerja keuangan bank, karena dapat menurunkan pendapatan dan laba, sementara kewajiban pembayaran bunga kepada nasabah tetap harus dipenuhi. Dengan demikian, risiko kredit sangat dipengaruhi oleh kualitas aset bank, yang ditentukan oleh tingkat kelancaran kredit, kondisi kesehatan bank, serta profitabilitas dari aktivitas penyaluran pinjaman.(Silitonga & Manda, 2022)

2.2.4 Sistem Kredit Scoring

Credit scoring adalah suatu proses yang melibatkan pengumpulan, analisis, dan pengklasifikasian berbagai variabel terkait dengan kredit untuk menilai keputusan kredit. Penilaian ini didasarkan pada perbandingan antara informasi debitur saat ini dengan debitur sebelumnya (Anderson, 2020). Credit scoring dapat dianggap sebagai suatu proses statistik yang mengubah data calon debitur menjadi informasi yang dapat digunakan untuk membuat keputusan kredit. Pemberian kredit harus dilakukan dengan bijaksana, mengingat tingginya tingkat risiko yang terkait. Proses credit scoring tidak hanya bertujuan untuk menilai kelayakan debitur dalam membayar kembali kredit, tetapi juga untuk menganalisis data kredit bermasalah atau Non-Performing Loan (NPL).(Dzulhijjah et al., 2024)

2.2.5 Kolektibilitas Kredit

Kolektibilitas merupakan cerminan dari kemampuan dan ketepatan debitur dalam memenuhi kewajiban pembayaran pokok serta bunga kredit sesuai dengan perjanjian yang telah disepakati. Selain itu, kolektibilitas juga mencerminkan tingkat kemungkinan bank untuk memperoleh kembali dana yang telah disalurkan, baik melalui pemberian kredit, investasi dalam surat berharga, maupun bentuk penanaman dana lainnya.(Wahyu, 2020)

Menurut ketentuan Bank Indonesia, tingkat kolektibilitas pinjaman diklasifikasikan ke dalam lima kategori, yaitu: (1) kredit lancar, (2) kredit dalam perhatian khusus (special mention), (3) kredit kurang lancar, (4) kredit diragukan, dan (5) kredit macet. Klasifikasi ini digunakan sebagai alat ukur untuk menilai kualitas aset produktif perbankan serta menjadi dasar dalam penetapan cadangan kerugian penurunan nilai (CKPN).(Wahyu, 2020)

Kolektibilitas kredit diklasifikasikan berdasarkan ketepatan pembayaran pokok dan bunga oleh debitur, yang berfungsi untuk menilai kualitas kredit dan tingkat risiko pembiayaan bank. Kredit lancar adalah kredit dengan pembayaran pokok dan bunga yang dilakukan tepat waktu sesuai perjanjian, mencerminkan kemampuan dan komitmen debitur yang baik. Kredit kurang lancar menunjukkan adanya keterlambatan pembayaran sekitar tiga bulan dari jadwal, menandakan penurunan kemampuan debitur meskipun peluang pelunasan masih ada. Sementara itu, kredit diragukan menggambarkan keterlambatan lebih dari enam bulan atau dua kali jadwal pembayaran, yang menandakan risiko gagal bayar tinggi dan membutuhkan tindakan penagihan serta evaluasi ulang agunan. Klasifikasi ini membantu bank dalam memantau kualitas kredit, menetapkan cadangan kerugian penurunan nilai (CKPN), dan menjaga stabilitas keuangan melalui penerapan prinsip kehati-hatian dalam pengelolaan risiko kredit.(Wahyu, 2020)

2.2.6 Klasifikasi

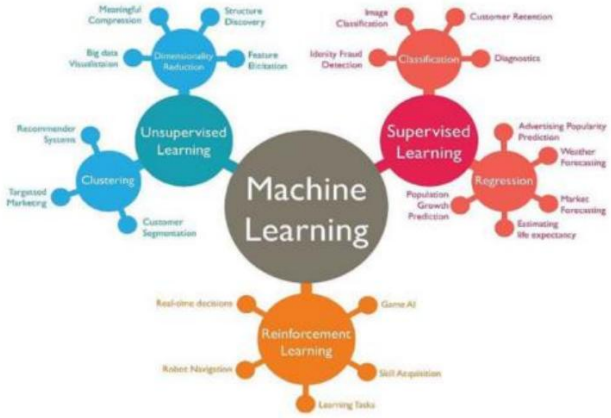
Penerapan model klasifikasi berbasis teknik Machine Learning (ML) di berbagai bidang telah menjadi tren penelitian yang signifikan dalam beberapa

tahun terakhir. Hal ini disebabkan oleh kemampuannya yang terbukti menghasilkan tingkat akurasi yang sangat tinggi, bahkan mencapai lebih dari 96%. Proses klasifikasi menggunakan teknik ML dilakukan dengan membagi data historis ke dalam data pelatihan (training data) dan data pengujian (testing data), yang selanjutnya diproses untuk membentuk suatu model. Model tersebut kemudian dimanfaatkan sebagai dasar dalam memberikan rekomendasi atau prediksi secara otomatis untuk berbagai kebutuhan di masa mendatang.(Puji Catur Siswipraptini 1*, Ahmad Saputra Fadiarora 2, 2023)

Klasifikasi merupakan salah satu teknik dalam data mining yang bertujuan untuk membangun suatu model atau fungsi yang mampu menjelaskan serta membedakan antar kelas atau konsep. Model yang dihasilkan selanjutnya dimanfaatkan untuk menentukan atau memprediksi kelas dari suatu objek yang belum diketahui label kelasnya. Proses pembentukan model tersebut didasarkan pada hasil analisis terhadap data pelatihan (training data) yang telah tersedia.(Pritasari Palupiningsih, 2018)

