

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

*Credit scoring* merupakan suatu sistem penilaian risiko kredit yang banyak digunakan untuk membantu lembaga keuangan atau perusahaan pemberi kredit dalam mengevaluasi calon peminjam, baik individu ataupun perusahaan yang kemungkinan gagal melakukan pembayaran. *Credit scoring* dapat memutuskan apakah calon peminjam tersebut berhak mendapatkan kredit atau tidak. Tujuan dari *credit scoring* ini adalah untuk mengklasifikasikan calon peminjam berkualitas (*good credit*) dan calon peminjam tidak berkualitas (*bad credit*) yang kemungkinan gagal bayar (Bastos, 2022). *Credit scoring* telah menjadi faktor penting dalam industri keuangan, terutama dalam proses pengambilan keputusan pemberian kredit. *Credit scoring* membantu lembaga keuangan dalam mengevaluasi kemungkinan peminjam untuk membayar pinjaman secara tepat waktu yaitu dengan menganalisis data keuangan, perilaku pembayaran, dan informasi kredit lainnya. Oleh karena itu, sangat penting untuk mengembangkan model *credit scoring* yang dapat digunakan secara akurat dan efektif (Xiao, et al., 2021).

*Credit scoring* memiliki tingkat urgensi yang tinggi dalam dunia keuangan karena beberapa alasan. Pertama, dapat membuat keputusan pemberian kredit secara cepat dan efisien. Kedua, dengan mengidentifikasi risiko kredit, lembaga keuangan dapat mengelola portofolio mereka dengan lebih baik untuk menentukan seberapa besar kredit yang akan diberikan dan dapat memantau risiko kredit. Ketiga, lembaga keuangan dapat menilai historis kredit, stabilitas finansial, dan prospek bisnis (Budianto, 2023). Meskipun *credit scoring* memberikan banyak manfaat, terdapat permasalahan yang perlu diatasi. Salah satu permasalahan utamanya adalah masalah klasifikasi.

Klasifikasi dalam *credit scoring* mencakup penentuan apakah seseorang layak atau tidak layak menerima pinjaman berdasarkan skor kredit mereka. Klasifikasi merupakan suatu bentuk analisis data yang mengekstrak model (atau fungsi) serta mendeskripsikan dan membedakan kelas atau konsep data. Klasifikasi juga dikenal

sebagai *supervised learning* (Han, et al., 2012). Klasifikasi sangat penting digunakan khususnya dalam bidang ekonomi.

Pertumbuhan ekonomi Indonesia pada triwulan I 2023 tercatat sebesar 5,03% dan sedikit meningkat dari pada triwulan sebelumnya yaitu 5,01% ini berdasarkan data dari Badan Pusat Statistik (BPS). Data tersebut dapat terlihat melalui peningkatan mobilitas penduduk, stabilitas daya beli masyarakat yang terjaga, terkendalinya laju inflasi, performa ekspor-impor yang tetap kuat, dan kelangsungan aktivitas produksi yang stabil. Oleh karena itu, konsumsi masyarakat baik itu individu maupun perusahaan menjadi meningkat dan lebih percaya diri dalam mengajukan pinjaman untuk berbagai keperluan (Susilawati & Putri, 2019). Hal ini diperkuat dengan adanya indikasi peningkatan dalam penyaluran kredit baru oleh lembaga perbankan pada bulan Agustus 2023 jika dibandingkan dengan bulan sebelumnya. (Bank Indonesia, 19 Sep 2023).

Semakin tinggi jumlah transaksi kredit, maka akan semakin tinggi juga risiko kredit bermasalah. Kredit dikatakan kredit bermasalah atau *non-performing loan* (NPL) apabila kualitas kredit kurang lancar, diragukan, atau macet. Kredit bermasalah ini akan mempengaruhi kondisi keuangan perusahaan karena menurunnya pendapatan dan memungkinkan terjadinya penurunan laba, sehingga perusahaan akan mengalami kerugian (Firmanto, 2019). Oleh karena itu, perusahaan harus lebih cermat dalam memilih calon peminjam yang berkualitas agar dapat mengurangi risiko kredit. Salah satu cara dalam mengurangi risiko kredit yaitu dengan melakukan *credit scoring*.

