BAB I
PENDAHULUAN

1.1 Latar Belakang

Semakin meningkatnya permintaan pinjaman dari debitur mendorong setiap bank untuk mengembangkan teknik pengambilan keputusan yang lebih baik dan efisien. Salah satu kriteria bank yang baik adalah bank yang dapat meminimalkan risiko kredit. Dalam setiap penyaluran kredit terkandung risiko di dalamnya yang dinamakan risiko kredit. Pengambilan keputusan yang tepat dalam penyaluran kredit dapat menekan kerugian bank yang disebabkan oleh kredit macet. Kegagalan bank dalam mendeteksi secara dini debitur yang berpotensi macet dapat menyebabkan kerugian tersebut. Maka, perlu dibuat sebuah model yang dapat membedakan antara debitur "baik" dan "buruk" dalam hal kemampuan membayar berdasarkan profilnya.

Nasabah menurut Kamus Besar Bahasa Indonesia (KBBI) adalah orang yang biasa berhubungan dengan atau menjadi pelanggan bank (dalam hal keuangan), dapat juga diartikan sebagai orang yang menjadi tanggungan asuransi, perbandingan pertalian. Sementara itu menurut Undang-Undang No. 24 Tahun 2004 tentang Lembaga Penjamin Simpanan membagi nasabah menjadi dua yaitu, nasabah penyimpan dan nasabah debitur. Nasabah penyimpan adalah nasabah yang menempatkan dananya di bank dalam bentuk simpanan berdasarkan perjanjian bank dengan nasabah yang bersangkutan. Nasabah debitur adalah nasabah yang memperoleh fasilitas kredit atau pembiayaan berdasarkan prinsip

syariah atau yang dipersamakan dengan itu berdasarkan perjanjian bank dengan nasabah yang bersangkutan.

Debitur "buruk" adalah debitur yang berpotensi untuk masuk kedalam kategori *Non Performing Loan* (NPL), sedangkan debitur baik adalah debitur yang tidak berpotensi masuk kedalam kategori NPL. Debitur masuk kedalam kategori NPL jika lama tunggakannya diatas 91 hari. Rasio NPL menjadi salah satu indikator penilaian kesehatan bank yang dilakukan oleh BI dan Otoritas Jasa Keuangan (OJK). Bank Indonesia (BI) melalui peraturan Bank Indonesia (PBI) menetapkan bahwa rasio NPL ideal adalah sebesar 5%.

Saat ini perkembangan teknologi dalam penyimpanan data memungkinkan setiap bank memiliki jumlah data yang sangat besar. Pembangunan model prediktif pun harus diikuti dengan kemampuan metode yang dapat digunakan untuk dataset yang besar. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait sebagai database besar. Salah satu teknik dalam data mining adalah klasifikasi yang termasuk ke dalam bidang ilmu statistika. Klasifikasi menghasilkan suatu model berdasarkan data yang dianalisis kemudian melakukan prediksi berdasarkan pola data sebelumnya. Proses klasifikasi terdiri dari 2 fase; pertama *learning process,* yaitu melakukan analisis pada dataset *training* dengan menggunakan algoritma klasifikasi. Fase ini menghasilkan model yang disajikan dalam bentuk aturan klasifikasi atau pola. Lalu, Fase kedua adalah

penggunaan model klasifikasi dan penggunaan dataset *testing* untuk menghitung akurasi dari model yang telah dibuat [1].

Pada kenyataannya di lapangan sering terjadi data tidak lengkap (*missing value*). Adanya *missing value* pada data yang dianalisis dapat menurunkan keakuratan dan kualitas dalam pengambilan keputusan. Menghapus observasi yang mengandung *missing value* dapat merugikan karena akan menyebabkan hilangnya informasi yang telah didapatkan. Dibutuhkan waktu dan biaya yang tidak sedikit untuk mengumpulkan data. Oleh karena itu, dibutuhkan *missing value* ini harus diimputasi sebelum dilakukan analisis.

