

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil dari penelitian yang telah peneliti lakukan dengan membandingkan 2 kinerja algoritma, dapat disimpulkan bahwa:

1. Perbandingan Performa XGBoost dan C4.5

Algoritma C4.5 menunjukkan performa yang lebih unggul dibandingkan XGBoost dalam memprediksi dan mengklasifikasikan risiko kredit berdasarkan data pemohon pinjaman sebelum menggunakan hyperparameter tuning GridSearchCV.

- Pada rasio 90:10, algoritma C4.5 mencapai akurasi tertinggi sebesar 96.20% dengan nilai ROC (AUC) 0.9712, sementara XGBoost hanya mencapai akurasi 93.85% dengan nilai ROC (AUC) 0.9655.
- Pada rasio 80:20, algoritma C4.5 tetap unggul dengan akurasi 96.20% dibandingkan XGBoost 94.37%.
- Untuk rasio 70:30 dan 60:40, C4.5 konsisten menunjukkan akurasi tinggi sebesar 96.19%, sementara XGBoost masing-masing mencapai 94.30% dan 94.23%.

Selain itu, algoritma C4.5 memiliki nilai Recall tertinggi sebesar 99.98%, yang menunjukkan kemampuan lebih baik dalam mengidentifikasi pemohon pinjaman berisiko tinggi.

2. Efektivitas Setelah Menggunakan Optimisasi Performa Model GridSearchCV

Dengan penerapan GridSearchCV 10-fold cross-validation, performa algoritma XGBoost meningkat secara signifikan dan hampir sama dengan performa algoritma C4.5. Namun, algoritma C4.5 tetap unggul di semua rasio data.

- Pada rasio 90:10, algoritma C4.5 mencapai akurasi 96.20% dengan nilai ROC (AUC) 0.9712, sementara XGBoost mencapai akurasi 96.20% dengan nilai ROC (AUC) 0.9713.
- Pada rasio 80:20, 70:30, dan 60:40, algoritma C4.5 tetap unggul dengan nilai akurasi konsisten 96.20% dan nilai ROC (AUC) maksimum 0.9730 dibandingkan XGBoost yang nilainya sedikit lebih rendah.

Berdasarkan hasil evaluasi, algoritma C4.5 dapat direkomendasikan sebagai algoritma yang unggul dalam hal akurasi, efektivitas, dan efisiensi waktu untuk mengklasifikasikan risiko kredit pada dataset yang digunakan untuk memprediksi resiko kredit yang baru. Sementara itu, algoritma XGBoost tetap menjadi alternatif kompetitif, terutama setelah optimasi parameter menggunakan GridSearchCV, yang mampu menghasilkan performa yang hampir sama dengan algoritma C4.5. Diantara semua rasio yang di uji yang mendapatkan hasil evaluasi tertinggi terbanyak dan unggul adalah rasio 80:20. Kesimpulan ini menjawab rumusan masalah penelitian dengan menegaskan perbandingan performa kedua algoritma pada berbagai rasio data serta keunggulan algoritma dalam memproses data besar dengan efisien.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat menjadi pertimbangan untuk penelitian selanjutnya maupun untuk implementasi praktis:

- **Eksplorasi Algoritma Lain:** Penelitian selanjutnya disarankan untuk mengeksplorasi algoritma lain dan mungkin juga melakukan perbandingan antar algoritma, seperti Random Forest, SVM, KNN, atau Neural Networks, untuk menemukan algoritma yang lebih optimal pada data berdimensi tinggi.

- **Optimasi Fitur:** Teknik feature engineering seperti PCA atau RFE dapat digunakan untuk mengurangi dimensi data dan meningkatkan efisiensi waktu eksekusi, terutama untuk algoritma yang memerlukan komputasi tinggi.
- **Uji Dataset Beragam:** Pengujian pada dataset dengan karakteristik berbeda, seperti data umum atau baru serta dataset nya bisa dari negara indonesia, diperlukan untuk menguji generalisasi algoritma pada dataset yang berbeda.
- **Tuning Parameter Lanjutan:** Teknik seperti Random Search atau Bayesian Optimization dapat digunakan untuk meningkatkan efisiensi dalam menemukan kombinasi parameter terbaik.
- **Implementasi Sistem Nyata:** Hasil penelitian dapat diimplementasikan dalam sistem berbasis lain seperti aplikasi dekstop atau mobile untuk analisis risiko kredit lebih lanjut secara real-time dan dinamis.

