BAB I
PENDAHULUAN

1.1 Latar Belakang

Kredit adalah penyediaan uang atau tagihan atas kesepakatan dan persetujuan pinjam-meminjam antara bank dengan pihak lain yang mewajibkan peminjam untuk melunasi hutangnya dengan pemberian bunga setelah jangka waktu tertentu (Hanun dan Zailani, 2020). Iriadi dan Leidiyana (2013) menyebutkan bahwa kredit yang diajukan oleh debitur mempunyai risiko. Hal ini dikarenakan adanya kemungkinan debitur yang mendaftar namun mengalami kesulitan dalam pembayarannya sehingga menyebabkan macetnya kredit. Sebelum menyetujui pinjaman yang diajukan oleh debitur, perbankan dapat melakukan analisis kredit terhadap debitur untuk menentukan apakah permohonan kredit akan disetujui atau tidak disetujui. Analisis kredit merupakan analisis yang berkaitan dengan faktor-faktor yang mempengaruhi kelancaran atau tidaknya pengembalian kredit.

Hanun dan Zailani (2020) mengatakan pengambilan keputusan disetujui atau tidaknya seorang debitur dapat menggunakan data historis debitur yang disetujui oleh bank. Namun, perlu diperhatikan bahwa tidak semua debitur yang disetujui adalah pembayar kredit yang baik. Ada beberapa debitur yang sudah disetujui tetapi terlambat membayar setelah beberapa bulan sehingga menyebabkan risiko terjadinya kredit macet.

Risiko kredit macet dapat diminimalkan dengan analisis menggunakan teknik *data mining*. *Data mining* merupakan proses dengan menggunakan statistika, matematika, kecerdasan buatan dan *machine learning* untuk mengekstrak dan mengidentifikasi informasi yang berguna dan pengetahuan yang relevan dari *database* besar. *Data mining* adalah sekumpulan proses yang mengeksplorasi nilai tambah dari kumpulan data berupa pengetahuan yang tidak dapat dipelajari secara manual (Bramer, 2007). Pada prosesnya, *data mining* mengekstrak informasi berharga dari sejumlah besar data dengan

menganalisis keberadaan pola dan keterkaitan hubungan tertentu (Iriadi dan Nuraeni, 2016).

Salah satu proses pemecahan dalam *data mining* adalah proses klasifikasi. Klasifikasi dalam *data mining* adalah proses mengklasifikasikan suatu objek atau konsep tertentu ke dalam sekumpulan kategori berdasarkan sifat dari objek atau konsep yang bersangkutan (Hanun dan Zailani, 2020). Metode klasifikasi bertujuan untuk mempelajari fungsi yang berbeda. Setiap fungsi memetakan masing-masing data yang dipilih ke salah satu kelompok kelas yang telah ditentukan sebelumnya (Hanun dan Zailani, 2020). Metode klasifikasi yang baik akan mengurangi kemungkinan kesalahan klasifikasi atau risiko kesalahan alokasi yang rendah (Johnson dan Wichern, 2007). Klasifikasi termasuk dalam *supervised learning*. Metode *supervised learning* didasarkan pada data latih. Pada data latih, jarang ditemukan jumlah distribusi data yang sama untuk masing-masing kelas. Jika salah satu kelas memiliki jumlah yang sangat besar dibandingkan kelas lainnya, maka data dianggap tidak seimbang (*imbalanced data*) (Arifiyanti dan Wahyuni, 2020).

