

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi finansial telah mendorong transformasi sistem layanan kredit dari proses bersifat manual menjadi berbasis digital. Digitalisasi ini memungkinkan proses evaluasi kelayakan kredit dilakukan secara lebih cepat dan efisien. Peningkatan penggunaan layanan kredit digital berdampak pada bertambahnya volume pengajuan pinjaman yang harus dianalisis secara akurat dan konsisten. Berdasarkan laporan Otoritas Jasa Keuangan (OJK) tahun 2024, penyaluran kredit di Indonesia mengalami pertumbuhan sekitar 8% dibandingkan tahun sebelumnya [1]. Kondisi tersebut menunjukkan perlunya sistem evaluasi risiko kredit yang mampu mengelola data dalam skala besar secara adaptif.

Seiring meningkatnya jumlah pengajuan kredit, lembaga keuangan dihadapkan pada tantangan dalam mengelola data berdimensi tinggi yang kompleks, seperti profil demografis, riwayat keuangan, dan pola pembayaran. Kesalahan klasifikasi dapat meningkatkan risiko kredit bermasalah (*Non-Performing Loan/NPL*) serta menimbulkan kerugian finansial. Sebaliknya, prediksi yang kurang tepat juga dapat merugikan pemohon yang sebenarnya layak memperoleh kredit. Oleh karena itu, metode *machine learning* semakin dioptimalkan dalam sistem penilaian risiko kredit untuk meningkatkan ketepatan prediksi dan konsistensi pengambilan keputusan berbasis data [2][3].

Berbagai algoritma klasifikasi telah diterapkan dalam evaluasi kelayakan kredit. *Logistic Regression* dikenal memiliki interpretabilitas yang baik dalam menjelaskan hubungan antarvariabel, sedangkan *LightGBM* unggul dalam menangani data berskala besar serta pola non-linier yang kompleks [4]. Namun, penggunaan model tunggal sering kali belum mampu menangkap pola tidak linier dan memiliki antar variabel pada data kredit yang beragam, terutama pada kondisi distribusi kelas yang tidak seimbang. Studi sebelumnya menunjukkan bahwa penanganan data tidak seimbang serta penerapan pendekatan *ensemble learning* dapat meningkatkan stabilitas dan akurasi prediksi risiko kredit [5], [6].

Penelitian ini bertujuan untuk mengembangkan model klasifikasi persetujuan kredit menggunakan metode *Soft Voting Ensemble* yang mengombinasikan algoritma *Logistic Regression* dan *LightGBM*. Penggabungan kedua algoritma tersebut dilakukan untuk memanfaatkan keunggulan masing-masing dalam memproses karakteristik data kredit yang kompleks dan tidak seimbang. *Logistic Regression* digunakan untuk menangkap hubungan linier antar variabel secara interpretatif, sedangkan *LightGBM* diterapkan untuk mengidentifikasi pola non-linier serta interaksi fitur dalam skala data yang besar. Melalui pendekatan *ensemble* ini, diharapkan model yang dihasilkan mampu meningkatkan akurasi, stabilitas, serta kemampuan generalisasi dalam proses evaluasi kelayakan pemberian kredit.

1.2 Rumusan Masalah

Dengan mengacu pada bagian latar belakang di Subbab 1.1, masalah yang menjadi fokus penelitian ini dirancang sebagai berikut:

1. Bagaimana membangun model klasifikasi persetujuan kredit menggunakan algoritma *Logistic Regression* dan *LightGBM*?
2. Bagaimana menerapkan metode *Soft Voting Ensemble* untuk menggabungkan kedua model dalam meningkatkan hasil klasifikasi persetujuan kredit?
3. Seberapa baik performa model *Soft Voting Ensemble* dalam mengklasifikasikan persetujuan kredit berdasarkan metrik evaluasi *accuracy*, *precision*, *recall*, dan *F1-score*?

1.3 Batasan Masalah

Untuk menyesuaikan pembahasan dengan rumusan masalah di Subbab 1.2, studi ini menetapkan batasan masalah sebagai berikut:

1. Penelitian ini berfokus pada pemodelan klasifikasi persetujuan kredit menggunakan dataset *Financial Risk for Loan Approval* yang bersumber dari Kaggle (Loan.csv), sehingga temuan yang diperoleh

hanya menggambarkan pola dan karakteristik yang terdapat dalam dataset tersebut.

