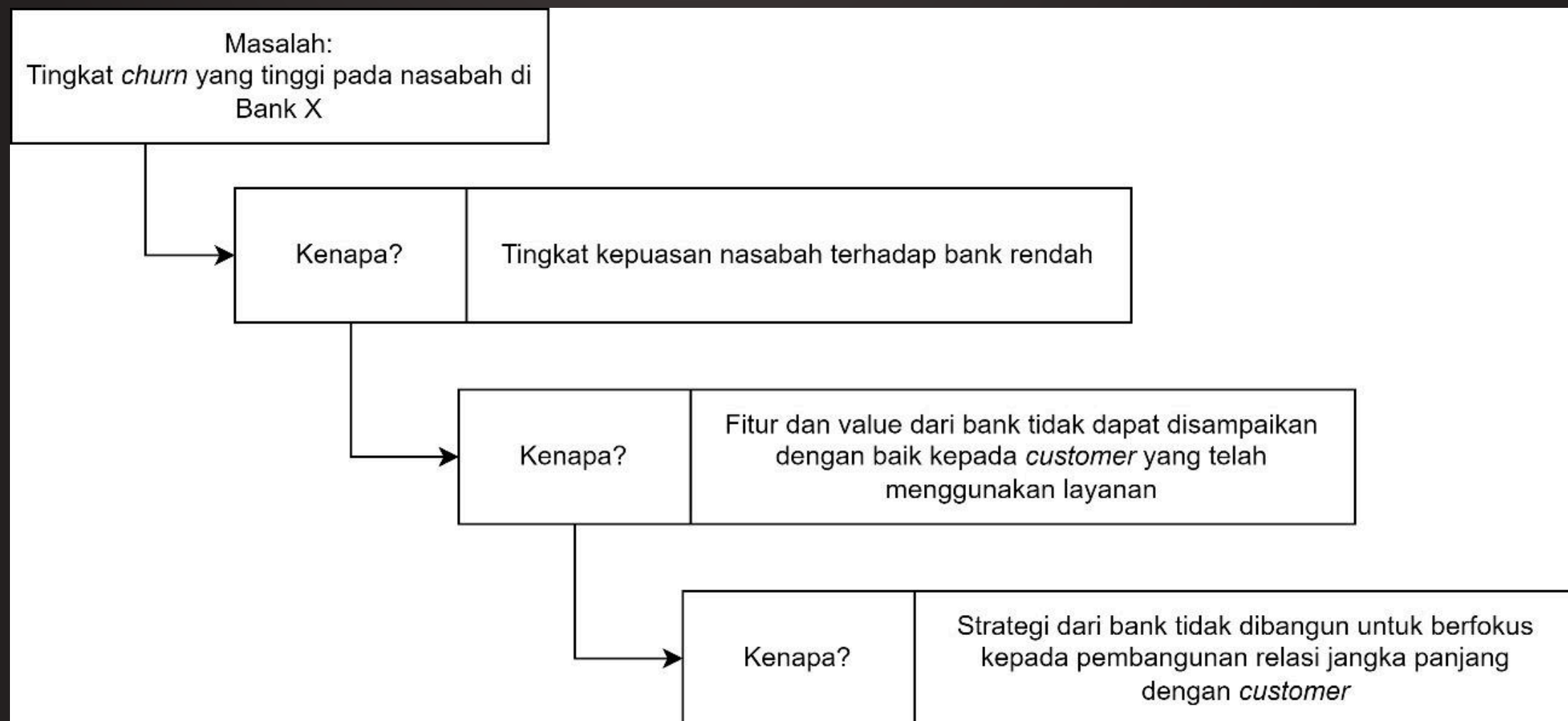

PEMBANGUNAN MODEL PREDIKSI DAN STRATEGI UNTUK MENGELOLA TINGKAT CHURN YANG TINGGI DI BANK X



BACKGROUND

Berdasarkan data yang dikelola European Banking Federation, jumlah Bank di Benua Eropa mengalami penurunan yang drastis sebesar 33% dari tahun 2009 ke 2020. Salah satu faktor yang menyebabkannya adalah tingkat persaingan yang tinggi antar bank. Oleh karena itu, pemilik Bank harus berinovasi untuk dapat mempertahankan nasabah lama dan menarik nasabah baru. Maka, akan dibangun model prediksi untuk mengetahui perilaku nasabah di masa depan dan merumuskan strategi bank