Ketidakseimbangan kelas merupakan salah satu faktor yang paling berpengaruh dalam kemampuan prediksi suatu klasifikasi. Metode klasifikasi seperti *Classification Tree* cenderung membuat model *learning* dengan akurasi prediksi yang bias untuk kelas minoritas jika dibandingkan kelas mayoritas. Hal ini dikarenakan *Classification Tree* dirancang untuk asumsi distribusi kelas yang seimbang (Prasetya, 2022). Data tidak seimbang adalah suatu keadaan data yang tidak seimbang antara kelas data pertama dengan kelas data lainnya. Ketidakseimbangan data menjadi masalah dalam klasifikasi karena model klasifikasi cenderung membuat prediksi tentang kelas data yang besar (mayoritas) dibandingkan dengan kelas data yang kecil (minoritas). Hal ini mengakibatkan akurasi prediksi yang baik untuk kelas mayoritas dan akan menghasilkan akurasi prediksi yang buruk untuk kelas minoritas pada data latih (Fitriani, Yasin, dan Tarno, 2021). Salah satu cara untuk menangani ketidakseimbangan data adalah dengan melakukan teknik *Random Oversampling*. Teknik *Random Oversampling* dapat menangani

ketidakseimbangan data dengan cara menambah data pada kelas minoritas tanpa mengurangi jumlah data pada kelas lainnya.

Decision Tree disebut sebagai *Classification Tree* jika variabel targetnya bertipe kategorik. Sedangkan jika variabel targetnya berupa numerik, metode ini sering disebut dengan metode *Regression Tree* (Mardika, Mukid, dan Yasin, 2016). *Classification Tree* adalah metode yang dapat mengubah data yang sangat besar menjadi *rules* dalam bentuk struktur pohon yang mewakili keputusan, dengan *rules* yang mudah dipahami dalam bahasa alami (Bramer, 2007). Data *Classification Tree* biasanya direpresentasikan dalam bentuk tabel variabel dan *record*. Variabel menentukan parameter yang dibuat sebagai kriteria saat pembentukan pohon. Variabel yang menunjukkan data solusi untuk setiap item data disebut variabel target sedangkan variabel memiliki nilai yang disebut *instance*. Proses dalam *Classification Tree* yaitu mengubah data yang berbentuk tabel menjadi model pohon, mengubah model pohon menjadi *rules* dan menyederhanakan *rules* (Bramer, 2007). Metode *Classification Tree* merupakan metode non parametrik sehingga tidak perlu memperhatikan asumsi normalitas data. Struktur data dapat dilihat secara visual untuk mempermudah eksplorasi data dan pengambilan keputusan berdasarkan model yang diperoleh. *Classification Tree* bertujuan untuk menghasilkan klasifikasi yang akurat dan menginterpretasikan prediksi data baru di setiap kategori yang terdapat dalam respon (Malta dan Sutikno, 2019).

Penelitian ini dilakukan untuk menangani ketidakseimbangan kelas pada data kredit debitur salah satu bank di Kalimantan Barat dengan teknik *Random Oversampling* menggunakan metode klasifikasi, yaitu *Classification Tree*. Hasil dari penelitian ini dapat digunakan sebagai dasar pengambilan keputusan penilaian kelayakan debitur di suatu instansi.

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang, maka rumusan masalah dalam penelitian ini yaitu:

1. Apakah penerapan metode *Classification Tree* pada *imbalanced data* lebih baik dengan teknik *Random Oversampling* atau tanpa teknik *Random Oversampling*?
2. Variabel apa saja yang berpengaruh pada kelayakan pemberian kredit?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dibuat, maka tujuan penelitian yang ingin dicapai dalam penelitian ini yaitu:

1. Untuk membandingkan apakah penerapan metode *Classification Tree* pada *imbalanced data* lebih baik dengan teknik *Random Oversampling* atau tanpa teknik *Random Oversampling*.
2. Untuk menentukan variabel yang berpengaruh pada kelayakan pemberian kredit.

1.4 Batasan Masalah

Berdasarkan uraian pada latar belakang, maka batasan masalah dalam penelitian ini yaitu:

1. Tingkat evaluasi kinerja model klasifikasi menggunakan *confusion matrix*.
2. Proporsi data latih dan data uji yang digunakan adalah proporsi 70:30, 80:20 dan 90:10. Pada proporsi 70:30, diperoleh data latih sebanyak 560 data dan data uji sebanyak 240 data. Pada proporsi 80:20, diperoleh data latih sebanyak 640 data dan data uji sebanyak 160 data. Pada proporsi 90:10, diperoleh data latih sebanyak 720 data dan data uji sebanyak 80 data.