2.2.7 Konsep Machine Learning

Salah satu bidang Kecerdasan Buatan yang semakin popular saat ini adalah Machine Learning.Machine Learning (ML) atau pembelajaran mesin merupakan salah satu cabang dari kecerdasan buatan yang berfokus pada kemampuan sistem untuk memperoleh pengetahuan dari data (learn from data). Bidang ini menitikberatkan pada pengembangan algoritma yang memungkinkan sistem belajar secara otomatis tanpa perlu instruksi pemrograman berulang dari manusia. Agar proses pembelajaran (training) dapat menghasilkan performa optimal pada tahap testing, diperlukan ketersediaan data yang valid dan berkualitas sebagai dasar pembelajaran.(Cholissodin & Soebroto, 2021)



Gambar 2. 1 Map Cakupan dari Machine Learning.*(Cholissodin & Soebroto, 2021)*

Jika Machine Learning dianalogikan sebagai sebuah kendaraan bermotor, maka data merupakan bahan bakar utamanya. Hal ini disebabkan karena algoritma pembelajaran mesin hanya dapat berfungsi apabila tersedia data yang dapat digunakan untuk membangun metode penyelesaian masalah. Dengan demikian, penerapan berbagai teknik dalam machine learning menuntut adanya data; tanpa data, algoritma tidak dapat dijalankan. Secara umum, data dalam pembelajaran mesin dibagi menjadi dua kategori, yaitu data training dan data testing. Data training digunakan untuk melatih algoritma agar mampu mengenali pola, sedangkan data testing dipakai untuk mengevaluasi kinerja algoritma yang telah dilatih ketika berhadapan dengan data baru yang belum pernah ditemui sebelumnya (Fikriya, Irawan, & Soetrisno, 2022).Lebih lanjut, machine learning terdiri atas beragam algoritma dengan fungsi dan tujuan yang berbeda. Beberapa jenis algoritma yang paling banyak digunakan mencakup Supervised Learning, Unsupervised Learning, Semi-Supervised Learning, dan Reinforcement Learning, masing-masing dengan karakteristik serta penerapan yang khas.(Alfarizi et al., 2023)

2.2.8 Extreame Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) merupakan algoritma *ensemble learning* yang menggunakan *decision tree* sebagai model dasar dan dikenal memiliki kapabilitas yang kuat dalam memodelkan data numerik yang bersifat kompleks, terutama data yang mengandung *noise* dan *outlier*. Algoritma ini

menerapkan pendekatan *boosting* dengan membangun model secara iteratif, di mana setiap *decision tree* yang dihasilkan pada tahap berikutnya difokuskan untuk mengoreksi kesalahan prediksi yang dihasilkan oleh model sebelumnya.(Ainul idham, 2025)

XGBoost bekerja dengan menggabungkan sejumlah *decision tree* sebagai *weak learner* yang dibangun secara iteratif dan bertahap, di mana setiap model baru berperan dalam memperbaiki kesalahan prediksi yang dihasilkan oleh model sebelumnya, sehingga menghasilkan model akhir dengan tingkat akurasi dan generalisasi yang lebih tinggi (Ibrahem Ahmed Osman et al., 2021).

Keunggulan utama XGBoost terletak pada kinerjanya yang tinggi, efisiensi dalam penggunaan waktu, serta pemanfaatan memori yang lebih optimal. Berkat karakteristik tersebut, XGBoost telah diaplikasikan secara luas di berbagai bidang penelitian, termasuk kesehatan, penilaian risiko kredit, hingga analisis metagenomik. (Ilmiah & Pendidikan, 2024)

Algoritma XGBoost dapat diterapkan baik untuk permasalahan klasifikasi maupun regresi. Mekanisme kerjanya dilakukan dengan menggabungkan hasil prediksi dari sejumlah pohon lemah menjadi sebuah model pohon yang lebih kuat melalui proses iteratif. Pada setiap iterasi, XGBoost memanfaatkan nilai residual untuk memperbaiki atau menyesuaikan pohon yang telah dibangun sebelumnya. Proses ini dikenal dengan istilah optimisasi fungsi loss (loss function optimization).(Yaqin, 2022)

Secara matematis, model prediksi XGBoost dirumuskan sebagai berikut:

$$\hat{y}_i = \sum_{k=1}^{\mathrm{K}} f_k(x_i), \;\; f_k \in \mathcal{F} \qquad \textbf{(2. 1)}$$

Keterangan:

$\hat{y}_i$ adalah nilai prediksi data ke-i

f_k adalah fungsi pohon keputusan ke-k

$\mathcal{F}$ merupakan ruang seluruh kemungkinan pohon keputusan

K menyatakan jumlah pohon dalam model

Persamaan (2.1) menunjukkan bahwa nilai prediksi akhir diperoleh dari penjumlahan kontribusi seluruh pohon keputusan yang dibangun secara bertahap.