IDENTIFIKASI MASALAH



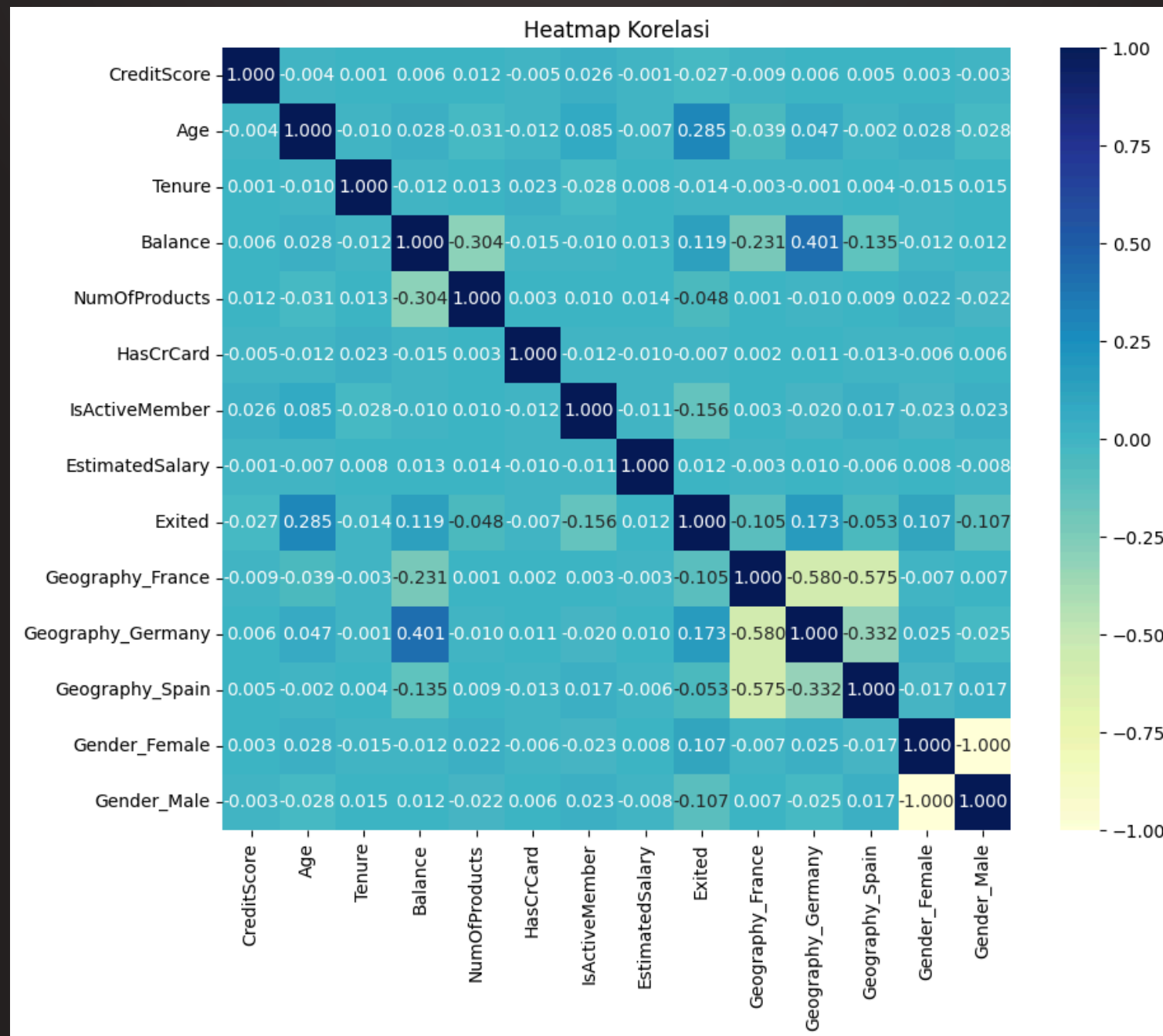
Dengan tools 5 whys, diperoleh akar penyebab dari tingkat churn yang tinggi di Bank X adalah **strategi dari bank tidak dibangun untuk berfokus kepada pembangunan relasi jangka panjang dengan customer**

PENGOLAHAN DATA

Nama Kolom	Deskripsi
CustomerId	ID unik dari setiap nasabah bank
Surname	Nama belakang atau nama keluarga dari nasabah
CreditScore	Nilai yang merepresentasikan risiko kredit nasabah dalam melakukan pinjaman/kredit
Geography	Negara tempat nasabah berdomisili
Gender	Jenis kelamin dari setiap pelanggan
Age	Usia dari setiap pelanggan
Tenure	Lama pelanggan telah menjadi nasabah bank dalam tahun
Balance	Saldo yang terdapat pada rekening nasabah
NumOfProducts	Jumlah produk dari bank yang digunakan nasabah
HasCrCard	Menunjukkan apabila nasabah mempunyai kartu kredit (1) atau tidak mempunyai kartu kredit (0)
IsActiveMember	Menunjukkan apabila nasabah masih aktif (1) atau tidak aktif (0)
EstimatedSalary	Gaji per tahun dari nasabah
Exited	Menunjukkan apabila pelanggan telah melakukan <i>churn</i> (1) atau tidak <i>churn</i> (0)

