

BAB I

PENDAHULUAN

1.1. Latar Belakang

Saat ini, sektor perbankan beroperasi dalam lingkungan yang sangat kompetitif, di mana menjaga loyalitas nasabah memiliki tingkat kepentingan yang setara dengan upaya menarik nasabah baru. Berdasarkan analisis data pelanggan serta kajian literatur, diketahui bahwa biaya akuisisi nasabah baru umumnya jauh lebih besar dibandingkan dengan biaya mempertahankan nasabah yang sudah ada. Oleh sebab itu, kemampuan mendeteksi indikasi awal potensi *churn* menjadi aspek strategis yang krusial agar program retensi dapat dijalankan secara lebih efektif dan tepat sasaran.

Perilaku churn dipengaruhi kombinasi banyak faktor: demografi, aktivitas rekening, produk yang dimiliki, hingga status keaktifan nasabah. Pada dataset yang saya gunakan, fitur-fitur seperti umur, jumlah produk, status anggota aktif, dan saldo menunjukkan pola yang relevan terhadap kejadian churn. Pola-pola ini tidak selalu linier atau mudah ditangkap oleh analisis sederhana sehingga pendekatan machine learning lebih cocok untuk menggali hubungan kompleks tersebut.

Algoritma berbasis pohon keputusan seperti Decision Tree, Random Forest, dan XGBoost banyak digunakan karena masing-masing memiliki keunggulan berbeda: Decision Tree mudah dipahami, Random Forest menawarkan kestabilan melalui pendekatan *ensemble*, sedangkan XGBoost mampu melakukan *boosting* untuk memperbaiki kesalahan secara bertahap. Meski demikian, penerapan nyata sering menghadapi tantangan karena kinerja model sangat bergantung pada pemilihan *hyperparameter* serta kualitas tahap pra-pemrosesan data. Tanpa adanya optimasi yang tepat, model berisiko menghasilkan prediksi yang menyesatkan meskipun secara angka tampak memiliki akurasi tinggi.

Grid Search merupakan pendekatan sistematis yang sederhana namun efektif untuk melakukan penyesuaian *hyperparameter* dengan cara menguji berbagai kombinasi nilai yang telah ditentukan sebelumnya. Dalam penerapannya pada sektor perbankan, metode ini berfungsi menemukan konfigurasi optimal yang mampu menyeimbangkan akurasi, *recall* pada kelas churn, serta beban komputasi. Walaupun prosesnya relatif memakan waktu, penggunaan Grid Search terbukti dapat meningkatkan kemampuan model dalam

mengidentifikasi nasabah berisiko, sehingga strategi retensi dapat diarahkan secara lebih tepat dan terfokus.

Dalam praktik penelitian ini, data yang dipakai berasal dari dataset publik berupa 10.000 entri dengan fitur terstruktur (demografi dan perilaku rekening). Ketersediaan data yang relatif bersih (tanpa missing value signifikan) memungkinkan saya fokus pada langkah-langkah penting lain: encoding kategori, normalisasi, analisis korelasi, dan pembagian data untuk skenario 70:30 dan 80:20. Langkah-langkah ini diperlukan agar evaluasi model valid dan hasil optimasi hyperparameter dapat diandalkan.

Hasil pengujian awal menunjukkan bahwa meskipun akurasi keseluruhan beberapa model tampak mirip, perbedaan nyata muncul pada metrik yang lebih sensitif terhadap kelas minoritas seperti PR-AUC dan recall untuk kelas churn. Ini menegaskan bahwa mengevaluasi model hanya dari akurasi tidak cukup; metrik yang menilai kemampuan model mendekripsi churn (kelas minoritas) harus dijadikan prioritas ketika tujuan praktisnya adalah mengurangi kehilangan nasabah.

Selain mengevaluasi performa, aspek interpretabilitas juga penting agar tim bisnis atau manajemen bank dapat memahami faktor yang mendorong churn dan merancang tindakan retensi. Decision Tree menyediakan aturan yang mudah dibaca, sementara Random Forest dan XGBoost dapat dilengkapi analisis feature importance untuk memberi gambaran kontribusi setiap fitur terhadap keputusan model. Kombinasi interpretasi dan performa inilah yang diharapkan menjadi output penelitian.

