

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan penelitian yang dilakukan adapun kesimpulan dari perbandingan kinerja model Logistic Regression dan XGBoost dalam memprediksi skor kredit (credit scoring) pada data debitur kredit mobil PT Graha Mazindo Mandiri:

1. Kedua algoritma machine learning, yaitu Logistic Regression dan XGBoost, berhasil diterapkan untuk memprediksi skor kredit dengan menggunakan data sekunder dari PT Graha Mazindo Mandiri. Adapun proses penerapan model diawali dengan melakukan pengumpulan data debitur yang kemudian diolah melalui tahap preprocessing data. Pada tahap ini, data kategorikal diubah menjadi format numerik menggunakan Label Encoding untuk memastikan konsistensi seluruh fitur, sementara distribusi data dinormalisasi melalui teknik scaling dengan *StandardScaler* dan penyeimbangan kelas menggunakan teknik SMOTE untuk mengatasi ketidakseimbangan data dengan rasio kelas positif (lancar) terhadap negatif (macet) sebesar 2:1. Setelah dilakukan preprocessing selanjutnya adalah tahapan pemodelan dimana pada model Logistic Regression diterapkan dengan parameter regularisasi ($C=0.001$) dan bobot kelas 'balanced' untuk mencegah overfitting, sedangkan XGBoost menggunakan parameter seperti $n_estimators=100$, $learning_rate=0.1$, dan $max_depth=4$ untuk mengoptimalkan performa. Kedua algoritma dilatih pada train set sebanyak 1.604 entri dan dievaluasi pada test set sebanyak 402 entri, dan menghasilkan prediksi yang mendukung pengambilan keputusan kredit.
2. Hasil yang ditemukan dalam penerapan *credit scoring* yaitu dimana hasil analisis menunjukkan bahwa XGBoost menghasilkan accuracy 98%, recall 98% dan ROC-AUC score 0,99 sementara Logistic Regression mencapai accuracy 90%, recall 90% dan ROC-AUC score 0,96. Adapun fitur-fitur

utama yang memengaruhi kelayakan kredit meliputi *DATA_BLACKLIST*, *OCCUPATION_TYPE*, *CNT_PAYMENT*, *AMT_CREDIT*, *AMT_ANNUITY*. Penanganan ketidakseimbangan data menggunakan SMOTE meningkatkan representasi kelas minoritas, sehingga model lebih sensitif terhadap debitur berisiko macet. Solusi yang dihasilkan mencakup penggunaan model XGBoost untuk mendeteksi risiko kredit secara akurat, yang dapat mengurangi risiko *Non-Performing Loan* (NPL) dengan mengidentifikasi debitur berisiko tinggi.

3. Penentuan metode terbaik dilakukan berdasarkan evaluasi menggunakan metrik recall dan ROC-AUC Score, dimana ROC-AUC mengukur kemampuan model untuk membedakan dua kategori, yang membandingkan True Positive Rate dengan False Positive Rate. Sedangkan recall mengukur kemampuan model dalam mendeteksi semua potensi default (gagal bayar). ROC-AUC lebih robust pada dataset tidak seimbang karena mengevaluasi performa model di berbagai ambang batas klasifikasi. XGBoost terbukti sebagai metode terbaik untuk memprediksi skor kredit pada dataset ini dengan menghasilkan ROC-AUC (0,99) dan recall (0,98) mengungguli Logistic Regression yang memperoleh ROC-AUC (0,96) dan recall (0,90). Keunggulan XGBoost terletak pada kemampuannya menangkap pola non-linier dan interaksi kompleks antar fitur, serta efektivitasnya dalam menangani ketidakseimbangan data setelah penerapan SMOTE. Logistic Regression, meskipun memberikan performa yang baik dengan cara kerja yang sederhana dan mudah dipahami, model ini terbatas karena menganggap hubungan antar variabel bersifat linear yang menyebabkan performa lebih rendah dibandingkan XGBoost. Oleh karena itu, XGBoost lebih direkomendasikan untuk pengaplikasian credit scoring, terutama untuk mendukung pengambilan keputusan kredit yang lebih akurat dan efisien.

5.2. Saran

Berdasarkan hasil penelitian yang telah dilakukan, berikut adalah saran untuk pengembangan lebih lanjut

1. Untuk meningkatkan hasil penelitian dan penerapan model credit scoring di PT Graha Mazindo Mandiri, disarankan untuk mengoptimalkan hyperparameter model XGBoost dengan teknik seperti grid search atau random search guna mencapai performa yang lebih robust.
2. Penelitian lanjutan dapat dilakukan dengan menguji model pada dataset dari jenis kredit lain, seperti kredit rumah atau mikro, untuk memvalidasi generalisasi model.
3. Selain itu, pemantauan berkala terhadap performa model dengan data terbaru sangat penting untuk menjaga akurasi prediksi seiring perubahan pola kredit. Menggabungkan pendekatan ensemble learning dengan algoritma lain seperti Random Forest atau LightGBM juga dapat menjadi langkah inovatif untuk meningkatkan ketepatan dan keandalan sistem credit scoring.

