

BAB I

PENDAHULUAN

1.1. Latar Belakang

Dalam sektor keuangan, pemberian kredit merupakan salah satu aspek penting yang memungkinkan individu dan bisnis mendapatkan dana untuk berbagai tujuan, mulai dari kebutuhan pribadi hingga investasi (Waluyo et al., 2020). Namun, setiap pemberian kredit mengandung risiko gagal bayar, sehingga lembaga keuangan perlu menerapkan sistem penilaian kredit yang tepat untuk mengurangi risiko tersebut (Fatimah et al., 2020). Salah satu metode yang digunakan dalam penilaian risiko kredit adalah *Credit Scoring*.

Credit Scoring adalah pendekatan matematis yang digunakan untuk menilai kelayakan kredit individu atau perusahaan dengan mengklasifikasikan mereka ke dalam kategori tertentu berdasarkan profil risiko mereka (Hakim et al., 2020). Model scoring ini umumnya dinyatakan dalam bentuk fungsi probabilitas yang menghubungkan variabel-variabel tertentu dengan kemungkinan gagal bayar (Wahyu & Budianto, 2023). Secara matematis, pendekatan yang sering digunakan adalah *Logistic Regression*, yang merepresentasikan hubungan antara variabel independen (fitur keuangan peminjam) dan variabel dependen (kelayakan kredit) (Daffaa et al., 2025). Di sisi lain, XGBoost merupakan algoritma berbasis pohon keputusan yang memanfaatkan *gradient boosting*, yang bekerja dengan membangun model secara iteratif dan mengoptimalkan hasil untuk meningkatkan akurasi prediksi dengan menyesuaikan bobot kesalahan pada setiap iterasi (Herni Yulianti et al., 2022). Pemanfaatan kedua algoritma ini dalam *credit scoring* sangat krusial untuk meminimalisir risiko gagal bayar serta meningkatkan efisiensi sistem keuangan.

Berbagai penelitian sebelumnya telah dilakukan guna meningkatkan akurasi model *credit scoring*. Diantaranya penelitian oleh Ainul Yaqin mengusulkan penggunaan metode Machine Learning untuk klasifikasi, membandingkan algoritma XGBoost dan Regresi logistik. Data yang digunakan berasal dari penilaian kredit Australia yang diambil dari situs UCI Machine Learning,

dengan 15 variabel. Penulis memilih variabel dengan korelasi di atas 0,3 untuk meningkatkan performa model dan mempercepat proses pelatihan. Variabel yang dipilih meliputi: *BankCustomer* (apakah debitur memiliki rekening bank), A7 (etnis), A8 (tahun bekerja) dengan korelasi tertinggi 0,7, A9 (apakah debitur pernah gagal bayar), dan A10 (apakah debitur bekerja). Setelah dilatih dan diuji menggunakan *stratified k-fold cross-validation*, XGBoost mencapai rata-rata akurasi 85,51%, sementara regresi logistik menghasilkan rata-rata akurasi 85,94%. *Logistic regression* menunjukkan keunggulan pada indikator akurasi, recall, dan F1 Score, sementara XGBoost hanya unggul dalam hal presisi. Selain itu, XGBoost cenderung lebih mengalami overfitting dibandingkan *logistic regression*. Berdasarkan hasil tersebut, dapat disimpulkan bahwa algoritma terbaik untuk penilaian kredit dalam penelitian ini adalah *Logistic regression*. meskipun begitu kedua algoritma ini mampu mengklasifikasikan kelayakan seseorang untuk menerima pinjaman dengan baik, sehingga dapat digunakan untuk membantu analisis kredit (Yaqin, 2022).

Pada penelitian (Amrin & Pahlevi, 2022), beberapa metode klasifikasi dalam data mining diterapkan, termasuk algoritma Logistic Regression dan Support Vector Machine untuk memprediksi kelayakan kredit menggunakan 481 data kredit kendaraan bermotor. Variabel input dalam penelitian ini mencakup 13 variabel diantaranya Status Perkawinan, Status Tempat Tinggal, Kepemilikan Rumah, Jumlah Tanggungan, Umur, Pekerjaan, Status Pekerjaan, Status Perusahaan, Pendidikan, Lama Tinggal, Kondisi Rumah Penghasilan, dan Uang Muka, Hasil penelitian menunjukkan bahwa model *Logistic Regression* menghasilkan akurasi 94,81% dan AUC 0,987, mengungguli SVM yang memiliki akurasi 94,19% dan AUC 0,978. Uji T-Test mengonfirmasi bahwa kinerja *Logistic Regression* secara signifikan lebih baik.