1.2 Permasalahan Penelitian

Dalam kaitannya dengan prediksi debitur berpotensi gagal bayar, salah satu penggunaan *Decision Tree* sebagai analisis pendahuluan bagi Regresi Logistik dapat meningkatkan akurasi dari model Regresi Logistik yang dihasilkan. Namun pada kenyataannya di lapangan sering terjadi *missing value,* sehingga perlu dilakukan estimasi untuk *missing value* tersebut. Pada penelitian ini akan dilakukan penggabungan metode *Amelia II , Decision Tree,* dan Regresi logistik untuk mendapatkan model dengan akurasi yang tinggi pada data tidak lengkap.

1.3 Identifikasi Masalah

Untuk menghasilkan metode yang memiliki akurasi tertinggi dalam pengklasifikasian debitur dapat ditemukan identifikasi masalah berikut :

1) Seberapa akurat penggabungan metode Amelia II, *Decision Tree* dan Regresi Logistik dalam melakukan prediksi debitur berpotensi gagal bayar pada data tak lengkap?

1.4 Tujuan Penelitian

Tujuan dari penelitian adalah untuk menghasilkan algoritma yang mampu memprediksi risiko debitur dengan kondisi data tidak lengkap menggunakan metode *Amelia II, Decision Tree,* dan Regresi Logistik.

1.5 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah :

1) Dari sisi keilmuan, kontribusi utama dari penelitian ini adalah dihasilkannya metode gabungan dari *Amelia II*, *Decision Tree*, dan Regresi Logistik untuk prediksi debitur berpotensi gagal bayar dengan data tak lengkap dengan akurasi tinggi.
2) Implementasi dari metode ini di lingkungan bisnis dalam suatu bank dapat mengurangi resiko terjadinya debitur gagal bayar. Sehingga bank dapat menghindari kerugian sebelum kerugian itu terjadi.

1.6 Ruang Lingkup Penelitian

Batasan masalah dari penelitian ini adalah sebagai berikut :

1) Metode yang digunakan untuk estimasi *missing value* adalah Amelia II

2) Metode *Decision Tree* yang digunakan sebagai seleksi dan pengkategorian ulang variabel adalah CART
3) Metode untuk pembuatan model klasifikasi menggunakan Regresi Logistik
4) Data yang dianalisis adalah *German Credit Dataset* yang diambil dari UCI *repository of machine learning.*
5) Penelitian ini hanya sampai dengan usulan metode, tidak sampai dengan pembuatan aplikasi.

1.7 Sistematika Penulisan

Tesis ini menjelaskan mengenai penggabungan metode Amelia II, CART, dan Regresi Logistik untuk mengklasifikasikan debitur kredit pada dataset dengan *missing value.* Isi dari tesis dijelaskan dalam beberapa bab yaitu :

1) Bab II : Tinjauan Pustaka menjelaskan mengenai penelitian terdahulu yang menjadi dasar dalam penelitian ini. Teknik perhitungan dari metode yang digunakan yaitu Amelia II, CART, dan Regresi Logistik. Di bab ini juga dijelaskan mengenai kelebihan dan kekurangan dari metode yang digunakan.
2) Bab III : Metodologi Penelitian menjelaskan mengenai tahapan penelitian yang dilakukan. Diagram alur tahapan penelitian dan penjelasan mengenai setiap tahap yang dilakukan. Dijelaskan juga mengenai data yang digunakan dalam penelitian.
3) Bab IV : Hasil Penelitian dan Pembahasan menjelaskan mengenai hasil penelitian dari setiap tahapan yang dilakukan. Proses imputasi dari dataset yang mengandung *missing value.* Proses Pereduksian dan pengkategorian ulang atribut dan pohon klasifikasi yang dihasilkan oleh CART dari masing –

masing dataset. Analisis secara keseluruhan dari hasil pereduksian dan pengkategorian ulang menggunakan CART. Model yang dihasilkan oleh regresi logistik untuk masing – masing dataset dan akurasi yang dihasilkan dalam bentuk nilai CCR. Penyajian nilai CCR dari masing – masing dataset dalam bentuk grafik.

4) Bab V : Kesimpulan dan Saran berisi kesimpulan yang dihasilkan dari hasil penelitian dan saran yang untuk penelitian selanjutnya.
5) Daftar Pustaka berisi penelitiaan lain yang mendukung penelitian ini.
6) Lampiran berisi peta sebaran missing value dan keluaran hasil analisis regresi logistik untuk masing – masing dataset.