

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi saat ini meningkat cukup pesat, hingga munculnya *Financial Technology* atau *Fintech* yang merupakan salah satu perkembangan teknologi bidang ekonomi dan keuangan. Seperti kebanyakan perusahaan *Financial Technology* yang dimana mempunyai fokusnya masing-masing seperti; Pembayaran (*payment*), Peminjaman (*Lending*), Perencanaan Keuangan (*Personal Finance*), Investasi Ritel, dll. Oleh karena itu saat ini peminjaman online [1] lebih banyak dilirik oleh masyarakat, karena peminjaman keuangan dapat dilakukan dengan cara online tanpa harus bertemu secara langsung.

Dalam konteks ini, pentingnya untuk mencari sebuah solusi yang efektif untuk dapat mengidentifikasi dan memitigasi risiko kredit macet. Salah satunya dengan menggunakan salah satu pendekatan yang menjanjikan yaitu dengan penggunaan teknik *machine learning*, lebih tepatnya dengan menggunakan algoritma *Random Forest*. Dengan menggunakan teknik ini terbukti lebih efektif dalam berbagai bidang, lebih tepatnya pada prediksi risiko kredit. Model *Random Forest* [2] ini juga dikenal dapat menggabungkan prediksi dari beberapa pohon keputusan untuk menghasilkan hasil prediksi yang lebih akurat dan stabil. Keunggulan utama ini yang membuat metode ini sangat cocok untuk menangani jumlah besarnya data, variabel yang berkorelasi dan menghindari *overfitting*.

Penelitian yang menunjukkan dengan *machine learning Random Forest* memiliki sebuah kemampuan untuk dapat menangani data yang kompleks dan tidak seimbang, serta dapat memberikan hasil yang lebih baik dibandingkan dengan beberapa metode tradisional seperti regresi logistik. Misalnya, pada sebuah studi yang dilakukan oleh yangudongnan Guo menunjukkan bahwa penggunaan *Weighted Random Forest* [3] untuk hal klasifikasi risiko kredit dapat meningkatkan sebuah akurasi klasifikasi data kredit secara signifikan. Selain itu, penelitian lain yang dilakukan oleh Amrin et al [2] berhasil mendapatkan hasil dengan algoritma

Random Forest mampu untuk memberikan sebuah hasil prediksi hingga 91% dalam pengelolaan risiko kredit.

Di Indonesia, dimana penggunaan pinjaman kian semakin meningkat dari tahun ke tahun, penerapan *Random Forest* untuk klasifikasi risiko pinjaman dapat menjadi salah satu langkah yang relevan. Dengan adanya data historis peminjaman yang cukup beragam, algoritma ini diharapkan agar dapat membantu lembaga keuangan dalam menilai sebuah kelayakan kredit [4] dan dapat memprediksi kemungkinan terjadinya gagal bayar. Penelitian sebelumnya yang dilakukan oleh Prasojo & Haryatmi menunjukkan hasil bahwa model *Random Forest* dapat menilai dengan *Area Under the Curve* (AUC) sebesar 0.83 yang mana menunjukkan bahwa kinerja dari klasifikasi bernilai baik dalam menganalisis kredit.

Oleh karena itu, penelitian ini dapat bertujuan untuk dapat menerapkan dengan menggunakan algoritma *Random Forest* dalam mengatasi klasifikasi risiko pinjaman keuangan yang terjadi di Indonesia, seharusnya semua lembaga keuangan dapat membuat sebuah keputusan yang lebih bijak dalam hal pemberian kredit dan mengurangi potensi kerugian akibat terjadinya gagal bayar.

1.2 Rumusan Masalah

1. Bagaimana kinerja algoritma *Random Forest* dalam mengklasifikasikan risiko pinjaman keuangan pada dataset yang tidak seimbang?
2. Bagaimana pengaruh penggunaan teknik *oversampling KMeansSMOTE* terhadap akurasi model dalam memprediksi risiko gagal bayar pinjaman?
3. Fitur-fitur apa saja yang memiliki kontribusi signifikan dalam meningkatkan akurasi prediksi risiko pinjaman keuangan menggunakan algoritma *Random Forest*?
4. Bagaimana *Random Forest* dibandingkan dengan algoritma klasifikasi lainnya dalam hal akurasi dan efisiensi dalam memprediksi risiko pinjaman?