Berdasarkan alasan-alasan tersebut, penelitian ini mengusulkan penerapan Grid Search untuk mengoptimalkan Decision Tree, Random Forest, dan XGBoost dalam memprediksi customer churn pada bank. Tujuannya praktis: menemukan konfigurasi model yang seimbang antara deteksi churn yang baik, stabilitas generalisasi, dan interpretasi yang mendukung pengambilan keputusan retensi nasabah. Hasil yang diharapkan adalah rekomendasi model dan setting hyperparameter yang bisa langsung digunakan sebagai dasar strategi retensi.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah saya jelaskan dalam skripsi, masalah penelitian ini dirumuskan sebagai berikut:

1. Bagaimana cara mengidentifikasi nasabah yang berisiko churn pada bank menggunakan data demografi dan perilaku rekening yang tersedia.
2. Sejauh mana pemilihan dan optimasi hyperparameter pada model Decision Tree, Random Forest, dan XGBoost dapat meningkatkan kemampuan deteksi churn, khususnya pada kelas minoritas.

3. Model mana (dan dengan konfigurasi hyperparameter apa) yang memberikan trade-off terbaik antara kemampuan mendeteksi churn (recall / PR-AUC), kestabilan umum (ROC-AUC) dan efisiensi waktu komputasi untuk dataset yang digunakan.

1.3. Tujuan Penelitian

Penelitian ini bertujuan untuk memprediksi customer churn pada bank menggunakan perbandingan optimasi algoritma grid search dengan decision tree, random forest, xgboost. Menemukan konfigurasi model paling efektif dengan mempertimbangkan metrik deteksi seperti recall dan PR AUC, kemampuan membedakan kelas melalui ROC AUC, serta efisiensi komputasi, membandingkan performa ketiga model sebelum dan sesudah optimasi untuk menyajikan rekomendasi praktis penerapannya pada sistem deteksi churn di lingkungan perbankan, serta menyediakan insight interpretatif berupa fitur yang paling berkontribusi dan aturan keputusan utama agar hasil model dapat ditindaklanjuti oleh pihak non-teknis di bank.

1.4. Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini meliputi:

1. Manfaat praktis bagi bank: menyediakan model prediksi churn yang telah dioptimalkan sehingga tim retensi dapat menargetkan intervensi secara lebih tepat dan efisien.
2. Manfaat teknis bagi pengembangan ilmu: membandingkan efektivitas Grid Search pada tiga algoritma pohon populer dan mendokumentasikan hubungan antara konfigurasi hyperparameter, metrik performa khusus kelas minoritas, dan waktu komputasi.
3. Manfaat bagi penelitian selanjutnya: menyajikan baseline eksperimen (pra-pemrosesan, grid hyperparameter, metrik evaluasi) yang dapat diperluas dengan teknik optimasi lain (mis. Randomized Search, Bayesian Optimization) atau fitur engineering lebih lanjut.
4. Manfaat edukatif bagi penulis: memperdalam pemahaman praktik implementasi machine learning end-to-end untuk masalah klasifikasi tidak seimbang dan keterkaitannya dengan kebutuhan bisnis nyata.

1.5. Batasan Penelitian

Penelitian ini memiliki Batasan penelitian yaitu sebagai berikut:

1. Penelitian ini dibatasi pada penggunaan Grid Search untuk optimasi hyperparameter tiga model pohon keputusan, yaitu Decision Tree, Random Forest, dan XGBoost, dalam memprediksi customer churn pada bank.

2. Evaluasi performa model dibatasi pada metrik yang relevan, yaitu precision, recall, F1-score, PR-AUC, accuracy, ROC-AUC, serta pengukuran waktu komputasi, tanpa mengevaluasi metrik lain di luar ruang lingkup tersebut.
3. Penelitian hanya mempertimbangkan dua skenario pembagian data (70:30 dan 80:20), serta analisis terbatas pada pemilihan fitur, pra-pemrosesan data, dan konfigurasi Grid Search, tanpa mencakup implementasi sistem operasional di lingkungan bank.