**LAPORAN FINAL PROJECT**

**PREDICTIVE PIONEERS**

**Background**

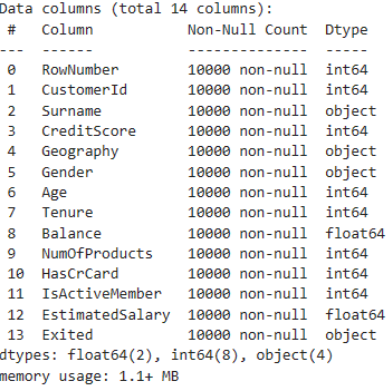
* Problem :

Bank Manduro, sebuah layanan jasa perbankan, menghadapi tantangan signifikan karena jumlah nasabah yang churn pada bulan November meningkat sebesar **20,4%**. Dampaknya akan terasa pada **pendapatan bank** yang berasal dari biaya layanan, bunga pinjaman, serta berbagai produk keuangan yang dimanfaatkan oleh nasabah.

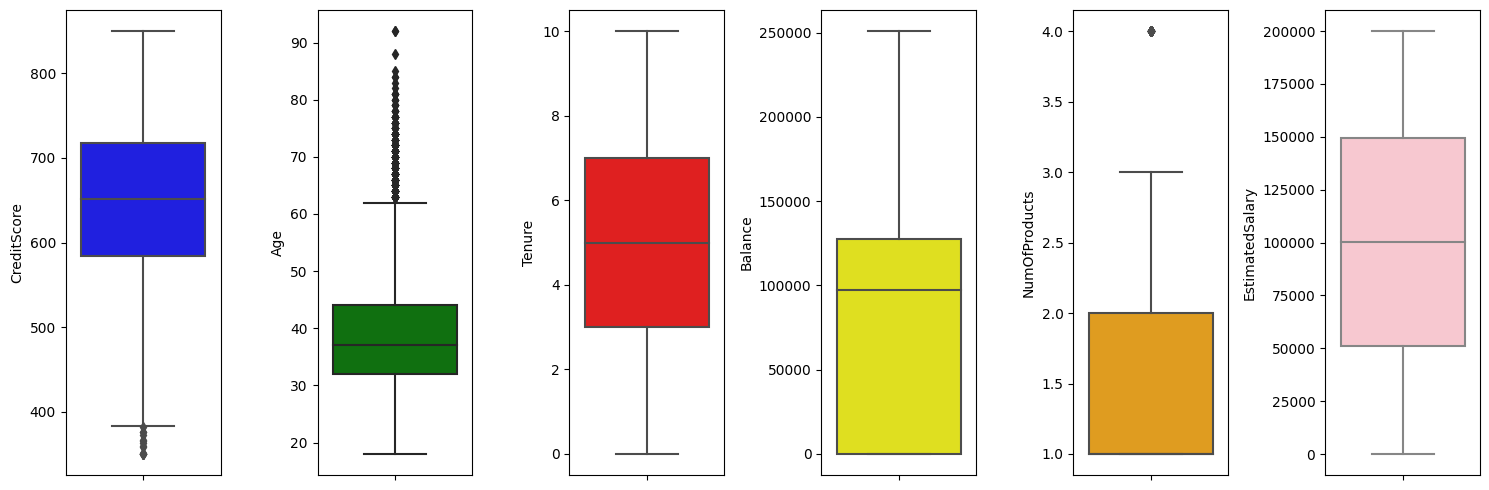
* Objective :
* **Mencari faktor** yang mempengaruhi tingkat churn nasabah.
* **Membuat model machine learning** yang dapat memprediksi nasabah yang memiliki kecenderungan churn
* Goals : **Mereduksi tingkat churn** pada Bank Manduro sebesar **3,4%** dalam waktu 1 bulan
* Business Metrics : ***Churn rate***