1.5 Tinjauan Pustaka

Kelancaran dalam pembayaran kredit akan meningkatkan dan mengembangkan kegiatan ekonomi negara. Pinjaman yang diberikan oleh

bank dalam bentuk kredit berasal dari modal masyarakat, sehingga memiliki risiko (*risk asset*) yang tinggi dengan tanpa dilunasinya kredit tepat pada waktunya yang dikenal sebagai *Non-Performance Loan* (NPL) (Nursyahriana, Hadjat & Tricahyadinata, 2017). Prinsip pemberian kredit oleh bank dikenal dengan analisis 5C, yaitu *Character, Capacity, Capital, Condition of economy* dan *Collateral* (Hadiwidjaya & Wirasasmita, 2007).

Kredit macet adalah keadaan dimana debitur mengalami kesulitan dalam membayar kewajibannya kepada bank karena faktor kesengajaan atau faktor eksternal di luar kendali debitur (Hohedu & Dewi, 2019). Kredit macet dapat menyebabkan kerugian pada bank. Hal ini berarti kerugian karena tidak diterimanya dana yang disalurkan dan pendapatan bunga yang tidak dapat diterima. Akibatnya, bank kehilangan kesempatan untuk memperoleh bunga, yang menyebabkan penurunan total pendapatan (Nursyahriana, Hadjat & Tricahyadinata, 2017).

Martha, Andani, dan Rizki (2022) melakukan penelitian tentang Perbandingan Metode *K-Nearest Neighbor*, Regresi Logistik Biner, dan Pohon Klasifikasi pada Analisis Kelayakan Pemberian Kredit. Penelitian ini membandingkan hasil akurasi dari metode *K-Nearest Neighbor*, Regresi Logistik Biner, dan Pohon Klasifikasi. Hasil penelitian menunjukkan bahwa metode Pohon Klasifikasi menjadi metode terbaik karena memiliki tingkat akurasi yang paling baik, yaitu dengan akurasi sebesar 87,68%.

Sartika dan Sensuse (2017) telah melakukan penelitian tentang perbandingan algoritma klasifikasi pada studi kasus pengambilan keputusan pemilihan pola pakaian. Kesimpulan dari penelitian ini yaitu, pada kasus pengambilan keputusan pemilihan pola pakaian menyatakan bahwa algoritma klasifikasi *Decision Tree* memiliki tingkat akurasi paling tinggi dengan tingkat akurasi sebesar 75.6%.

Fitriani, Yasin, dan Tarno (2021) telah melakukan penelitian terkait penanganan klasifikasi kelas data tidak seimbang dengan *Random Oversampling* pada *Naive Bayes*, dengan studi kasus status Peserta KB IUD di Kabupaten Kendal. Kesimpulan dari penelitian ini yaitu hasil klasifikasi status

keberhasilan program KB tahun 2018 menggunakan *Random Oversampling* merupakan hasil klasifikasi terbaik dengan nilai sensitivitas sebesar 0,974 dan *G-mean* sebesar 0,784 pada data latih.

1.6 Metodologi Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari salah satu bank yang ada di Kalimantan Barat. Adapun variabel yang digunakan dalam penelitian ini yaitu sebanyak 11 variabel, yaitu 10 variabel independen yang mencakup *limit*, *rate*, tenor, total angsuran, jenis kebutuhan, gaji, premi dan biaya admin, instansi, jenis kredit dan usia. Sedangkan untuk variabel target berjumlah sebanyak 1 variabel, yakni kolektibilitas. Adapun teknik analisis data yang digunakan dalam penelitian ini adalah klasifikasi menggunakan metode *Classification Tree* dengan teknik *Random Oversampling* untuk mengatasi *imbalanced data*. Proses analisis dalam penelitian ini menggunakan *software* R Studio. Pada penelitian ini, akan dilakukan proses analisis *data mining* menggunakan metode *Classification Tree* untuk data debitur lancar dan non lancar pada data kredit yang diperoleh dari salah satu bank di Kalimantan Barat. Proses yang dilakukan adalah sebagai berikut:

