BAB I
PENDAHULUAN

Bab ini berisi latar belakang penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, ruang lingkup penelitian, serta sistematika penulisan skripsi.

1.1. Latar Belakang

Dalam dunia keuangan, pemberian kredit merupakan salah satu layanan utama yang memungkinkan individu dan bisnis mendapatkan akses pendanaan. Agar proses ini berjalan dengan optimal, lembaga keuangan menggunakan *credit scoring*, yaitu sistem penilaian untuk menentukan apakah seseorang layak menerima pinjaman (Zhang dkk., 2018). *Credit scoring* menjadi sangat penting karena membantu bank dan perusahaan keuangan mengurangi risiko gagal bayar serta meningkatkan efisiensi dalam pengambilan keputusan kredit.

Seiring perkembangan teknologi, metode tradisional dalam menilai kredit, seperti analisis manual oleh analis kredit, dianggap kurang efisien dan memiliki keterbatasan dalam menangani jumlah data yang besar (Madaan dkk., 2021). Oleh karena itu, banyak perusahaan mulai beralih ke *Machine Learning*, yang memungkinkan pemrosesan data kredit secara lebih cepat dan akurat (Dai dkk., 2021). Dalam penelitian (Dai dkk., 2021), *Support Vector Machine* (SVM) ditemukan memiliki performa lebih baik dibandingkan *Random Forest* dalam beberapa kasus, tetapi *Random Forest* tetap menjadi pilihan karena kemampuannya dalam menangani dataset yang kompleks dengan lebih baik.

Algoritma *Random Forest* telah menjadi salah satu metode paling populer dalam klasifikasi *credit score*. Dibandingkan dengan metode tradisional seperti *Decision Tree* atau *Logistic Regression*, *Random Forest* memiliki akurasi lebih tinggi, mampu menangani beragam fitur, dan lebih tahan terhadap *overfitting* (Breiman, 2001). Penelitian oleh Madaan dkk., (2021) menunjukkan bahwa *Random Forest* mencapai akurasi 79,8% dalam prediksi risiko kredit, lebih tinggi dibandingkan *Decision Tree* yang hanya mencapai 73,1%.

Beberapa studi lain juga membuktikan keunggulan *Random Forest* dalam klasifikasi kredit. Zhang dkk. (2018) mengusulkan sebuah model credit score baru bernama NCSM (*Novel Credit Scoring Model*) yang berbasis pada optimasi algoritma *Random Forest*. Hasil

penelitian tersebut menunjukkan bahwa model yang di usulkan terbukti lebih akurat dibandingkan dengan model-model klasifikasi popular lainnya. Selain itu, penelitian oleh Dai dkk. (2021) menunjukkan bahwa *Random Forest* memiliki akurasi hingga 95%, lebih baik daripada metode *Support Vector Machine* (SVM) yang hanya mencapai 86% dalam dataset tertentu.

Meskipun banyak penelitian telah membuktikan keunggulan *Random Forest* dalam klasifikasi *credit score*, studi seperti yang dilakukan oleh Zhang dkk., (2018) menunjukkan bahwa performa model NCSM yang dihasilkan sangat bergantung pada pengaturan *hyperparameter*-nya. *Hyperparameter* dapat dipahami sebagai konfigurasi untuk mengontrol perilaku algoritma yang nilainya ditetapkan sebelum proses pelatihan (Kim, 2016). Pentingnya pengaturan ini telah ditekankan oleh Breiman (2001) dalam artikelnya yang memperkenalkan *Random Forest*, di mana ia membahas parameter kunci seperti jumlah pohon (*trees*) yang perlu dibangun. Mengingat hal tersebut, penelitian ini berfokus pada pencarian konfigurasi *hyperparameter* terbaik meliputi *n_estimators* (jumlah pohon), *max_depth* (kedalaman maksimum), dan *max_features* (jumlah fitur pada setiap pemisahan) untuk meningkatkan akurasi model *Random Forest* dalam klasifikasi *credit score*.

Berdasarkan pustaka *Scikit-Learn*, model *Random Forest* adalah hasil dari pelatihan algoritma *Random Forest* pada data yang dimasukkan. Setelah model ini dilatih menggunakan data latih, ia akan menyimpan beberapa pohon keputusan yang dibuat secara acak. Setiap pohon tersebut akan memberikan hasil prediksi. Untuk menentukan hasil akhir, model mengambil suara terbanyak dari semua pohon (voting mayoritas) untuk memutuskan kelas mana yang paling sesuai. Dengan cara ini, model dapat memberikan hasil klasifikasi yang lebih akurat dan stabil.