**EDA & Insight**

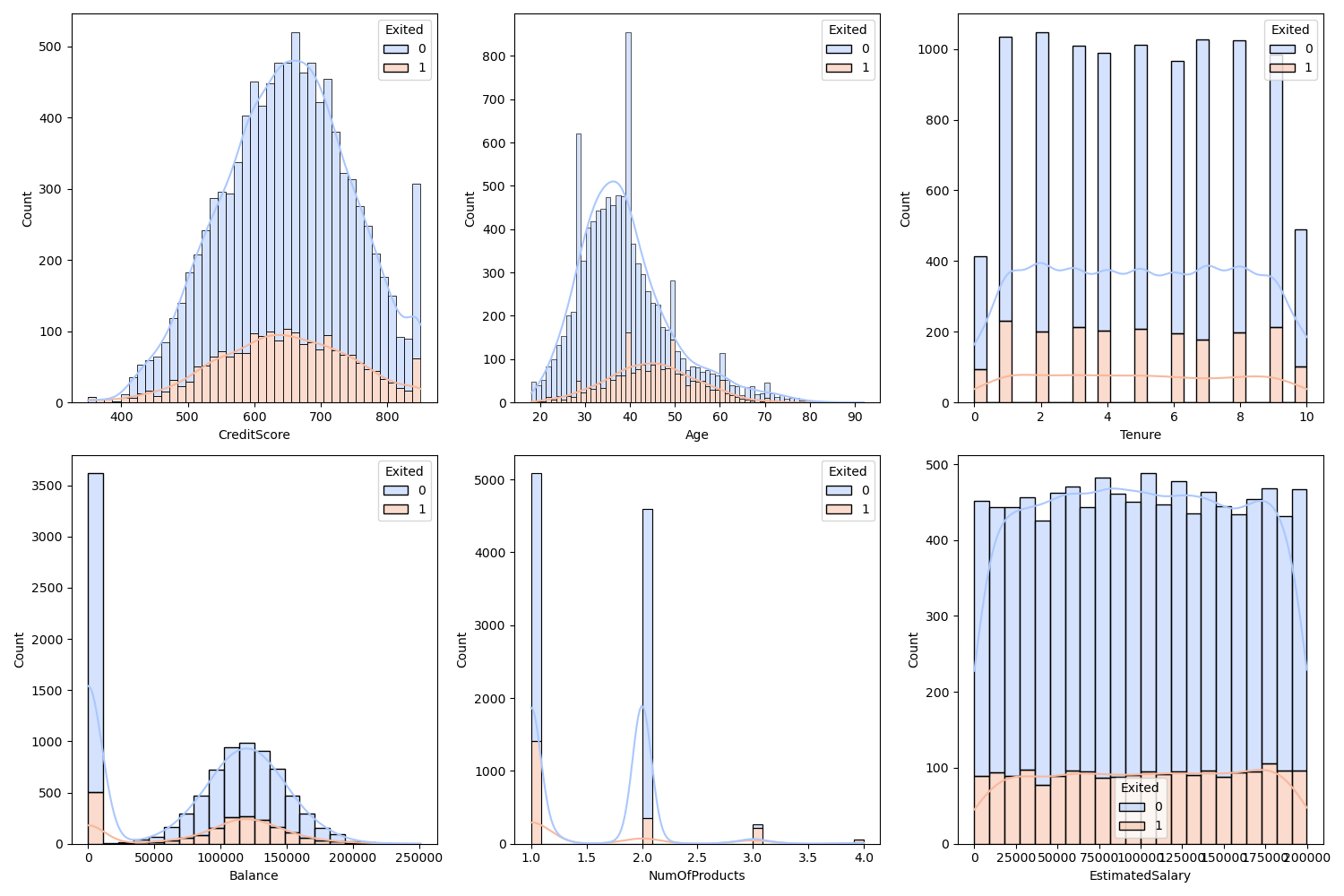
* Menampilkan info dataset



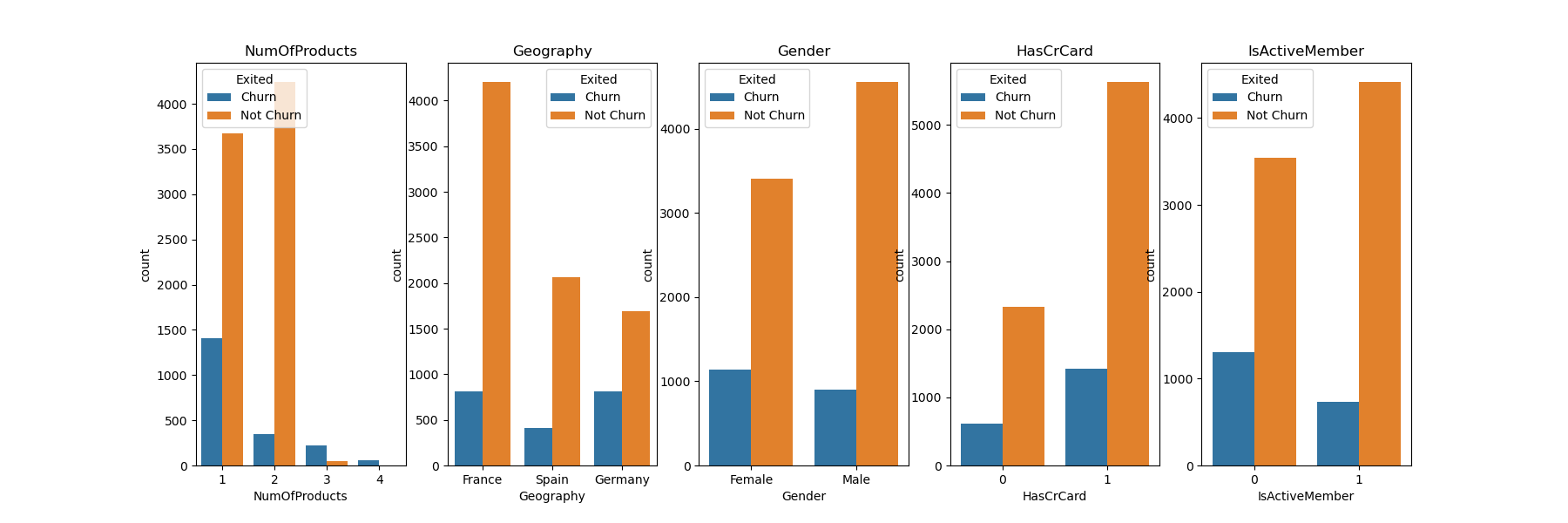
* **Semua tipe sudah sesuai dan tidak ada kolom NULL**

****

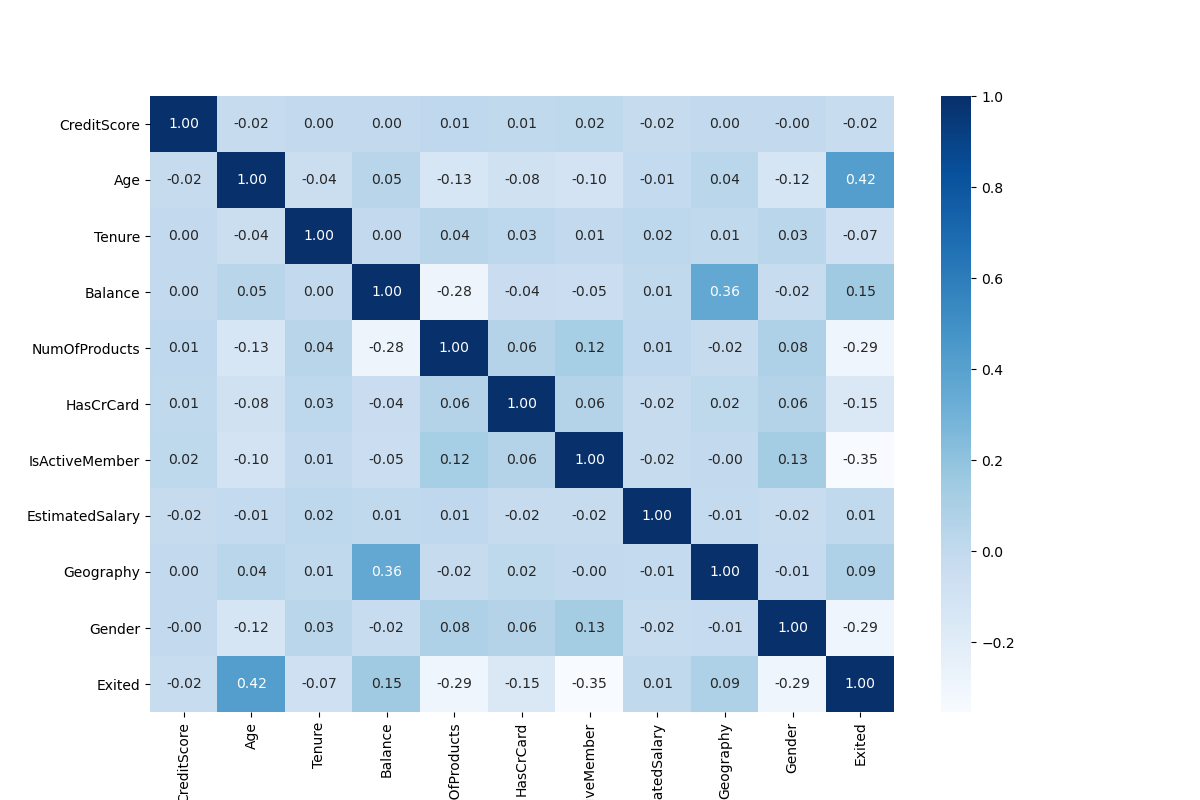
* Tiga variabel yaitu **Tenure, Balance, EstimatedSalary tidak memiliki *outlier***
* Variabel **CreditScore, Age, dan NumOfProduct memiliki nilai ekstrem** atau pencilan sehingga dapat berpengaruh dalam proses pemodelan
* **Variabel Age memiliki outlier paling banyak** sehingga dibutuhkan analisis lebih lanjut dengan analisis signifikansi, transformasi data, atau mengabaikan nilai *outlier.*