1.3 Batasan Masalah

1. Penelitian ini hanya menggunakan dataset pinjaman keuangan yang terdiri dari 20.000 entri data dengan berbagai atribut terkait pemohon pinjaman, seperti skor kredit, rasio utang terhadap pendapatan, dan status pekerjaan. Sumber dataset diambil dari platform *Kaggle*.
2. Penelitian ini tidak membandingkan *Random Forest* dengan seluruh algoritma *machine learning* yang ada, tetapi akan membandingkan dengan beberapa algoritma klasifikasi lainnya yang relevan seperti *Regresi Logistik* dan *Support Vector Machine (SVM)*.
3. Jenis klasifikasi yang digunakan dalam penelitian ini adalah klasifikasi *biner*, dengan tujuan memprediksi apakah pinjaman disetujui atau tidak berdasarkan risiko gagal bayar.
4. Penelitian ini mengimplementasikan algoritma *Random Forest* untuk klasifikasi risiko pinjaman menggunakan platform *Python* dan alat bantu seperti *Scikit-learn*. Teknik *oversampling KMeansSMOTE* diterapkan untuk menangani ketidakseimbangan kelas dalam dataset.

1.4 Tujuan Penelitian

1. Menganalisis kinerja algoritma *Random Forest* dalam mengklasifikasikan risiko pinjaman keuangan menggunakan data sintesis yang tidak seimbang.
2. Mengidentifikasi pengaruh fitur-fitur yang memiliki kontribusi signifikan terhadap peningkatan akurasi prediksi risiko pinjaman keuangan.
3. Mengevaluasi performa model *Random Forest* dengan menerapkan metode *oversampling* dan *undersampling* menggunakan metrik evaluasi seperti *accuracy*, *precision*, *recall*, *F1-Score*, dan *AUC-ROC*.
4. Bagaimana pengaruh penggunaan teknik *oversampling KMeansSMOTE* terhadap akurasi model dalam memprediksi risiko gagal bayar pinjaman

1.5 Manfaat Penelitian

1. Memberikan kontribusi metodologis dalam pengembangan model prediktif untuk risiko peminjaman dengan memanfaatkan *machine learning*.
2. Menyediakan referensi yang berguna bagi lembaga keuangan dalam menciptakan sistem penilaian risiko yang lebih akurat dan efisien.
3. Menghasilkan pemahaman yang mendalam mengenai penerapan teknik *sampling* pada data yang memiliki ketidakseimbangan kelas.

1.6 Sistematika Penulisan

Bab I Pendahuluan menjelaskan latar belakang masalah yang menjadi dasar penelitian, rumusan masalah yang akan dipecahkan, batasan masalah untuk memperjelas ruang lingkup penelitian, tujuan yang ingin dicapai, manfaat yang diharapkan dari penelitian ini, serta sistematika penulisan yang menggambarkan struktur dokumen skripsi.

Bab II Tinjauan Pustaka membahas dasar-dasar teori yang mendukung penelitian ini. Dalam bab ini, dibahas teori-teori yang berkaitan dengan risiko finansial untuk peminjaman, teknik *machine learning* dengan fokus pada algoritma *Random Forest*, serta metode *sampling* untuk mengatasi ketidakseimbangan data. Selain itu, bab ini juga mengulas penelitian-penelitian sebelumnya yang relevan dengan topik yang diteliti.

Bab III Metodologi Penelitian menjelaskan langkah-langkah dan metode yang digunakan dalam penelitian. Dimulai dari alur penelitian, objek yang diteliti, teknik pengumpulan dan pengolahan data, implementasi algoritma, hingga metode pengujian yang akan diterapkan. Bab ini memberikan gambaran rinci tentang bagaimana penelitian akan dilaksanakan.

Bab IV Hasil dan Pembahasan menyajikan hasil implementasi sistem serta analisis dari pengujian yang telah dilakukan. Di dalam bab ini, akan diuraikan proses pengembangan sistem *approval risk for loan*, hasil evaluasi model *machine learning*, dan pembahasan mengenai performa sistem secara keseluruhan. Hasil

pengujian akan dianalisis untuk menilai seberapa efektif fitur-fitur yang digunakan untuk menentukan pembuatan persetujuan pinjaman.

Bab V Penutup menyajikan kesimpulan dari keseluruhan penelitian serta saran untuk pengembangan sistem di masa mendatang. Kesimpulan ini akan menjawab rumusan masalah yang telah ditetapkan, sementara saran akan memberikan rekomendasi untuk penelitian yang akan datang