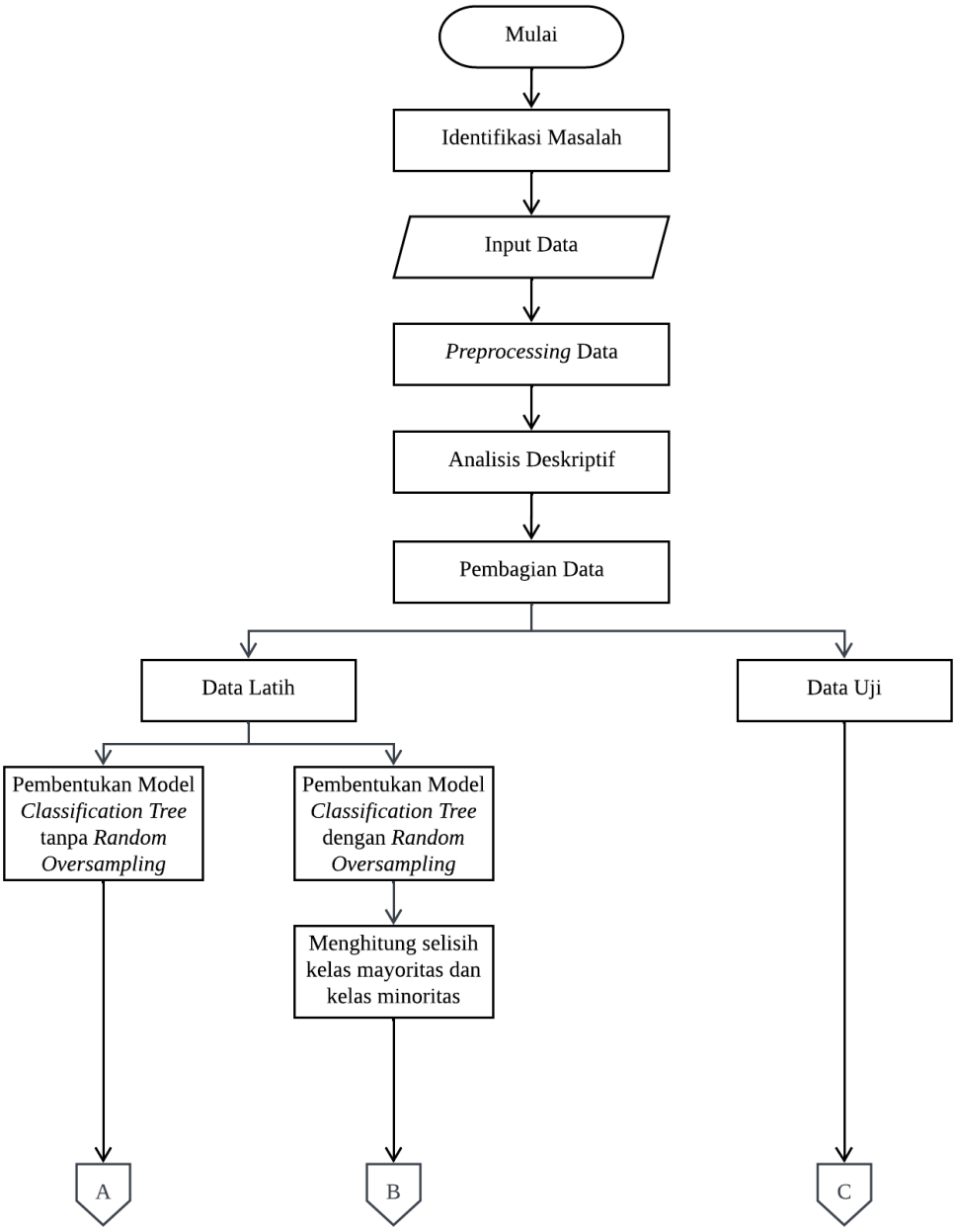
a) Identifikasi masalah,
b) Input data,
c) *Preprocessing data*,
d) Analisis deskriptif,
e) Pembagian data latih dan data uji,
f) Pembentukan model *Classification Tree* tanpa *Random Oversampling* menggunakan data latih,
g) Pembentukan model *Classification Tree* dengan *Random Oversampling* menggunakan data latih, dengan meliputi berbagai proses dari *Random Oversampling*,
h) Validasi model *Classification Tree* tanpa *Random Oversampling* dan *Classification Tree* dengan *Random Oversampling* menggunakan data uji,