- **Penghapusan atribut CustomerId dan Surname** karena memiliki variabel yang unik sehingga tidak berguna untuk model prediktif yang melakukan prediksi berdasarkan korelasi antar record.
- **Tipe data object64 diubah menjadi bentuk string** untuk memudahkan pemrosesan data.
- **Data kategorikal dipisahkan dengan menggunakan one-hot encoding.** Atribut Geography dipisah ke dalam France, Germany, dan Spain. Atribut Gender dipisah ke dalam 2 kelompok yakni Male dan Female. Pemisahan atribut tersebut akan menghasilkan tipe data biner

PEMBANGUNAN MODEL



Faktor yang mempengaruhi tingkat churn berdasarkan heatmap korelasi adalah sebagai berikut.

- Age
- Balance
- IsActiveMember
- Geography_France
- Geography_Germany
- Gender_Female
- Gender_Male

Ketujuh atribut akan digunakan untuk pembangunan model prediksi

PEMBANGUNAN MODEL

K-Nearest Neighbors

Setiap data atau records diperlakukan sebagai titik dan dikelompokkan sesuai tetangga terdekat.

Random Forest

Menggabungkan beberapa decision tree untuk memperoleh hasil prediksi yang akurat

Gradient Boosting Machine

Menggabungkan algoritma atau metode prediktif lain sehingga model lebih akurat

Support Vector Machine

Mengelompokkan data menjadi kelas-kelas yang berbeda berdasarkan jarak radial antar data

ANALISIS MODEL

Tingkat Akurasi

Model	Tingkat Akurasi
K-Nearest Neighbors	81,73%
Random Forest	82,67%
Gradient Boosting Machine	83,7%
Support Vector Machine	83,27%

Memberikan gambaran umum dari tingkat ketepatan model. Model yang paling baik berdasarkan tingkat akurasi adalah Gradient Boosting Machine

Confusion Matrix

Model	Prediksi Benar		Prediksi Salah	
	TRUE	FALSE	Kesalahan Tipe I	Kesalahan Tipe II
K-Nearest Neighbors	258	2194	179	369
Random Forest	220	2260	113	407
Gradient Boosting Machine	225	2286	87	402
Support Vector Machine	162	2336	37	465

Memberikan gambaran lebih rinci terkait tingkat akurasi. Model yang paling baik untuk melihat perilaku konsumen adalah model dengan kesalahan tipe II paling sedikit, yakni K-Nearest Neighbors

ANALISIS MODEL

Cost-Benefit Analysis

Model	Loss Insentif	Loss Churn	Total loss	Benefit nasabah tanpa insentif	Benefit nasabah dengan insentif	Total Benefit	Profit
K-Nearest Neighbors	\$ 179.000	\$ 846.689	\$ 1.025.689	\$ 591.994	\$ 5.516.157	\$ 6.108.151	\$ 5.082.462
Random Forest	\$ 113.000	\$ 933.882	\$ 1.046.882	\$ 504.801	\$ 5.516.157	\$ 6.020.958	\$ 4.974.076
Gradient Boosting Machine	\$ 87.000	\$ 922.409	\$ 1.009.409	\$ 516.274	\$ 5.516.157	\$ 6.032.431	\$ 5.023.022
Support Vector Machine	\$ 37.000	\$ 1.066.966	\$ 1.103.966	\$ 371.717	\$ 5.516.157	\$ 5.887.874	\$ 4.783.909

Berdasarkan hasil dari confusion matrix, dilakukan analisis biaya dan keuntungan. Diperoleh model yang memiliki tingkat profit tertinggi adalah K-Nearest Neighbors. Oleh karena itu, **model K-Nearest Neighbors dipilih sebagai model prediktif yang digunakan oleh Bank X**



KESIMPULAN

Strategi yang disarankan untuk digunakan oleh Bank X adalah menggunakan model prediktif K-Nearest Neighbors untuk memprediksi perilaku nasabah yang akan melakukan churn.

Berdasarkan hasil prediksi tersebut, dapat diberikan insentif atau program lainnya untuk mempertahankan nasabah di Bank X

THANK YOU