Tujuan utama algoritma XGBoost adalah meminimalkan fungsi objektif (*objective function*) yang terdiri dari fungsi loss dan fungsi regularisasi. Fungsi objektif tersebut dirumuskan sebagai berikut:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(yi, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \quad \textbf{(2. 2)}$$

Keterangan:

$l(y_i, \hat{y}_i)$ adalah fungsi loss yang mengukur kesalahan prediksi

$\hat{y}_i^{(t-1)}$ adalah prediksi pada iterasi sebelumnya

$f_t(x_i)$ adalah pohon keputusan pada iterasi ke-t

$\Omega(f_t)$ adalah fungsi regularisasi

Fungsi regularisasi digunakan untuk mengendalikan kompleksitas model agar tidak mengalami *overfitting*. Fungsi regularisasi pada XGBoost dirumuskan sebagai berikut:

$$\Omega(f_t) = \gamma \mathrm{T} + \frac{1}{2}\lambda \sum_{j=1}^{\mathrm{T}} w_j^2 \quad \textbf{(2. 3)}$$

Keterangan:

T adalah jumlah daun pada pohon keputusan

w_j adalah bobot pada daun ke-j

γ merupakan parameter penalti kompleksitas struktur pohon

λ merupakan parameter regularisasi L2

Persamaan (2.3) menunjukkan bahwa semakin kompleks struktur pohon dan semakin besar bobot daun, maka nilai penalti akan semakin besar, sehingga model

terdorong untuk membentuk struktur yang lebih sederhana dan stabil.(Ibrahem Ahmed Osman et al., 2021)

Proses optimasi dalam XGBoost dilakukan dengan menggunakan pendekatan ekspansi Taylor orde kedua terhadap fungsi loss. Dengan pendekatan ini, fungsi loss didekati menggunakan gradien pertama dan gradien kedua sebagai berikut:

$$g_i = \frac{\partial l(y_i, \hat{y}_i)}{\partial \hat{y}} \quad \textbf{(2. 4)}$$

$$h_i = \frac{\partial^2 l(y_i, \hat{y}_i)}{\partial \hat{y}_i^2} \quad \textbf{(2. 5)}$$

Penggunaan gradien orde kedua memungkinkan XGBoost untuk memperhitungkan tingkat perubahan kesalahan secara lebih akurat, sehingga proses optimasi menjadi lebih cepat dan stabil dibandingkan metode Gradient Boosting konvensional.(Ramaneswaran et al., 2021)

Bobot optimal pada setiap daun pohon keputusan dihitung menggunakan persamaan berikut:

$$w_j^* = \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad \textbf{(2. 6)}$$

Keterangan:

I_j adalah himpunan data yang berada pada daun ke-j

w_j^* adalah bobot optimal daun

Nilai fungsi objektif minimum yang diperoleh setelah optimasi pohon keputusan dapat dirumuskan sebagai berikut:

$$\mathcal{L}^* = -\frac{1}{2}\sum_{j=1}^{\mathrm{T}} \frac{\left(\sum_{i \in I_j} g_i\right)^2}{\sum_{i \in Ij} h_i + \lambda} + \gamma \mathrm{T} \quad \textbf{(2. 7)}$$

Persamaan (2.7) digunakan sebagai dasar dalam menentukan struktur pohon keputusan terbaik dengan cara memilih pemisahan (split) yang memberikan nilai penurunan fungsi objektif terbesar.(Abdurrahman & Sintawati, 2020)

Pada permasalahan klasifikasi multikelas, XGBoost menghasilkan skor mentah (*logit*) untuk setiap kelas. Untuk mengubah skor tersebut menjadi probabilitas kelas, digunakan fungsi Softmax. Fungsi Softmax dirumuskan sebagai berikut:

$$\mathrm{P}(y = k|x_i) = \frac{\exp\left(\hat{y}_{i,k}\right)}{\sum_c^C \exp\left(\hat{y}_{i,k}\right)} \quad \textbf{(2. 8)}$$

Keterangan:

$P(y = k \mid x_i)$ adalah probabilitas data ke-i termasuk ke kelas ke-k

$\hat{y}_{i,k}$ adalah skor prediksi XGBoost untuk kelas ke-k

C adalah jumlah kelas

Fungsi Softmax memetakan skor prediksi ke dalam bentuk probabilitas dengan total nilai sama dengan satu. Kelas dengan skor terbesar akan memiliki probabilitas tertinggi.(Ramaneswaran et al., 2021)

Dalam klasifikasi multikelas, fungsi Softmax digunakan bersama fungsi *loss* multiclass log loss yang dirumuskan sebagai berikut:

$$l(y_i, \hat{y}_i) = -\sum_{k=1}^{C} y_{i,k} \log\left(\mathrm{P}\left(y = k \mid x_i\right)\right) \quad \textbf{(2. 9)}$$

Keterangan:

$y_{i,k}$ adalah label aktual dalam bentuk *one-hot encoding*

$P(y = k \mid x_i)$ adalah probabilitas hasil fungsi Softmax

Gradien pertama dan kedua dari fungsi *loss* Softmax inilah yang digunakan dalam proses pembentukan pohon keputusan dan perhitungan bobot daun, sehingga fungsi Softmax terintegrasi langsung ke dalam mekanisme optimasi XGBoost.(Ramaneswaran et al., 2021)

Berdasarkan mekanisme tersebut, algoritma XGBoost membangun model secara iteratif hingga mencapai jumlah pohon maksimum atau hingga nilai fungsi objektif tidak lagi mengalami penurunan yang signifikan. Pendekatan ini menjadikan XGBoost sebagai algoritma yang memiliki performa tinggi, stabil

terhadap overfitting, serta mampu menangani data dengan kompleksitas yang tinggi.