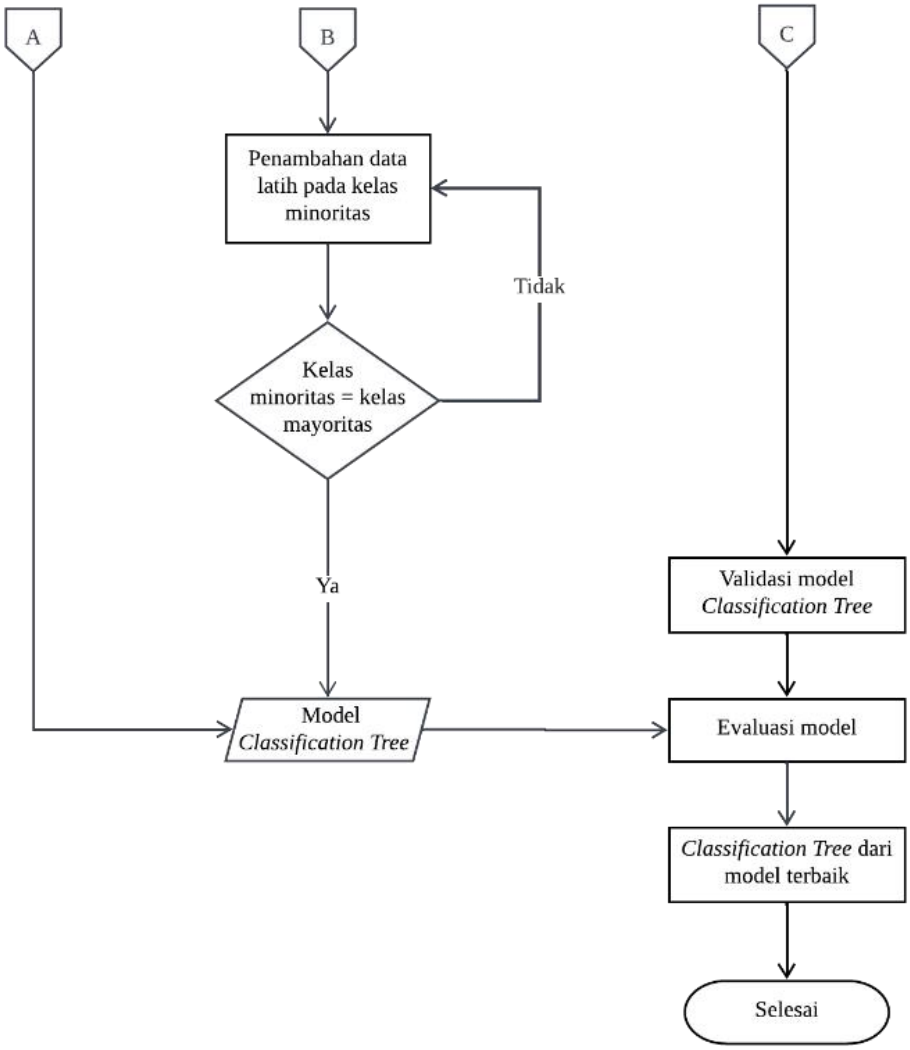
i) Evaluasi model *Classification Tree* tanpa *Random Oversampling* dan *Classification Tree* dengan *Random Oversampling*,

j) Menampilkan *Classification Tree* dari model terbaik,

k) Menarik kesimpulan.

Alur dari penelitian ini dapat dilihat pada Gambar 1.1.





Gambar 1.1 *Flowchart* Penelitian