2. Penelitian dibatasi pada pembangunan model klasifikasi menggunakan *Logistic Regression* dan *LightGBM* serta penggabungannya melalui metode *Soft Voting Ensemble* tanpa membandingkan dengan algoritma atau metode ensemble lainnya.
3. Evaluasi performa model hanya menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk menilai efektivitas klasifikasi persetujuan kredit, tanpa membahas implementasi sistem nyata atau matriks tambahan.

1.4 Tujuan Penelitian

Dengan mengacu pada rumusan dan batasan masalah yang telah dijelaskan, fokus penelitian ini dirancang sebagai berikut:

1. Membangun model klasifikasi persetujuan kredit menggunakan algoritma *Logistic Regression* dan *LightGBM*.
2. Menerapkan metode *Soft Voting Ensemble* untuk menggabungkan kedua model guna meningkatkan hasil klasifikasi persetujuan kredit.
3. Menganalisis performa model *Soft Voting Ensemble* dalam klasifikasi persetujuan kredit berdasarkan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

1.5 Manfaat Penelitian

Manfaat yang didapatkan dari penelitian ini adalah :

1. Manfaat Teoritis

Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan ilmu di bidang *data mining* dan *machine learning*, khususnya pada penerapan metode *Soft Voting Ensemble* untuk klasifikasi persetujuan kredit. Hasil penelitian ini dapat menjadi referensi akademik terkait efektivitas penggabungan algoritma

Logistic Regression dan *LightGBM* dalam meningkatkan performa model klasifikasi pada data keuangan dengan pola yang kompleks.

2. Manfaat Praktis

Penelitian ini diharapkan dapat membantu lembaga keuangan atau penyedia layanan kredit dalam meningkatkan kualitas proses evaluasi kelayakan kredit berbasis data. Model yang dihasilkan dapat digunakan sebagai pendukung keputusan (*decision support system*) untuk mengurangi risiko kesalahan dalam pemberian kredit, sehingga dapat menurunkan potensi kredit bermasalah (*Non-Performing Loan* atau *NPL*) serta meningkatkan efisien proses analisis kredit.

1.6 Sistematika Penulisan

Sistematika penulisan pada penelitian ini disusun untuk memberikan gambaran menyeluruh mengenai alur pembahasan dan struktur isi laporan. Penyusunan laporan dibagi ke dalam lima bab utama, dimana setiap bab menguraikan tahapan serta komponen penting dalam pelaksanaan penelitian. Adapun susunan sistematika penulisan dalam penelitian ini dijelaskan sebagai berikut:

BAB I PENDAHULUAN

Bab ini memaparkan Latar belakang pentingnya sistem prediksi persetujuan kredit, rumusan dan batasan masalah, tujuan serta manfaat penelitian, dan sistematika penulisan. Selain itu, dijelaskan urgensi penerapan *machine learning* untuk meminimalkan risiko kesalahan dalam pengambilan keputusan kredit.

BAB II TINJAUAN PUSTAKA

Bab ini berisi landasan teori yang mendukung penelitian, meliputi *konsep machine learning*, klasifikasi, penanganan data tidak seimbang, *feature selection*, serta *ensemble learning*. Uraian juga mencakup teori *Logistic Regression* dan *LightGBM*, *Soft Voting Ensemble*, penelitian terdahulu, dan kerangka pemikiran penelitian.

BAB III METODE PENELITIAN

Bab ini menjelaskan tahapan penelitian mulai dari deskripsi dataset persetujuan kredit, *preprocessing data*, *feature selection* menggunakan *SelectKBest*, pembangunan model *baseline*, penanganan imbalance dengan *SMOTE*, *hyperparameter tuning*, penerapan *Soft Voting Ensemble*, hingga evaluasi performa menggunakan metrik *accuracy*, *precision*, *recall* dan *F1-score*.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil eksperimen dari model *baseline* dan model *ensemble*, disertai analisis perbandingan performa sebelum dan sesudah penerapan *SMOTE*, *tuning* dan *Soft Voting*. Pembahasan difokuskan pada kemampuan model dalam mengklasifikasi kelayakan kredit secara akurat.

BAB V PENUTUP

Bab ini berisi kesimpulan yang menjawab rumusan masalah berdasarkan hasil penelitian serta saran untuk pengembangan model dan peluang penerapan sistem prediksi persetujuan kredit pada institusi keuangan di masa mendatang.