

Nama : Al Fitra Nur Ramadhani

NIM : 202210370311264

Mata Kuliah : Pemodelan dan Simulasi Data B

Laporan Analisis Prediksi Churn Pelanggan Bank

Deskripsi

Laporan ini menganalisis data pelanggan bank untuk memprediksi kemungkinan churn (pelanggan yang keluar) menggunakan pendekatan machine learning. Dataset berisi 10.000 entri dengan 18 kolom, termasuk informasi seperti CreditScore, Age, Geography, Gender, Balance, dan status Exited (0 untuk tidak churn, 1 untuk churn). Tujuan utama adalah membangun model prediktif untuk mengidentifikasi pelanggan yang berisiko churn dan memberikan wawasan untuk strategi retensi.

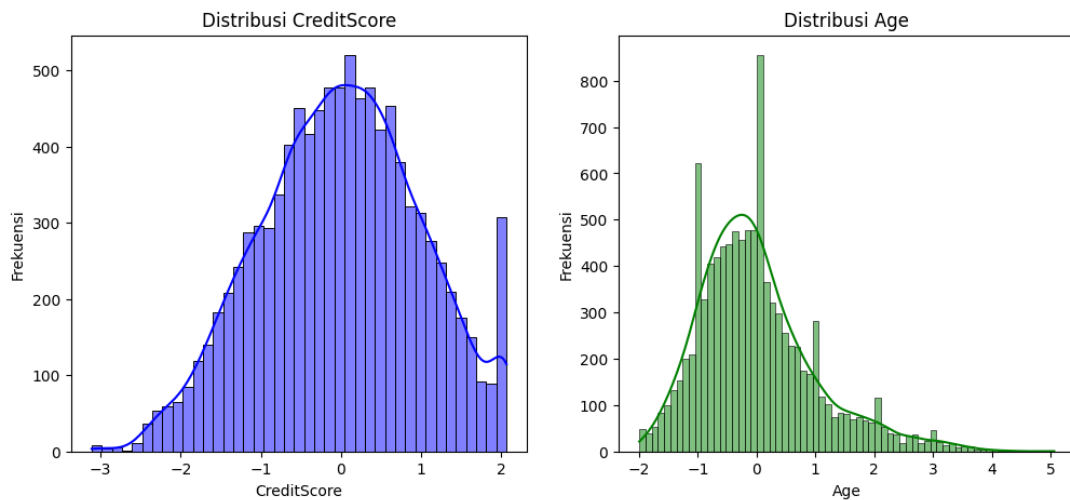
Implementasi Step-by-Step

1. Pengumpulan dan Pemahaman Data Dataset diambil dari file BankCustomerChurnRecords.csv. Informasi awal menunjukkan tidak ada data yang hilang, dengan kolom numerik (CreditScore, Age, dll.) dan kategorikal (Geography, Gender, Card Type). Jumlah nilai unik per kolom dianalisis untuk memahami variasi data.
2. Preprocessing Data Kolom numerik distandardisasi menggunakan StandardScaler untuk menormalkan distribusi. Kolom kategorikal diencode menggunakan LabelEncoder. Proses ini memastikan data siap untuk pemodelan.
3. Analisis Eksplorasi Data (EDA) Distribusi CreditScore dan Age divisualisasikan melalui histogram dengan kurva KDE. Scatter plot dibuat untuk melihat hubungan antara CreditScore dan Age berdasarkan status Exited.
4. Pemodelan dan Evaluasi Model RandomForestClassifier digunakan dengan parameter yang diatur untuk mencegah overfitting (max_depth=4, min_samples_split=15, min_samples_leaf=8). Data dibagi menjadi 80% training dan 20% testing. Teknik SMOTE diterapkan untuk menangani ketidakseimbangan kelas. Model dievaluasi menggunakan cross-validation, akurasi, classification report, dan confusion matrix.

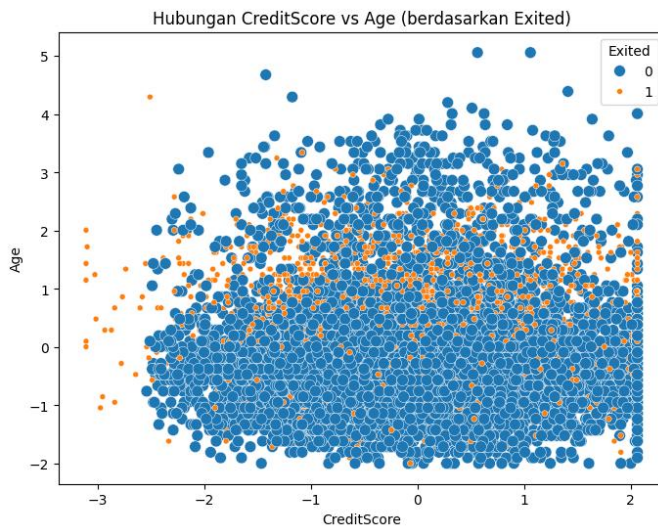
5. Simulasi Prediksi Simulasi dilakukan untuk lima pelanggan dengan karakteristik beragam. Probabilitas churn dan non-churn divisualisasikan dalam bentuk bar chart untuk memberikan gambaran risiko churn per pelanggan.

Hasil Analisis

- Distribusi Data Histogram menunjukkan CreditScore dan Age telah distandardisasi, dengan CreditScore memiliki distribusi mendekati normal (mean sekitar 0), sedangkan Age sedikit miring ke kanan, menunjukkan lebih banyak pelanggan muda.