2.2.9 Pembagian Data Latih dan Data Uji

Pembagian antara data latih dan data uji merupakan metode yang lazim digunakan untuk mengevaluasi kinerja suatu algoritma. Baik data latih maupun data uji terdiri atas pasangan input dan output. Data latih berfungsi untuk membangun serta melatih model, sedangkan data uji digunakan sebagai data baru dengan nilai output yang disembunyikan dari algoritma guna menilai kemampuan generalisasi model. Dalam praktiknya, jika jumlah data relatif kecil, proporsi pembagian yang umum digunakan adalah 80% untuk data latih dan 20% untuk data uji. Sementara pada dataset berukuran besar, proporsi yang sering diterapkan adalah 70% data latih dan 30% data uji.(Fikriaziz et al., 2024)

2.2.10 Normalisasi Data (Min-Max Normalization)

Normalisasi data merupakan suatu proses transformasi nilai dari data mentah ke dalam rentang tertentu agar lebih terstandarisasi. Tujuan dari normalisasi adalah untuk mengurangi perbedaan skala antar variabel sehingga setiap atribut memiliki kontribusi yang seimbang dalam proses analisis maupun pemodelan. Dengan demikian, normalisasi berperan penting dalam meningkatkan stabilitas, akurasi, dan efisiensi algoritma machine learning maupun metode analisis statistik. Salah satu teknik yang paling sederhana dan banyak digunakan adalah Min-Max Normalization. Metode ini melakukan transformasi linear terhadap nilai asli data sehingga menghasilkan rentang baru, biasanya antara 0 hingga 1 atau -1 hingga 1, tergantung kebutuhan. Transformasi tersebut menjaga proporsi relatif antar data, sehingga nilai setelah dinormalisasi tetap mencerminkan hubungan skala dari data sebelum diproses. (Selle et al., 2022)

Dalam konteks implementasi, Min-Max Normalization sangat bermanfaat ketika data memiliki perbedaan skala yang signifikan. Misalnya, atribut dengan nilai ribuan dapat mendominasi atribut lain dengan skala puluhan jika tidak dilakukan normalisasi. Oleh karena itu, penerapan metode ini membantu

menghindari bias model terhadap variabel tertentu, serta mempercepat konvergensi dalam algoritma pembelajaran berbasis gradien.(Selle et al., 2022)

Persamaan matematis dari Min-Max Normalization dapat dituliskan sebagai berikut:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad \textbf{(2. 10)}$$

Keterangan:

$x =$ hasil transformasi suatu nilai

$min =$ nilai terkecil pada data set fitur X

$max =$ nilai terbesar pada data set fitur X

2.2.11 Random Oversampling

Dalam proses klasifikasi, salah satu permasalahan yang sering muncul adalah ketidakseimbangan kelas (class imbalance problem). Kondisi ini terjadi ketika jumlah data pada satu kelas jauh lebih banyak dibandingkan kelas lainnya Ketidakseimbangan ini menyebabkan model cenderung mempelajari pola dari kelas mayoritas dan mengabaikan kelas minoritas sehingga meningkatkan kemungkinan kesalahan klasifikasi pada kelas minoritas.(Yang et al., 2024)

Apabila data tidak seimbang, algoritma machine learning dapat menjadi bias terhadap kelas mayoritas. Model akan lebih sering memprediksi kelas mayoritas karena secara statistik memberikan kesalahan total yang lebih kecil, namun kemampuan mendeteksi kelas minoritas menjadi rendah.(Yang et al., 2024)

Salah satu teknik yang umum digunakan untuk mengatasi masalah tersebut adalah **Random Oversampling (ROS)**. Random Oversampling merupakan metode penyeimbangan data dengan cara menambahkan kembali data pada kelas minoritas melalui proses penggandaan sampel secara acak (random replication) hingga proporsi jumlah data antar kelas menjadi lebih seimbang.

Pada penelitian Cynthia Yang (2024), yang merancang proses pengembangan model melalui pembagian data menjadi data pelatihan dan data pengujian, kemudian melakukan validasi silang (cross-validation) pada data pelatihan. Dalam proses tersebut, teknik penyeimbangan kelas seperti random oversampling hanya diterapkan pada *training folds*, sedangkan *validation fold* dan *test set* tetap menggunakan distribusi data asli. Pendekatan ini bertujuan agar evaluasi performa model dilakukan pada data yang tidak mengalami manipulasi distribusi kelas sehingga hasil pengujian mencerminkan kemampuan generalisasi model secara objektif. (Yang et al., 2024)

2.2.12 Metrik Evaluasi Model

Confusion Matrix merupakan suatu representasi evaluasi model klasifikasi yang disajikan dalam bentuk tabel dan secara luas digunakan untuk menggambarkan kinerja (performance) suatu model klasifikasi. Tabel ini tersusun atas baris dan kolom yang jumlahnya disesuaikan dengan banyaknya kelas yang terlibat. Setiap sel dalam *Confusion Matrix* merepresentasikan hasil prediksi model, yang meliputi false positives (FP), yaitu data dengan kelas negatif yang diprediksi sebagai positif; false negatives (FN), yaitu data dengan kelas positif yang diprediksi sebagai negatif; true positives (TP), yaitu data dengan kelas positif yang berhasil diprediksi secara benar; serta true negatives (TN), yaitu data dengan kelas negatif yang teridentifikasi secara tepat. Selanjutnya, nilai-nilai tersebut digunakan sebagai dasar dalam perhitungan metrik evaluasi, salah satunya akurasi. (Nur et al., 2022)

Beberapa metrik yang paling umum digunakan antara lain Accuracy, Precision, Recall, dan F1-score, yang masing-masing memiliki kelebihan, keterbatasan, serta konteks penggunaan tertentu.