Beberapa metode yang dapat digunakan untuk klasifikasi dalam *credit scoring* yaitu, *Decision Tree*, Naïve Bayes, *K-Nearest Neighbor* (K-NN), *Support Vector Machines* (SVM), *Neural Networks*, *Logistic Regression* dan *Ensemble Methods* (Bastos, 2022). Salah satu metode dalam klasifikasi yang sering digunakan adalah *Decision Tree*. *Decision Tree* merupakan salah satu metode klasifikasi yang populer karena algoritmanya yang sederhana sehingga mudah dipahami dan mampu menangani tipe data campuran. Selain *Decision Tree*, *Logistic Regression* juga merupakan salah satu metode klasifikasi yang memerankan peran penting dalam mengevaluasi strategi *machine learning* dengan algoritmanya yang mudah dimengerti (Shah, et al., 2020). Algoritma *Decision Tree* ini mampu melakukan

klasifikasi data dengan berbagai keunggulan yaitu dapat mengolah data numerik dan diskrit, menangani atribut yang hilang, menghasilkan aturan-aturan yang mudah dimengerti, dan memiliki kecepatan eksekusi tercepat. Algoritma ini telah terbukti dapat memberikan tingkat akurasi dan kinerja yang sangat baik (Amin, et al., 2015). Beberapa penelitian yang melakukan *credit scoring* seperti yang dilakukan oleh Naufal, et al. (2023) yang mengusulkan metode Naïve Bayes dan *Decision Tree* untuk memprediksi potensi hilangnya nasabah bank. Akurasi terbaik diperoleh dari metode *Decision Tree* dengan nilai akurasi mencapai 93%. Pada penelitian Jadhav & Chane (2016) *Decision Tree* dinilai lebih akurat, memiliki tingkat kesalahan yang lebih rendah, dan lebih mudah dipahami dibandingkan K-NN dan Naïve Bayes. Pada kasus mendeteksi penipuan kartu kredit oleh Alenzi & Aljehane (2020) metode *Logistic Regression* memperoleh hasil akurasi tertinggi sebesar 97,2% dibandingkan dengan metode K-NN dan *Voting Classifier*. Selain itu, metode *Logistic Regression* pada penelitian Silva, et al. (2020) tentang *credit scoring* lembaga keuangan Portugis memperoleh hasil akurasi sebesar 89,79%.

Akan tetapi, model *Decision Tree* mudah dipengaruhi oleh *data noise* dan atribut data yang berlebihan sehingga membuat model *Decision Tree* menjadi lemah. Sedangkan pada model *Logistic Regression* kesulitan untuk bekerja dengan baik pada kumpulan data yang tidak seimbang (Zhang & Chen, 2021). Untuk mengatasi masalah tersebut dapat menggunakan metode *ensemble* seperti *Random Forest*, *Bagging (Bootstrap Aggregating)*, *AdaBoost (Adaptive Boosting)*, dan *Gradien Boosting* (Wang, et al., 2012). Penelitian Wang, et al. (2012) mengusulkan metode RS-Bagging DT dan Bagging-RS DT yang didasarkan pada dua strategi *ensemble* yaitu *bagging* dan *random subspace*. Hasil penelitian ini terbukti dapat meningkatkan akurasi *Decision Tree* dibandingkan dengan *Decision Tree* tunggal. Berdasarkan penelitian yang berjudul “*Application of Ensemble Models in Credit Scoring Models*” oleh Chopra & Bhilare (2018) metode *ensemble* juga terbukti mampu meningkatkan akurasi pada *Decision Tree*. Pada penelitian penilaian risiko kredit oleh Tian, et al. 2020 metode *Gradient Boosting Decision Tree* mampu meningkatkan nilai akurasi *Decision Tree* dan memperoleh nilai akurasi tertinggi yaitu 92,19%. Pada penelitian yang telah dilakukan oleh Bastos (2022) yang membahas mengenai prediksi *credit scores* menggunakan *Boosted Decision Tree*