* Dalam kategori CreditScore, nasabah dengan rentang **score 580-699** memiliki kecenderungan **CHURN** yang **besar**.
* Dalam kategori Age, nasabah yang cenderung **CHURN** rata-rata berusia **40-49** tahun
* Dalam kategori Balance, nasabah yang cenderung **CHURN** memiliki saldo sebanyak **0 ( tidak memiliki saldo )**

****

* Dalam kategori NumOfProducts, nasabah dengan 1 produk memiliki kecenderungan **CHURN** yang besar.
* Dalam kategori Geography, nasabah yang cenderung **CHURN** bertempat tinggal di **Germany**
* Dalam kategori Gender, nasabah yang cenderung **CHURN** ialah **Perempuan**
* Dalam kategori HasCrCard, nasabah yang cenderung **CHURN** ialah nasabah yang mempunyai **Credit Card**
* Dalam kategori IsActiveMember, nasabah yang cenderung **CHURN** ialah member yang **tidak aktif**

****

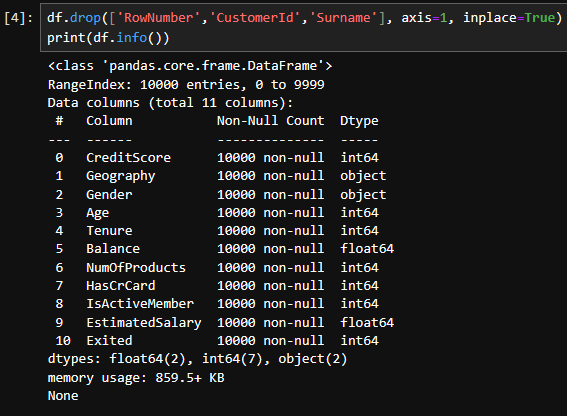
* Tidak ada korelasi yang kuat antar variabel (dibuktikan dengan nilai korelasi <0,5)
* Beberapa variabel memiliki korelasi negatif dengan variabel lainnya.
* Variabel yang paling berpengaruh terhadap *churn* yaitu Age dengan korelasi 0,42 yang menunjukkan bahwa semakin tua usia seseorang, semakin tinggi kemungkinan terjadi churn.
* Korelasi terbesar kedua terhadap churn yaitu dengan Balance sebesar 0,15 yang mengindikasikan pelanggan dengan saldo yang tinggi cenderung memiliki kecenderungan churn yang lebih rendah

