

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi yang pesat telah mendorong transformasi digital di berbagai bidang, termasuk sektor keuangan. Saat ini, lembaga keuangan semakin bergantung pada data untuk mendukung proses pengambilan keputusan, terutama dalam pemberian pinjaman atau kredit[1]. Risiko kredit menjadi salah satu aspek penting yang harus diperhatikan, karena kesalahan dalam menilai kemampuan nasabah dapat menyebabkan terjadinya kredit macet, kegagalan pembayaran pinjaman, suku bunga tinggi dan kerugian finansial yang signifikan. Oleh sebab itu, diperlukan metode yang mampu klasifikasi risiko kredit nasabah secara lebih akurat dan efisien[2].

Dalam konteks ini, diperlukan suatu pendekatan yang mampu mengklasifikasikan risiko kredit nasabah secara sistematis dan terukur. Pendekatan klasifikasi dalam *data mining* menjadi salah satu metode yang efektif dalam memprediksi risiko kredit, dengan cara mengelompokkan nasabah berdasarkan karakteristik seperti pendapatan, usia, lama masa bekerja, kepemilikan rumah, jumlah pinjaman, nilai pinjaman, tujuan pinjaman, dan persentase pendapatan pinjaman. Berbagai algoritma klasifikasi dapat digunakan untuk tujuan tersebut, di antaranya *Support Vector Machine (SVM)*, *Random Forest*, dan *XGBoost*, yang masing-masing memiliki keunggulan dalam hal interpretasi, kemampuan generalisasi, serta akurasi hasil prediksi [3][4].

Peneliti dalam kegiatan ini menggunakan dataset yang diambil dari situs *Kaggle*, yang berisi data nasabah dengan berbagai atribut yang berpengaruh terhadap status pinjaman atau *loan_status*. Penelitian ini bertujuan untuk menganalisis kinerja tiga algoritma klasifikasi yaitu *SVM*, *Random Forest*, dan *XGBoost* dalam memprediksi risiko kredit nasabah. Hasil penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan sistem pendukung keputusan di bidang keuangan, khususnya dalam meningkatkan ketepatan evaluasi

risiko kredit secara data-driven.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka dapat dirumuskan masalah penelitian ini sebagai berikut :

1. Apa saja variabel-variabel nasabah seperti pendapatan, usia, lama masa bekerja, kepemilikan rumah, jumlah pinjaman, Tujuan Pinjaman, nilai pinjaman, dan persentase pendapatan pinjaman yang paling dominan dan berpengaruh dalam memprediksi resiko kredit gagal bayar?
2. Bagaimana perbandingan kinerja dari ketiga algoritma klasifikasi tersebut berdasarkan metrik evaluasi seperti *Accuracy*, *Precision*, *Recall*, dan *F1-Score* dalam mengklasifikasikan nasabah ke dalam kategori risiko kredit?
3. Algoritma manakah di antara SVM, *Random Forest*, dan *XGBoost* yang memberikan performa terbaik dalam klasifikasi risiko kredit nasabah berdasarkan hasil evaluasi yang telah dilakukan?

1.3 Batasan Masalah

Dari rumusan masalah yang ditetapkan dalam penelitian ini, maka peneliti membuat batasan masalah yang diteliti sebagai berikut:

1. Penelitian ini hanya berfokus pada analisis kinerja tiga algoritma klasifikasi, yaitu SVM, *Random Forest*, dan *XGBoost*, tanpa membahas algoritma klasifikasi lain di luar ketiga metode tersebut.
2. Variabel yang digunakan sebagai fitur (variabel bebas) dibatasi pada atribut-atribut yang tersedia dalam dataset, seperti usia (*person_age*), pendapatan (*person_income*), lama masa bekerja (*person_emp_length*), kepemilikan rumah (*person_home_ownership*), Jumlah pinjaman (*loan_amnt*), Nilai Pinjaman (*loan_grade*), Tujuan Pinjaman (*loan_intent*), , serta persentase pendapatan pinjaman (*loan_percent_income*).
3. Evaluasi performa model dibatasi pada metrik *accuracy*, *precision*, *recall*, dan *F1-score* serta *feature importance*, tanpa membahas aspek lain seperti kompleksitas waktu komputasi atau efisiensi memori secara mendalam.

4. Dataset yang digunakan dalam penelitian ini diambil dari situs Kaggle, dengan jumlah 32.581 baris data dan 12 atribut tetapi hanya mengambil 9 atribut termasuk variabel target. Variabel target pada dataset ini adalah *loan_status* yang merepresentasikan memprediksi risiko kredit nasabah.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang ditetapkan, maka tujuan penelitian yang ingin dicapai sebagai berikut:

1. Menganalisis dan mengidentifikasi variabel-variabel nasabah (pendapatan, usia, lama masa bekerja, kepemilikan rumah, jumlah pinjaman, tujuan pinjaman, nilai pinjaman, dan persentase pendapatan pinjaman) yang paling dominan atau berpengaruh dalam klasifikasi risiko kredit melalui fitur *feature importance* pada model yang dibangun.
2. Mengevaluasi hasil klasifikasi dengan menggunakan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score* untuk menilai performa masing-masing algoritma dalam mengelompokkan nasabah ke dalam kategori risiko kredit.
3. Menentukan algoritma dengan kinerja terbaik di antara SVM, *Random Forest*, dan *XGBoost* dalam memprediksi risiko kredit nasabah berdasarkan hasil analisis perbandingan.

1.5 Manfaat Penelitian

Melalui penelitian ini, manfaat yang dapat diambil adalah:

1. Menambah wawasan dan referensi ilmiah dalam bidang data mining, khususnya terkait penerapan algoritma klasifikasi seperti SVM, *Random Forest*, dan *XGBoost* pada permasalahan risiko kredit.
2. Memberikan kontribusi dalam pengembangan kajian akademik mengenai analisis perbandingan kinerja algoritma klasifikasi, sehingga dapat menjadi dasar bagi penelitian selanjutnya yang ingin melakukan eksplorasi metode serupa dengan dataset berbeda.
3. Mendukung peningkatan pemahaman mengenai pentingnya pemilihan

algoritma yang tepat berdasarkan karakteristik data untuk memperoleh hasil prediksi yang lebih akurat dan efisien.

1.6 Sistematika Penulisan

Sistematika penulisan ini dirancang untuk memberikan gambaran uraian tentang isi setiap bab dalam penelitian ini. Penelitian ini terdiri dari lima bab, yaitu:

BAB I PENDAHULUAN

Bab ini berisi Latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan yang menjadi panduan dalam pelaksanaan penelitian.

BAB II TINJAUAN PUSTAKA

Bab ini membahas secara mendalam penelitian-penelitian sebelumnya atau terkait dan teori-teori yang menjadi dasar dalam penelitian yang akan digunakan sebagai landasan dalam analisis dan pembahasan.

BAB III METODE PENELITIAN

Bab ini menguraikan metode yang digunakan untuk menyelesaikan masalah penelitian, alur penelitian dan solusi yang ditawarkan termasuk langkah-langkah implementasi, penerapan metodologi yang telah dipilih.

BAB IV HASIL DAN PEMBAHASAN

Bab ini merupakan tahapan hasil pengujian yang diperoleh dari penelitian serta membahas evaluasi hasil tersebut berdasarkan analisis yang dilakukan.

BAB V PENUTUP

Bab ini menyajikan kesimpulan akhir dari penelitian serta saran-saran yang diberikan untuk pengembangan atau penelitian lanjutan di masa mendatang