dengan yang menjadi *Boosted*-nya itu adalah metode *AdaBoost*. Hasil dari penelitian tersebut membuktikan bahwa metode *AdaBoost* mampu meningkatkan akurasi *Decision Tree* dan tingkat keakuratan *Boosted Decision Tree* ini lebih tinggi daripada *Multilayer Perceptron* dan *Support Vector Machine (SVM)*. Selain dapat meningkatkan nilai akurasi dari *Decision Tree*, *AdaBoost* juga dapat meningkatkan nilai akurasi dari *Logistic Regression* (Zhang & Chen, 2021). Hal ini menunjukkan bahwa metode *AdaBoost* merupakan sebuah teknik yang kompetitif untuk meningkatkan kinerja model *Decision Tree* dan *Logistic Regression*. Oleh karena itu, berdasarkan pemaparan tersebut, peneliti tertarik mengkaji lebih dalam mengenai *credit scoring* menggunakan *ensemble learning* dengan membandingkan metode *Decision Tree – AdaBoost* dan *Logistic Regression – AdaBoost* dengan menggunakan variabel yang lebih beragam dan kompleks.

Data yang digunakan pada penelitian ini adalah data perusahaan *Home Credit*. *Home Credit* adalah lembaga keuangan non-bank yang menyediakan jasa pembiayaan atau pinjaman bagi mereka yang membutuhkan.

Tujuan dari penelitian ini adalah mengidentifikasi calon peminjam yang berpotensi memiliki risiko rendah untuk gagal membayar pinjaman. Perusahaan dapat menggunakan model prediktif untuk melakukan penilaian risiko yang lebih baik pada calon peminjam dengan memahami faktor-faktor yang mempengaruhi *default* atau keterlambatan pembayaran.

Penelitian ini diharapkan dapat memberikan kontribusi pada perusahaan penyedia layanan kredit baik itu lembaga perbankan maupun lembaga non-bank dalam mengelola risiko kredit di Indonesia. Penelitian ini juga diharapkan dapat menjadi acuan dalam membuat keputusan mengenai calon peminjam. Oleh karena itu, penelitian ini berjudul “Metode *Hybrid Decision Tree – Adaptive Boosting* (Studi Kasus: Klasifikasi *Credit Scoring*)”.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka rumusan masalah dalam penelitian ini adalah bagaimana hasil perbandingan model *Decision Tree – AdaBoost* dan model *Logistic Regression – AdaBoost* dalam menentukan klasifikasi *credit scoring* pada perusahaan *Home Credit*?

### 1.3 Tujuan Penelitian

Berdasarkan rumusan masalah tersebut, maka tujuan dalam penelitian ini adalah mengetahui model terbaik dengan menghitung tingkat akurasi dari model *Decision Tree – AdaBoost* dan model *Logistic Regression – AdaBoost* dalam menentukan klasifikasi *credit scoring* pada perusahaan *Home Credit*.

### 1.4 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut.

#### 1. Manfaat Penelitian Praktis

Penelitian ini dapat meningkatkan akurasi penilaian kredit. Perusahaan penyedia layanan kredit baik itu lembaga perbankan maupun lembaga non-bank dapat menggunakan hasil penelitian ini sebagai acuan dalam mengelola risiko kredit. Hasil penelitian ini dapat digunakan untuk mengembangkan atau memperbarui sistem *credit scoring* yang lebih efisien, sehingga meningkatkan keandalan dan akurasi dalam evaluasi kredit. Selain itu, penelitian ini dapat membantu dalam optimalisasi proses pemilihan calon peminjam dengan mengetahui model terbaik. Sehingga perusahaan penyedia layanan kredit dapat lebih selektif dalam proses seleksi calon peminjam dan dapat lebih cepat serta tepat dalam menentukan kelayakan kredit.

#### 2. Manfaat penelitian teoritis

Penelitian ini memberikan kontribusi pada literatur *data mining* dan *machine learning*, khususnya mengenai penggunaan metode *hybrid* dalam konteks *credit scoring*. Studi ini juga memperluas pengetahuan tentang efektivitas metode *hybrid Decision Tree – AdaBoost* dan *Logistic Regression – AdaBoost*, yang dapat menjadi dasar bagi penelitian selanjutnya dalam mengembangkan model-model yang lebih canggih. Selain itu, penelitian ini menyediakan evaluasi komprehensif mengenai performa berbagai model dalam klasifikasi *credit scoring*, yang dapat menjadi referensi bagi peneliti lain dalam memilih dan mengembangkan model yang sesuai.