Penelitian oleh (Delima et al., 2023) ini mengajukan penggunaan Algoritma *Extreme Gradient Boosting* (XGBoost) dan *Adaptive Boosting* (AdaBoost) untuk mengidentifikasi risiko pinjaman, yang bertujuan untuk mencegah penunggakan pembayaran utang. Data yang digunakan berasal dari platform kaggle dengan judul *Peer to Peer (P2P) Lending*. Analisis dilakukan dengan mempertimbangkan karakteristik peminjam menggunakan algoritma klasifikasi

yang berbasis pada data pinjaman historis. Dalam penelitian ini, terdapat 34 fitur yang digunakan. Fitur-fitur tersebut diterapkan dalam algoritma XGBoost dan AdaBoost untuk mengklasifikasikan peminjam ke dalam kategori baik dan buruk. Fitur yang paling berpengaruh dalam model pelatihan meliputi *int_rate* (suku bunga pinjaman), *installment* (pembayaran bulanan), dan *loan_amnt* (jumlah pinjaman). Berdasarkan evaluasi menggunakan kurva ROC, algoritma XGBoost menunjukkan performa klasifikasi yang lebih unggul dibandingkan AdaBoost dalam model pemberian pinjaman. Nilai AUC XGBoost sebesar 0,92 lebih tinggi daripada AdaBoost yang hanya 0,89, yang mengindikasikan kemampuan klasifikasi yang lebih baik.

Meskipun telah banyak penelitian mengenai penerapan algoritma dalam credit scoring, masih terdapat beberapa *reaserch gap* yang perlu dieksplorasi lebih lanjut. Salah satu tantangan utama adalah bagaimana mengoptimalkan pemilihan fitur (*feature selection*) dalam model scoring untuk meningkatkan akurasi prediksi. Selain itu, permasalahan data tidak seimbang (*imbalance class*) yang sering terjadi dalam kasus kredit masih menjadi kendala bagi beberapa model klasifikasi tradisional. Oleh karena itu, penelitian ini berupaya untuk mengeksplorasi model *credit scoring* yang lebih efektif dengan membandingkan metode *Logistic Regression* dan XGBoost, serta mengevaluasi kemampuan masing-masing algoritma dalam menangani data yang tidak seimbang dan mengidentifikasi faktor utama yang mempengaruhi kelayakan kredit.

Adapun pemilihan model Logistic Regression dan XGBoost pada penelitian ini memiliki alasan yang cukup jelas. Dimana Logistic Regression digunakan sebagai metode dasar yang sudah lama dipakai dalam credit scoring karena mudah dijelaskan, cepat dilatih, dan menghasilkan koefisien yang bisa langsung diinterpretasikan sebagai pengaruh setiap fitur terhadap risiko gagal bayar (Yaqin, 2022). Keunggulan utamanya adalah kemudahan interpretasi bagi analis kredit dan regulator. Namun, metode ini memiliki batasan karena mengasumsikan hubungan linier antara fitur dan target, serta kurang efektif untuk menangani interaksi variabel yang kompleks dan data yang tidak seimbang tanpa teknik tambahan. Maka dari itu model ini berfungsi menjadi

acuan untuk menilai apakah metode yang lebih kompleks diperlukan (Daffaa et al., 2025).

Di sisi lain, XGBoost mampu menangkap pola non-linier, interaksi antar fitur, dan ketidakseimbangan kelas yang sering muncul pada data kredit (Delima et al., 2023). Menariknya, meskipun mekanisme internalnya berbeda, XGBoost juga menggunakan fungsi logistik pada tahap output untuk menghasilkan probabilitas risiko, sehingga secara konsep tetap sejalan dengan Logistic Regression. Banyak penelitian, baik di Indonesia maupun internasional, menunjukkan bahwa XGBoost sering kali menghasilkan performa lebih baik dibanding Logistic Regression, Random Forest, maupun neural network sederhana pada kasus credit scoring bahkan dengan selisih AUC 0,05–0,15 (Dachi & Sitompul, 2023).

Logistic Regression dan XGBoost memiliki keterkaitan penting dalam proses prediksi skor kredit karena keduanya sama-sama digunakan untuk memperkirakan probabilitas gagal bayar, namun melalui pendekatan yang berbeda. Keterkaitan ini tampak dari tujuan utama keduanya yang sama, yaitu mengukur risiko kredit, di mana Logistic Regression berperan sebagai baseline yang mudah diinterpretasikan, sementara XGBoost menjadi metode yang lebih canggih untuk meningkatkan akurasi dan performa prediksi (Zedda, 2024).