**Insight** :

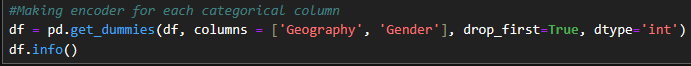
* **EstimatedSalary atau jumlah pendapatan tidak memiliki pengaruh terhadap churn, hal ini dapat dilihat dari distribusi yang hampir merata pada setiap jangkauan pendapatan atau EstimatedSalary.**
* **Usia memiliki pengaruh terhadap churn, hal ini dapat dilihat dari distribusi normal pada age atau usia, dengan nasabah yang melakukan exit atau churn umumnya di rentang usia 40-49 tahun.**
* **Balance memiliki pengaruh terhadap churn, hal ini dapat dilihat dari distribusi normal pada Balance, dengan nasabah yang melakukan exit atau churn umumnya memiliki jangkauan balance 0.**
* **HasActiveMember berpengaruh terhadap tingkat Churn. Hal ini dapat di lihat dari nasabah yang melakukan exited atau churn umumnya bukan member aktif.**
* **Tenure atau masa nasabah menggunakan layanan bank tidak memiliki pengaruh terhadap churn, hal ini dapat dilihat dari distribusi yang hampir merata pada setiap jangkauan tenure**
* **Credit score memiliki pengaruh terhadap churn, hal ini dapat dilihat dari distribusi normal pada creditscore, dengan nasabah yang melakukan exit atau churn memiliki jangkauan credit score 580-699.**
* **Berdasarkan analisa menggunakan kolom Geography, kami mendapatkan bahwa negara germany memiliki tingkat yang berpotensi churn terbanyak diantara kedua negara lainnya. Oleh karena itu, perlunya action yang harus dilakukan terhadap negara tersebut agar pelanggan dapat betah dengan product yang ditawarkan.**
* **Berdasarkan analisa menggunakan kolom Gender, didapatkan bahwa nasabah dengan jenis kelamin female lebih berpotensi untuk melakukan churn daripada male. Oleh karna itu, bank harus lebih memberikan banyak tindakan terhadap nasabah perempuan agar tetap menjadi nasabah pada bank tersebut. Jadi, nasabah dengan jenis kelamin female lebih berpotensi untuk exited (churn) dibandingkan dengan jenis kelamin male.**

**Pre-Processing**

**Feature Selection :**

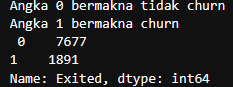
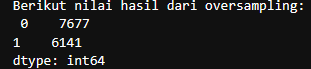
****

Kami hanya memilih kolom-kolom tertentu saja yang akan dijadikan fitur independen pada pemodelan nantinya

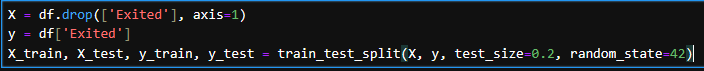


Kolom Geography dan Gender merupakan kolom yang berisikan data kategorik sehingga harus diubah menjadi numerik dahulu sebelum melanjutkan ke tahap pemodelan. Kami menggunakan One-Hot Encoding.