2.2.12.1 Accuracy

Accuracy adalah metrik yang paling sederhana dan intuitif untuk menilai kinerja model. Ia mengukur persentase prediksi yang benar dari keseluruhan data, baik untuk kelas positif maupun negatif. Rumusnya adalah:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad \textbf{(2. 11)}$$

Dalam kasus multi-class, jurnal Opitz (2024) menjelaskan bahwa Accuracy setara dengan micro precision, micro recall, dan micro F1, karena semua prediksi dihitung secara agregat tanpa membedakan tiap kelas. Hal ini membuat Accuracy populer karena mudah dipahami dan dihitung.

Namun, kelemahan utama *Accuracy* adalah ketergantungannya pada distribusi kelas (*prevalence*). Jika dataset tidak seimbang, metrik ini bisa menyesatkan. Misalnya, model yang selalu memprediksi kelas mayoritas tetap dapat memperoleh nilai Accuracy tinggi meskipun gagal mendeteksi kelas minoritas. Jurnal menekankan bahwa dalam situasi seperti ini, *Accuracy* sebaiknya dipadukan dengan metrik lain yang lebih sensitif terhadap ketidakseimbangan data agar evaluasi lebih adil dan bermakna.(Opitz, 2024)

2.2.12.2 Precision

Precision menilai kualitas prediksi positif dengan menghitung proporsi prediksi positif yang benar. Rumusnya:

$$Precision = \frac{TP}{TP + FP} \quad \textbf{(2. 12)}$$

Metrik ini penting bila konsekuensi dari false positive cukup besar, misalnya dalam sistem deteksi spam, di mana terlalu banyak false alarm akan mengganggu pengguna. Dalam jurnal Opitz (2024), Precision juga dibahas dalam bentuk Macro Precision, yakni rata-rata precision tiap kelas tanpa memperhitungkan ukuran kelas. Pendekatan ini memberi bobot yang sama bagi kelas mayoritas dan minoritas, sehingga lebih adil dalam kasus data tidak seimbang.

Meski begitu, Precision dipengaruhi oleh bias prediksi model dan distribusi kelas. Jika distribusi kelas berubah dari data latih ke data nyata,

nilai Precision juga bisa bergeser. Oleh karena itu, jurnal menyarankan adanya prevalence calibration, yaitu penyeimbangan distribusi kelas secara artifisial agar nilai Precision menjadi lebih stabil dan dapat dibandingkan antar model. Dengan demikian, Precision tetap berguna, tetapi harus dipahami dalam konteks distribusi kelas dan sering dipadukan dengan Recall untuk memperoleh gambaran menyeluruh.(Opitz, 2024)

2.2.12.3 Recall

Recall mengukur seberapa baik model mendeteksi semua kasus positif. Rumusnya:

$$Recall = \frac{TP}{TP + FN} \quad \textbf{(2. 13)}$$

Recall tinggi berarti model jarang melewatkan kasus positif, yang sangat penting di bidang seperti kesehatan atau keamanan, di mana kesalahan mendeteksi kasus positif (false negative) bisa berakibat fatal. Jurnal Opitz (2024) menyebut Macro Recall sebagai metrik yang sangat kuat karena memenuhi lima sifat ideal evaluasi: monotonicity, class sensitivity, decomposability, prevalence invariance, dan chance correction.

Artinya, Macro Recall benar-benar adil terhadap semua kelas karena tidak dipengaruhi distribusi data, serta mampu memberi baseline yang jelas bahkan jika dibandingkan dengan random classifier. Jurnal menekankan bahwa Macro Recall dapat dianggap sebagai Accuracy yang telah dikalibrasi agar semua kelas memiliki bobot yang sama. Hal ini menjadikan Recall sangat penting untuk dataset tidak seimbang, misalnya dalam deteksi penipuan atau penyakit langka, di mana yang terpenting adalah memastikan semua kasus positif terdeteksi meski ada risiko false positives meningkat.(Opitz, 2024)

2.2.12.4 F1-score

F1-score adalah harmonic mean dari Precision dan Recall, dengan formula:

$$F1 = \frac{2\,.\,Precision\,.Recall}{Precision + Recall} \quad \textbf{(2. 14)}$$

Metrik ini dirancang untuk menyeimbangkan trade-off antara Precision dan Recall, sehingga F1 hanya akan tinggi jika keduanya tinggi. Hal ini membuat F1 sangat populer, terutama dalam kasus di mana distribusi data tidak seimbang. Dalam jurnal Opitz (2024), Macro F1 sering dijadikan metrik utama dalam evaluasi model NLP karena menggabungkan kelebihan metrik macro (adil antar kelas) dan micro (mengukur performa keseluruhan).

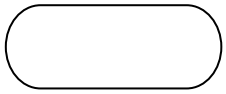
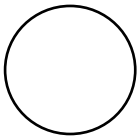
Namun, jurnal juga mengungkap bahwa Macro F1 tidak sepenuhnya memperlakukan semua kelas secara adil karena masih dipengaruhi distribusi kelas, kecuali dilakukan prevalence calibration. Selain itu, ada kebingungan yang muncul karena adanya dua definisi berbeda, yakni Macro F1 (rata-rata F1 tiap kelas) dan Macro F1' (harmonic mean dari Macro Precision dan Macro Recall). Perbedaan ini sering tidak dijelaskan dengan jelas dalam publikasi, sehingga bisa menimbulkan salah tafsir dalam perbandingan hasil penelitian.(Opitz, 2024)

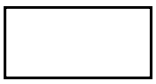
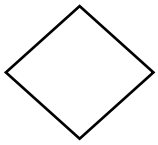
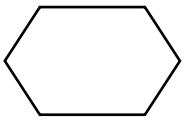
2.2.13 Flowcart Penelitian