Dengan penelitian ini, diharapkan dapat diperoleh pemahaman yang lebih mendalam mengenai optimasi model *Random Forest* dalam klasifikasi *credit score*, sehingga dapat menentukan konfigurasi *hyperparameter* yang paling optimal untuk meningkatkan akurasi model *Random Forest*.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka diperoleh rumusan masalah yaitu bagaimana memperoleh model *Random Forest* yang optimal dalam mengklasifikasikan *credit score*.

1.3. Tujuan Penelitian

Tujuan dari penelitian ini adalah menentukan kombinasi *hyperparameter* terbaik untuk meningkatkan akurasi model *Random Forest* dalam klasifikasi *credit score*.

1.4. Manfaat Penelitian

Manfaat dari penelitian ini adalah memberikan informasi tentang konfigurasi terbaik *Random Forest* yang dapat digunakan oleh perusahaan keuangan untuk meningkatkan akurasi klasifikasi *credit score* dan dapat menjadi referensi bagi penelitian selanjutnya terkait optimasi *hyperparameter* dalam algoritma *Random Forest* untuk *credit score*.

1.5. Ruang Lingkup Penelitian

Ruang lingkup dalam skripsi ini bertujuan untuk memberikan batasan yang jelas agar penelitian tetap fokus pada permasalahan yang dikaji dan sesuai dengan tujuan yang ingin dicapai. Sehingga pembahasan tidak akan melebar ke luar konteks yang telah ditentukan. Adapun ruang lingkup dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini berfokus pada pencarian model *Random Forest* yang paling optimal dalam mengklasifikasikan *credit score* melalui *tuning hyperparameter* serta analisis performa model.
2. Data yang digunakan dalam penelitian ini merupakan data perusahaan keuangan yang diperbarui pada tahun 2022. Pengumpulan data dilakukan dengan menggunakan dataset publik berjudul "*Credit Score Classification*" yang diunduh dari situs *Kaggle.com* yang terdiri dari 100.000 data nasabah.
3. Model *Random Forest* dikembangkan menggunakan bahasa pemrograman *Python* dengan memanfaatkan *library scikit-learn*, serta dilatih dan diuji dalam lingkungan *Jupyter Notebook*.

4. Pra-pemrosesan data dilakukan sebelum pelatihan model *Random Forest*, termasuk penanganan nilai yang hilang dan normalisasi data jika diperlukan.
5. Optimasi model *Random Forest* dilakukan dengan metode *tuning hyperparameter*, seperti *GridSearchCV* atau *RandomizedSearchCV*, untuk mendapatkan kombinasi parameter yang memberikan performa terbaik
6. Evaluasi model *Random Forest* dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F-score* untuk menilai performa klasifikasi.

1.6. Sistematika Penulisan

Sistematika penulisan memberikan gambaran laporan dari skripsi ini secara jelas. Berikut adalah sistematika penulisan dari skripsi ini:

BAB I PENDAHULUAN

Bab ini berisi latar belakang penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, ruang lingkup penelitian, serta sistematika penulisan skripsi.

BAB II TINJAUAN PUSTAKA

Bab ini berisi teori-teori yang mendukung penelitian, termasuk *credit score*, *data mining*, klasifikasi, Algoritma *Random Forest*, *Feature Importance*, *confusion matrix*, *Tools* dan *Library*, *Jupyter Notebook*, dan *state of the art*.

BAB III METODOLOGI PENELITIAN

Bab ini menguraikan metodologi penelitian yang digunakan, mulai dari pengumpulan data, pra-pemrosesan data, perancangan dan pelatihan model *Random Forest*, serta teknik evaluasi kinerja model *Random Forest* menggunakan metrik akurasi, presisi, *recall*, dan *F-score*.

BAB IV HASIL DAN ANALISIS

Bab ini menyajikan implementasi dan analisis model *Random Forest* dengan fokus pada pencarian konfigurasi *hyperparameter* terbaik untuk klasifikasi *credit score*. Pembahasan di dalamnya meliputi skenario pengujian, hasil evaluasi performa dari setiap

konfigurasi, serta analisis *feature importance* untuk menentukan fitur yang paling berpengaruh.

BAB V PENUTUP

Bab ini berisi kesimpulan dari penelitian yang telah dilakukan dan saran untuk penelitian selanjutnya.