**Handling imbalance class :**

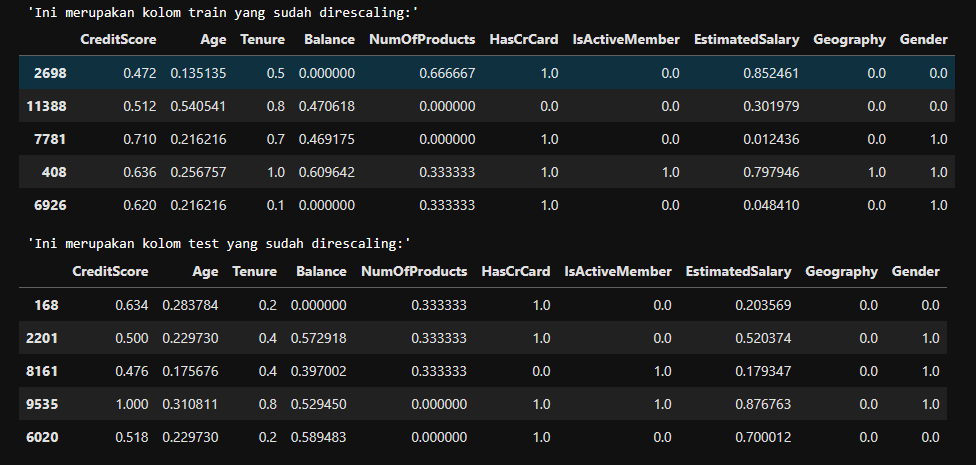
** **

Kami melakukan imbalancing class menggunakan SMOTE dengan menggunakan sampling strategy 0.8.



Kami melakukan splitting data dengan porsi train 80 : test 20.

Lalu kami melakukan scaling terhadap data train dan test menggunakan MinMaxScaler seperti gambar dibawah ini:



**Modeling**

**Model yang dipakai untuk melakukan klasifikasi yaitu RandomForest**

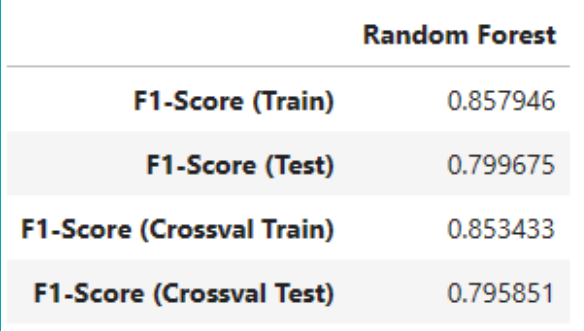
**Kenapa RandomForest?**

* **Kemampuan Menangani Ketidakseimbangan:** RandomForest memiliki kemampuan yang baik dalam menangani ketidakseimbangan kelas.
* **Performa yang Baik:** RandomForest memiliki performa yang kuat dan cenderung memberikan hasil yang baik dalam klasifikasi, terutama saat ada ketidakseimbangan kelas.
* **Pencegahan Overfitting:** RandomForest memiliki teknik regulasi bawaan dan pencegahan overfitting

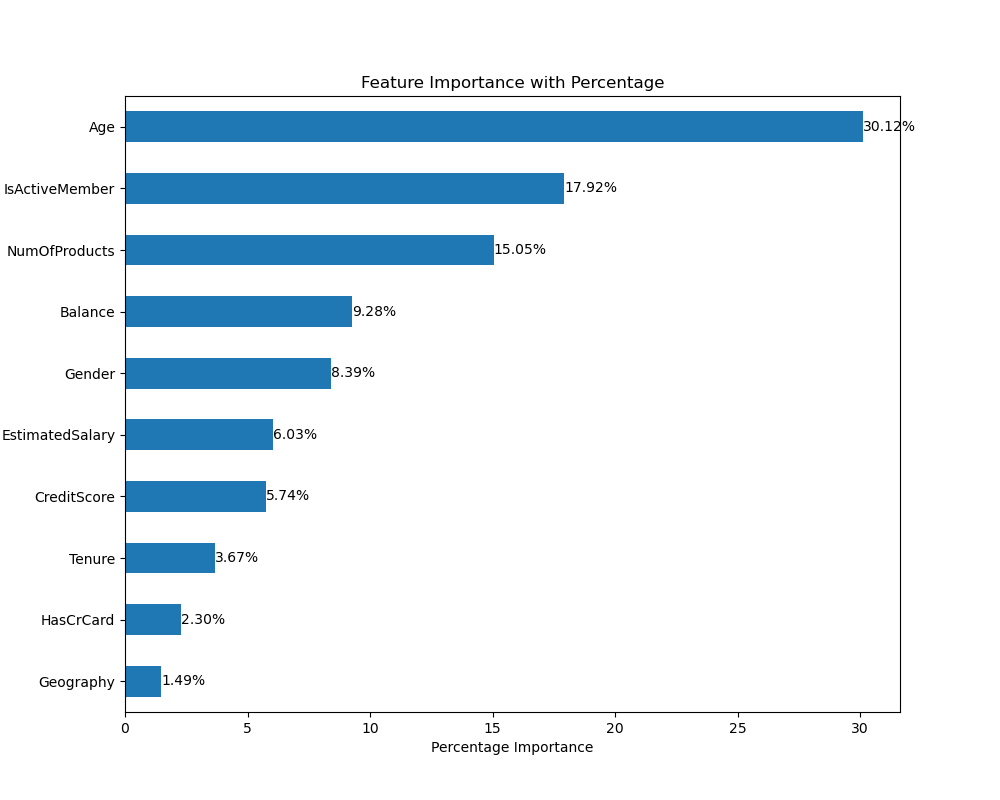
**Pemilihan metrics F1-Score, kenapa ?**