Selain itu, penelitian ini memberikan manfaat penting bagi perusahaan. Pertama, hasil penelitian dapat membantu menekan risiko kerugian akibat meningkatnya Non-Performing Loan (NPL) serta menjadi acuan bagi lembaga keuangan dalam menilai kelayakan calon peminjam, sehingga proses penilaian kredit menjadi lebih efisien (Yaqin, 2022). Kedua, penelitian ini memungkinkan model untuk mengidentifikasi fitur-fitur yang paling berpengaruh terhadap kelayakan kredit, sehingga mendukung pengambilan keputusan yang lebih optimal (Huang et al., 2021). Terakhir, analisis pola nasabah misalnya kecenderungan usia peminjam atau tingginya kredit macet pada jenis pekerjaan tertentu memberikan wawasan berbasis data yang dapat meningkatkan kualitas layanan dan membantu perusahaan dalam mengantisipasi risiko NPL (Pratiwi et al., 2022).

1.2. Identifikasi Masalah

Berdasarkan dari latar belakang masalah yang telah di uraikan maka berikut ini merupakan identifikasi masalah:

1. Perusahaan layanan kredit sering menghadapi masalah kredit macet, di mana pelanggan tidak mampu melunasi pinjamannya. Hal ini dapat menyebabkan kerugian finansial yang signifikan pada perusahaan.
2. Terdapat banyak algoritma *Machine Learning* yang dapat digunakan untuk memprediksi kredit scoring, Masalah yang muncul adalah bagaimana menentukan algoritma yang paling baik dalam memprediksi dengan tepat berdasarkan kriteria tertentu, seperti akurasi, kecepatan, dan interpretabilitas.
3. Dataset kredit sering kali didominasi oleh debitur yang lancar bayar, menyebabkan model cenderung bias terhadap kelas mayoritas dan kurang efektif dalam mendeteksi debitur berisiko. Ketidakseimbangan ini dapat menghambat identifikasi risiko kredit yang tinggi, berpotensi merugikan lembaga keuangan. Oleh karena itu, penanganan bias ini sangat penting untuk meningkatkan akurasi model dalam memprediksi risiko kredit.

1.3. Rumusan Masalah

Berdasarkan hasil dari uraian latar belakang tersebut maka diperoleh rumusan masalah sebagai berikut:

1. Bagaimana cara menerapkan metode *Logistic regression*, dan Xgboost dalam memprediksi hasil skor kredit.
2. Apa saja hasil yang ditemukan dari penerapan *credit scoring*, termasuk analisis model yang telah dilakukan serta solusi yang dapat diberikan untuk mengatasi permasalahan yang telah diidentifikasi.
3. Bagaimana menentukan metode terbaik dari hasil evaluasi yang dihasilkan dengan menggunakan algoritma *Logistic regression* dan Xgboost, dalam memprediksi skor kredit.

1.4. Tujuan

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Menerapkan metode *Logistic Regression* dan Xgboost dalam memprediksi skor kredit serta mengidentifikasi langkah-langkah dan metodologi yang diperlukan dalam proses penerapan model.
2. Mengidentifikasi dan menganalisis hasil yang diperoleh dari penerapan *credit scoring*, serta memberikan hasil solusi yang relevan berdasarkan temuan penelitian untuk meningkatkan strategi pengelolaan risiko kredit.
3. Mengevaluasi dan membandingkan hasil prediksi untuk menentukan metode terbaik dari kedua algoritma, yaitu *Logistic Regression* dan Xgboost dalam memprediksi *credit scoring*,

1.5. Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Penelitian ini hanya menggunakan data yang tersedia dari perusahaan PT Graha Mazindo Mandiri. Data yang digunakan sebagai studi kasus yaitu berupa data nasabah peminjam kredit di PT Graha Mazindo Mandiri yang dimana pada data tersebut mencakup beberapa variabel tertentu.
2. Penelitian ini merupakan simulasi sederhana yang berfokus pada penerapan algoritma machine learning, yaitu Logistic Regression dan XGBoost, untuk memprediksi skor kredit.
3. Evaluasi model dilakukan menggunakan metrik seperti akurasi, presisi, recall, f1 score dan ROC-AUC Score.

1.6. Manfaat Penelitian

Penelitian tentang perbandingan kinerja algoritma Machine Learning dalam memprediksi skor kredit sangat bermanfaat bagi peneliti. Ini memberikan wawasan mengenai kelebihan dan kekurangan berbagai metode, sehingga peneliti dapat mengembangkan pendekatan yang lebih baik di masa depan dan meningkatkan akurasi model. Temuan ini diharapkan menjadi dasar untuk eksplorasi lebih lanjut dalam analisis data dan pembelajaran mesin, serta membuka peluang inovasi dalam penilaian risiko kredit.