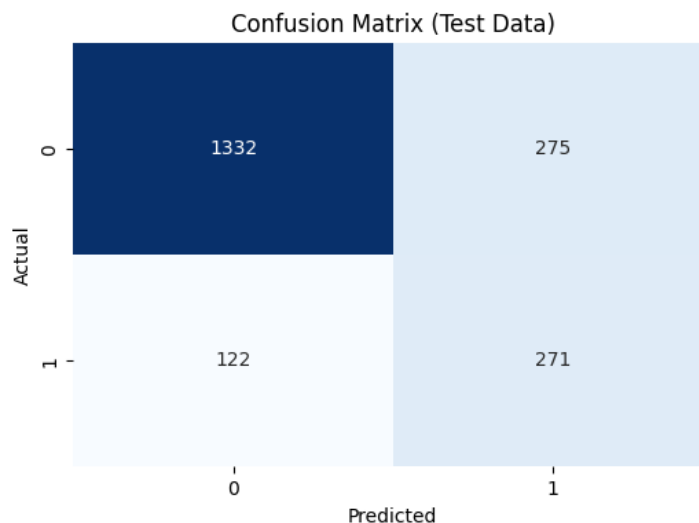


- Hubungan CreditScore dan Age Scatter plot mengindikasikan pelanggan yang churn (Exited=1, oranye) cenderung memiliki CreditScore dan Age yang lebih bervariasi, tetapi tidak ada pola pemisahan yang jelas antara kelompok churn dan non-churn.

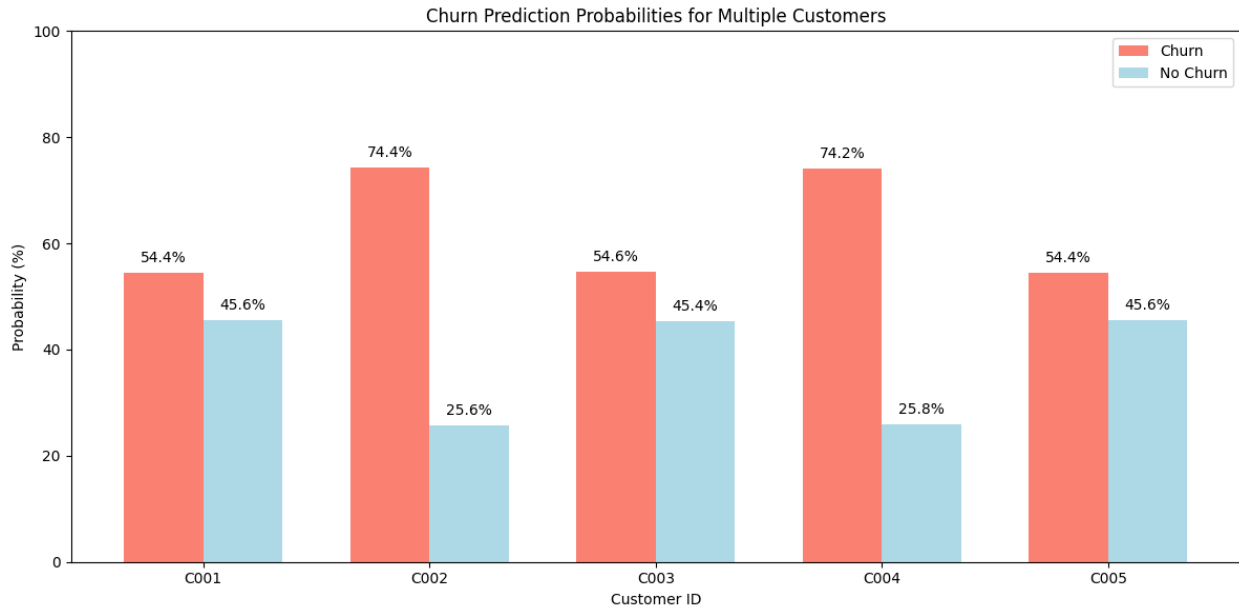


- Performa Model Model mencapai akurasi training 81% dan testing 80%, dengan cross-validation accuracy 0.81 ± 0.01 . Confusion matrix menunjukkan 1332 true negatives, 275 false positives, 122 false negatives, dan 271 true positives. Precision untuk kelas 1 (churn) adalah 0.50, dengan recall 0.69, menunjukkan model cukup baik dalam mendeteksi pelanggan yang churn.

Classification Report (Test Data):				
	precision	recall	f1-score	support
0	0.92	0.83	0.87	1607
1	0.50	0.69	0.58	393
accuracy			0.80	2000
macro avg	0.71	0.76	0.72	2000
weighted avg	0.83	0.80	0.81	2000



- Simulasi Prediksi Hasil simulasi untuk lima pelanggan menunjukkan probabilitas churn yang bervariasi: C001 (54.4% churn), C002 (74.4% churn), C003 (54.6% churn), C004 (74.2% churn), dan C005 (54.4% churn). Pelanggan C002 dan C004 memiliki risiko churn tertinggi.



Insight dan Kesimpulan

- Insight Pelanggan dengan risiko churn tinggi (seperti C002 dan C004) cenderung memiliki karakteristik seperti CreditScore rendah, usia lebih tua, saldo tinggi, dan kepuasan rendah. Faktor seperti ketidakaktifan (IsActiveMember=0) dan jumlah produk yang sedikit juga berkontribusi pada risiko churn.
- Kesimpulan Model RandomForestClassifier memberikan prediksi churn yang cukup akurat (80%) dan dapat digunakan untuk mengidentifikasi pelanggan berisiko. Bank dapat memprioritaskan strategi retensi pada pelanggan dengan probabilitas churn tinggi, seperti meningkatkan kepuasan pelanggan atau menawarkan produk tambahan untuk meningkatkan keterlibatan.