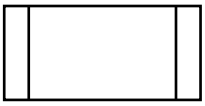
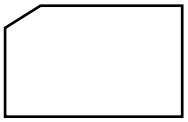
Flowchart dapat dipahami sebagai sebuah bagan yang merepresentasikan alur logis dari suatu program atau prosedur dalam sistem. Melalui simbol-simbol grafis yang telah distandarisasi, flowchart menyajikan urutan langkah, keputusan, maupun proses yang terjadi dalam sebuah algoritma atau prosedur kerja.Fungsi utama flowchart adalah sebagai alat bantu komunikasi yang memudahkan penyampaian ide, logika program, maupun prosedur sistem kepada berbagai pihak, baik pengembang, peneliti, maupun pengguna non-teknis. Dengan visualisasi yang jelas, flowchart mampu mengurangi kesalahpahaman,

mempercepat proses diskusi, serta mendukung dokumentasi yang sistematis.Selain itu, flowchart juga memiliki nilai praktis dalam perencanaan, pengembangan, dan evaluasi sistem, karena membantu mengidentifikasi alur kerja yang tidak efisien, mendeteksi kemungkinan kesalahan logika, serta mempermudah proses debugging dan perbaikan. Dengan demikian, flowchart tidak hanya berfungsi sebagai media komunikasi, tetapi juga sebagai instrumen analisis dan pengendalian kualitas dalam rekayasa perangkat lunak maupun proses bisnis.(Smrti et al., 2023)

Tabel 2. 2 Flowchart (Smrti et al., 2023)

No	Simbol	Nama	Keterangan
1.		*Flow Direction*	Simbol berupa garis panah yang berfungsi untuk menghubungkan satu simbol dengan simbol lainnya, sehingga membentuk alur proses yang runtut dan mudah diikuti.
2.		*Terminator*	Simbol yang digunakan untuk menandai titik awal (*start*) maupun titik akhir (*stop*) dari suatu proses dalam flowchart. Kehadirannya memberikan kejelasan batasan mengenai kapan suatu prosedur dimulai dan kapan berakhir.
3.		*Connector*	Simbol yang berperan untuk menghubungkan bagian-bagian proses yang terdapat dalam satu halaman atau lembar yang sama. Simbol ini membantu menjaga keterbacaan flowchart dengan menghindari garis alir yang terlalu rumit atau saling bersilangan.

No	Simbol	Nama	Keterangan
4.		*Processing*	Simbol yang menunjukkan aktivitas pengolahan data oleh komputer atau sistem. Biasanya digunakan untuk menggambarkan langkah-langkah perhitungan, manipulasi, atau transformasi data.
5.		*Decision*	Simbol yang merepresentasikan titik pengambilan keputusan dalam proses, di mana alur selanjutnya ditentukan berdasarkan kondisi tertentu. Simbol ini menekankan adanya percabangan logika dalam alur program atau sistem.
6.		*Input – Output*	Simbol yang digunakan untuk menggambarkan aktivitas pemasukan data (input) maupun keluaran data (output), tanpa memperhatikan jenis perangkat yang digunakan, baik itu keyboard, scanner, layar monitor, maupun printer.
7.		*Preparation*	Simbol yang menunjukkan proses persiapan data sebelum dimasukkan ke dalam media penyimpanan atau sebelum diproses lebih lanjut. Tahap ini penting untuk memastikan data berada dalam format yang sesuai dengan kebutuhan sistem.

No	Simbol	Nama	Keterangan
8.		*Predefine Proses*	Simbol yang digunakan untuk merepresentasikan eksekusi dari suatu prosedur atau komponen tertentu yang telah didefinisikan sebelumnya.
9.		*Display*	Simbol yang digunakan untuk merepresentasikan perangkat output yang menampilkan informasi hasil pemrosesan(seperti printer, plotter, layar, dan lainnya).
10.		*Punch Card*	Simbol yang digunakan untuk menunjukkan bahwa data output dicetak atau direkam ke dalam kartu berlubang, atau sebaliknya, input yang digunakan berasal dari kartu tersebut.

2.2.14 Python sebagai Tools Implentasi

Python merupakan bahasa pemrograman tingkat tinggi yang dikembangkan oleh Guido Van Rossum dan pertama kali dirilis pada tahun 1991. Hingga saat ini, Python menjadi salah satu bahasa pemrograman yang sangat populer. Karakteristik utamanya adalah sifatnya yang serbaguna (multi-purpose), sehingga dapat digunakan dalam berbagai bidang, termasuk pengembangan Machine Learning dan Deep Learning. Pemilihan Python dalam penelitian umumnya didasarkan pada sintaksisnya yang sederhana dan mudah dipahami, serta ketersediaan pustaka (library) yang sangat beragam dan lengkap. Selain itu, Python bersifat open source dan memiliki dukungan komunitas yang luas, sehingga terus berkembang dan mendapat pembaruan. Dalam praktik penulisan source code, Python dapat dijalankan melalui berbagai Integrated Development Environment (IDE), seperti Visual Studio Code, Sublime Text, maupun PyCharm.