Kami memilih metrics evaluation F1-SCORE karena metrics ini yang paling sejalan dengan business metrics kami yaitu menurunkan jumlah churn nasabah.

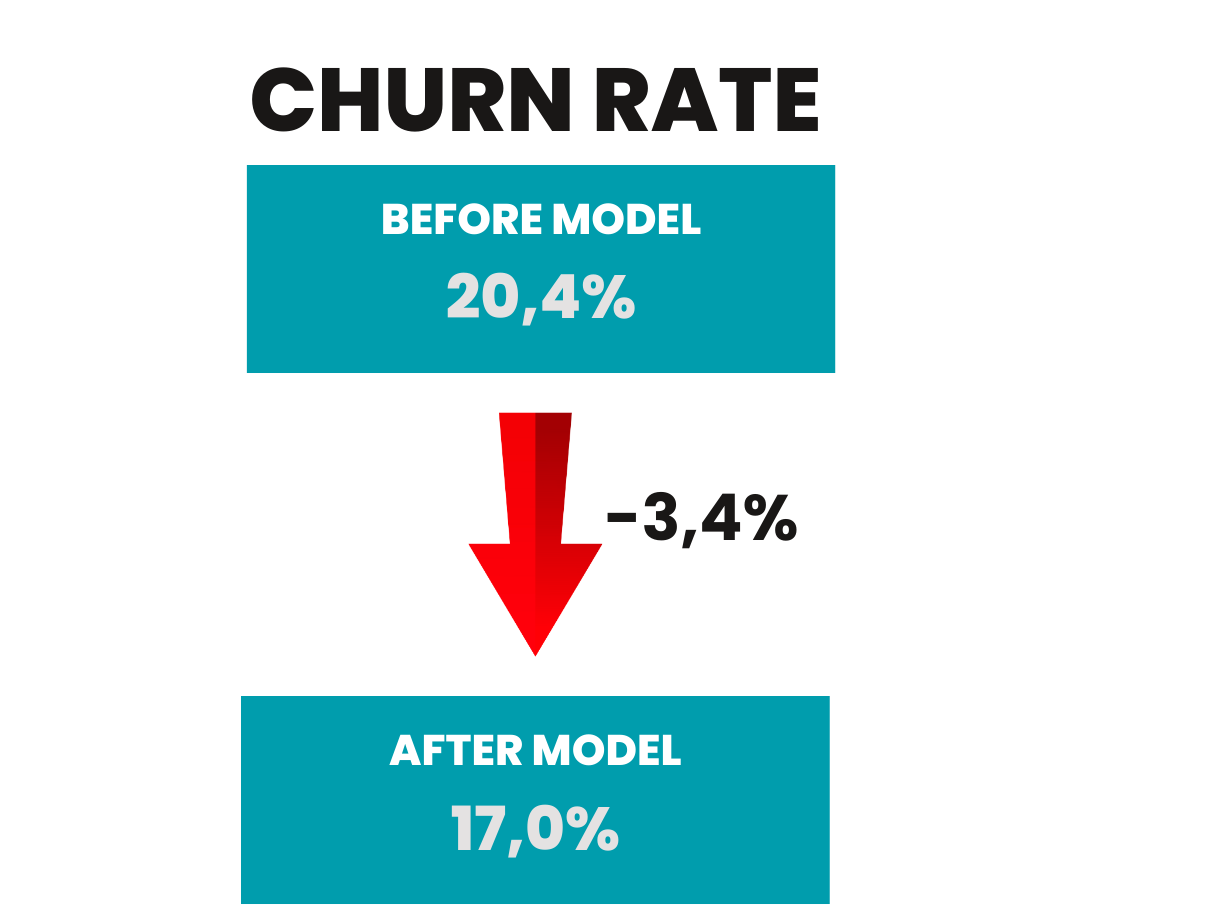
Penggunaan F1-SCORE kami harapkan dapat menghasilkan model dengan FN dan FP sekecil mungkin sehingga model ini dapat bekerja dengan maksimal untuk menjadi solusi bagi business metrics kami.