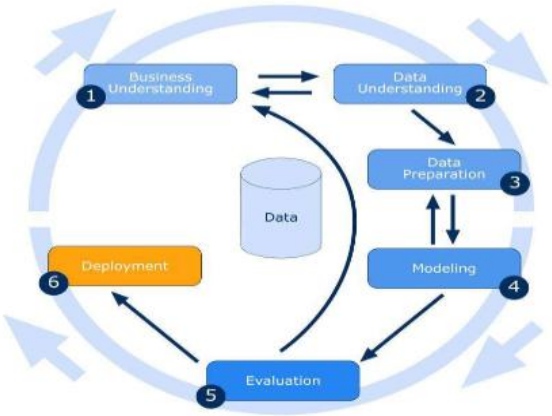
Selain itu, Python juga dapat digunakan pada IDE berbasis daring, misalnya Jupyter Notebook dan Google Colab, yang memudahkan proses eksperimen serta kolaborasi dalam penelitian dan pengembangan perangkat lunak.(Alfarizi et al., 2023)

2.2.15 Metodologi CRISP-DM

CRISP-DM (Cross Industry Standard Process for Data Mining) merupakan kerangka kerja standar yang banyak digunakan dalam praktik data mining. Model ini dirancang untuk memberikan alur kerja yang sistematis, di mana data yang tersedia akan diproses melalui serangkaian tahapan yang terstruktur, terdefinisi dengan jelas, serta efisien. Dengan adanya standar ini, proses data mining menjadi lebih konsisten, dapat direplikasi, serta memudahkan peneliti maupun praktisi dalam mengelola proyek analitik yang kompleks.(Hasanah et al., 2021)

Dalam penelitian ini, digunakan metode CRISP-DM (Cross Industry Standard Process for Data Mining) sebagai salah satu kerangka kerja dalam bidang data mining. Metode ini dinilai memiliki kapabilitas yang tinggi dalam memperluas rekayasa pengembangan aplikasi, mengingat kuatnya integrasi proses yang mendukung kebutuhan bisnis. Selain itu, CRISP-DM oleh berbagai proyek dipandang efektif dalam mengakomodasi serta mengintegrasikan beragam keterampilan lintas disiplin, khususnya dalam ranah Teknologi Informasi.(Puji Catur Siswipraptini 1*, Ahmad Saputra Fadiarora 2, 2023)

Metodologi ini mencakup enam tahapan utama, yaitu Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, serta Deployment. Setiap tahapan tersebut membentuk suatu proses yang saling berkesinambungan, di mana masing-masing fase memiliki peran spesifik yang dapat dijelaskan secara sistematis sesuai dengan alur metodologi.(Hasanah et al., 2021)



Gambar 2. 2 Metodologi CRISP-DM (Hasanah et al., 2021)

Berdasarkan diagram diatas merupakan enam tahapan utama dari CRISP-DM yang saling berhubungan. Berikut adalah penjelasan dari setiap tahapan :

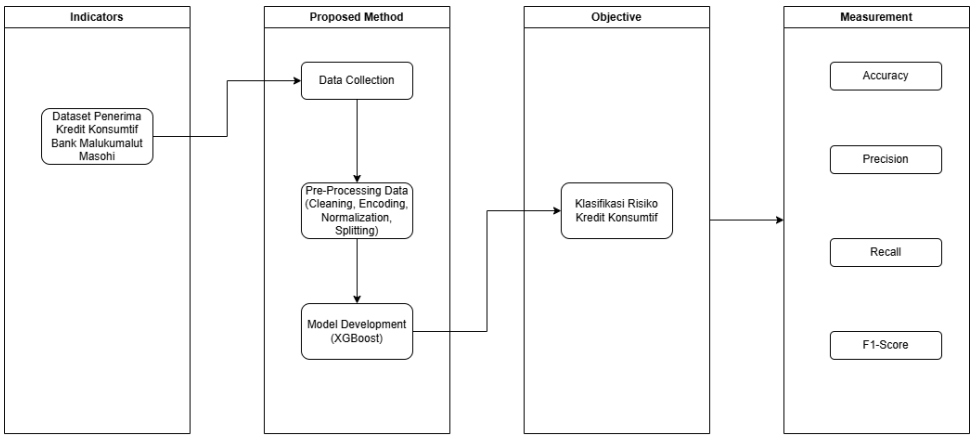
1. Business Understanding: Tahap ini berfokus pada keselarasan keseluruhan proses dengan kebutuhan serta tujuan bisnis. Tujuannya adalah memahami sasaran bisnis secara mendalam dan merumuskan permasalahan yang akan diselesaikan melalui pendekatan data mining.
2. Data Understanding: Pada fase ini dilakukan pengumpulan serta eksplorasi data untuk menelaah kualitas dan karakteristik yang dimilikinya. Selain itu, dilakukan pula identifikasi pola awal yang berpotensi memberikan wawasan awal bagi analisis lanjutan.
3. Data Preparation: Tahap ini mencakup pengolahan data agar siap digunakan dalam pemodelan, meliputi pembersihan (data cleaning), transformasi, serta seleksi atribut sehingga data berada pada format yang sesuai untuk proses analisis lebih lanjut.
4. Modeling: Pada tahap ini dipilih dan diterapkan teknik pemodelan yang relevan untuk melakukan prediksi maupun klasifikasi data. Selanjutnya, model diuji dan dievaluasi guna memastikan efektivitas serta tingkat akurasi yang diperoleh.
5. Evaluation: Fase ini bertujuan untuk menilai kinerja model berdasarkan sasaran yang telah ditentukan sebelumnya, memastikan kesesuaiannya

dengan kebutuhan bisnis, serta melakukan perbaikan atau iterasi apabila diperlukan.

6. Deployment: Tahap akhir ini merupakan proses implementasi model ke dalam lingkungan operasional untuk mendukung pengambilan keputusan bisnis. Selain itu, dilakukan pula pemeliharaan (maintenance) dan pemantauan (monitoring) agar kinerja model tetap optimal dalam jangka panjang.

2.3 Kerangka Pemikiran

Kerangka pemikiran merupakan landasan konseptual yang menjelaskan alur berpikir peneliti dalam menyelesaikan suatu permasalahan penelitian. Pada penelitian ini, kerangka pemikiran dibangun untuk menggambarkan bagaimana metode yang digunakan, yakni Extreme Gradient Boosting (XGBoost), dapat diterapkan dalam proses klasifikasi risiko kredit konsumtif di Bank Malukumalut Masohi.



Gambar 2. 3 Kerangka Pemikiran

Gambar di atas memperlihatkan alur penelitian yang akan dilakukan oleh penulis dalam studi ini, yang penjelasannya akan dipaparkan pada bagian berikut:

1. Indicators: Indikator utama dalam penelitian ini adalah dataset penerima kredit konsumtif yang diperoleh dari Bank Malukumalut Cabang Masohi. Dataset ini memuat berbagai atribut nasabah yang relevan dengan penilaian risiko kredit,

seperti data jumlah pinjaman dan jangka waktu kredit. Data tersebut menjadi dasar bagi model untuk mempelajari pola hubungan antara karakteristik nasabah dengan status kelayakan kreditnya. Informasi-informasi tersebut kemudian dilabeli dengan dua kategori utama, yaitu layak dan tidak layak menerima kredit selanjutnya. Dengan indikator ini, penelitian bertujuan untuk mengembangkan model yang mampu mengidentifikasi pola historis nasabah yang berpotensi menimbulkan risiko kredit tinggi, sehingga bank dapat melakukan tindakan mitigasi sejak dini.

2. Proposed Method: Tahapan metode yang diusulkan terdiri dari tiga langkah utama: data collection, pre-processing data, dan model development. Tahap pertama, data collection, dilakukan untuk mengumpulkan data kredit konsumtif dari sistem internal Bank Malukumalut Masohi. Tahapan ini merupakan fondasi awal karena kualitas dan kelengkapan data akan sangat menentukan akurasi hasil model. Dalam proses ini, peneliti juga memperhatikan aspek privasi dan keamanan data nasabah, sesuai dengan prinsip kerahasiaan perbankan.

   Tahap kedua, pre-processing data, bertujuan untuk menyiapkan data agar layak digunakan dalam proses pemodelan. Langkah ini mencakup data cleaning untuk menghapus data yang tidak konsisten, encoding untuk mengubah data kategorikal menjadi bentuk numerik, serta normalization untuk menyeragamkan skala antar variabel. Setelah itu, data dibagi menjadi tiga bagian, yaitu data latih (training set), data uji (testing set), dan data validasi agar proses evaluasi model menjadi objektif.

   Tahap ketiga adalah model development, yaitu pengembangan model klasifikasi menggunakan algoritma Extreme Gradient Boosting (XGBoost). Algoritma ini dipilih karena kemampuannya dalam menangani data tabular secara efisien, mengatasi overfitting melalui regularisasi, serta menghasilkan performa prediksi yang tinggi. Pada tahap ini, model dilatih untuk mempelajari pola-pola yang membedakan nasabah layak dan tidak layak berdasarkan fitur yang tersedia.
3. Objective: Tujuan utama dari penelitian ini adalah untuk melakukan klasifikasi risiko kredit konsumtif, yaitu menentukan apakah seorang penerima kredit tergolong layak atau tidak layak menerima pinjaman berikutnya berdasarkan

karakteristik datanya. Sistem klasifikasi ini diharapkan mampu memberikan rekomendasi berbasis data yang lebih cepat, objektif, dan akurat dibandingkan penilaian manual oleh analis kredit.

Selain itu, hasil penelitian ini diharapkan dapat membantu Bank Malukumalut Cabang Masohi dalam menekan angka kredit bermasalah atau Non-Performing Loan (NPL) dengan cara mengidentifikasi calon nasabah berisiko tinggi sebelum kredit disetujui. Dengan demikian, penelitian ini tidak hanya berkontribusi pada peningkatan efisiensi operasional bank, tetapi juga memperkuat penerapan prinsip prudential banking dalam proses penyaluran kredit.

4. Measurement (Evaluasi Kinerja): Tahap measurement atau evaluasi kinerja model merupakan proses penting untuk menilai sejauh mana model klasifikasi mampu melakukan prediksi sesuai kondisi sebenarnya. Evaluasi dilakukan menggunakan empat metrik utama, yaitu Accuracy, Precision, Recall, dan F1-Score, yang masing-masing memberikan sudut pandang berbeda terhadap performa model. Accuracy mengukur proporsi prediksi yang benar terhadap seluruh data uji, namun kurang representatif bila data tidak seimbang. Oleh karena itu, Precision dan Recall digunakan untuk analisis yang lebih spesifik. Precision menunjukkan ketepatan model dalam memprediksi nasabah “layak kredit”, sedangkan Recall menilai kemampuan model dalam mengenali seluruh nasabah yang benar-benar layak kredit. F1-Score digunakan untuk menyeimbangkan precision dan recall, terutama pada dataset dengan distribusi kelas tidak merata, sehingga menghasilkan penilaian yang lebih adil terhadap performa model. Dengan mengombinasikan seluruh metrik tersebut, evaluasi kinerja model menjadi lebih komprehensif dan objektif, tidak hanya menilai akurasi secara umum, tetapi juga kestabilan dan keandalan model dalam menghadapi ketidakseimbangan data.