Rentang antara train dan test tidak jauh yang artinya model **BEST FIT**

****

Fitur **Age** menjadi fitur terpenting dalam model **RandomForest**.

****

Terdapat penurunan tingkat churn rate sebesar **3,4%** setelah menggunakan model **RandomForest** untuk mendeteksi nasabah yang akan churn di kemudian hari.

**Recommendation**

Rekomendasi tindakan untuk bank, yaitu sebagai berikut:

1. **Wawasan dari Model**

Usia berperan signifikan dalam model **RandomForest**, menunjukkan bahwa usia Nasabah pada bank dapat menjadi faktor penting dalam **memprediksi churn**. Ini menandakan bahwa Nasabah yang berumur **lebih tua** akan berpotensi churn daripada Nasabah yang lebih muda

1. **Wawasan Bisnis**

Jumlah produk yang dimiliki pelanggan dan status keanggotaan aktif adalah key factor teratas yang mempengaruhi churn. Ini mengindikasikan bahwa penting bagi bank untuk mempertahankan pelanggan agar tetap memiliki lebih banyak produk dan tetap aktif.

1. **Strategi Retensi:** Fokus pada strategi retensi untuk **nasabah berusia 40-49 tahun**, karena mereka memiliki kecenderungan untuk churn. Bank dapat **menawarkan insentif khusus, layanan tambahan**, atau **program loyalitas** untuk mempertahankan mereka.
2. **Pengoptimalan Strategi Pemasaran:** Bank perlu mengoptimalkan strategi pemasaran dan layanan yang lebih spesifik di **Germany** seperti melakukan **kampanye pemasaran** dan **memberikan penawaran menarik** untuk daerah tersebut
3. **Peningkatan Keterlibatan Nasabah:** Dorong keaktifan nasabah dengan memberikan **dorongan atau insentif** bagi yang **kurang aktif dan hanya memiliki 1 produk**, seperti **diskon untuk transaksi rutin** atau **program loyalitas**.

**IMPACTS FOR BANK MANDURO**

* **Membantu mendeteksi** nasabah yang akan churn sebelum nasabah tersebut churn.
* **Membantu pekerjaan** lebih mudah dikarenakan tim bank dapat mengecek Nasabah yang memiliki potensi churn saja tanpa harus mengecek seluruh Nasabah
* **Membantu mengurangi** kehilanggan Nasabah, kepercayaan dan hilangnya pendapatan pada bank

**Pembagian Tugas**

| Nama | Stage 0 | Stage 1 | Stage 2 | Stage 3 | Stage 4 |
| --- | --- | --- | --- | --- | --- |
| Muhammad Irfan Karim | problem statement | multivariate analysis | feature engineering | ML evaluation | Modeling, preprocessing |
| Ariel nathania | role | business insight | feature engineering | unsupervised learning | recommendation |
| Veby pebiola br simanjorang | goal | multivariate analysis | data cleansing | unsupervised learning | recommendation |
| Deyyana aulia hakim | business metrics | univariate analysis | data cleansing | ML evaluation | eda & insight |
| Anita putri | business metrics | make git | feature engineering | unsupervised learning | background |
| Tiya novita | business metrics | descriptive statistics | data cleansing | ML evaluation | background |
| Setya dewa nugraha | objective | univariate analysis | feature engineering | unsupervised learning | eda & insight |
| robi m subhan | objective | - | - | - | - |