

# Final Project

## Stage 2 - Data Preprocessing

### Churn Modeling Dataset

Presented by



# 1. Data Cleansing



- A. Berdasarkan data.info() tidak ditemukan data yang kosong atau hilang.
- B. Tidak terdapat kolom yang duplikat dari semua kolom
- C. Kita menghapus data outlier sekitar ~141 Baris dengan menggunakan z-score.  
Jumlah baris sebelum memfilter outlier: 10000  
Jumlah baris setelah memfilter outlier: 9859
- D. Kita melakukan standarisasi untuk feature yang bertipe numeric agar skalanya cenderung sama.
- E. Kita melakukan feature Encoding terhadap 2 Feature:  
**Gender : label encoding**  
**Geography: One Hot Encoding**
- F. Karena data target kita tidak seimbang atau imbalance maka kami menggunakan oversampling SMOTE untuk membuat data seimbang.  
Data awal : Original Exited (0=5485    1= 1416)  
Data Setelah : Original Exited (0=5485    1= 5485)



## 2. Feature Engineering

### A. Feature Selection

- Dilakukan pemilihan fitur dengan menghapus kolom RowNumber, CustomerId, dan Surname
- Karena fitur EstimatedSalary menampilkan distribusi seragam untuk kedua jenis pelanggan, maka dapat dihilangkan juga.
- Setelah dilakukan analisis, fitur Tenure dan HasCrCard memiliki chi-kuadrat yang kecil dan nilai p lebih besar dari 0,05. sehingga dapat disimpulkan bahwa hipotesis awal kami bahwa kedua fitur ini tidak berpengaruh secara signifikan terhadap target. Maka kedua fitur juga akan dibuang atau dihilangkan.



## 2. Feature Engineering



### B. Feature Extraction

- **Balance\_to\_Salary\_Ratio:** Rasio antara saldo akun dan estimasi gaji pelanggan. Fitur ini mungkin dapat memberikan wawasan tentang bagaimana proporsi saldo akun pelanggan dibandingkan dengan besarnya estimasi gaji mereka. Jika rasio ini tinggi, mungkin menunjukkan bahwa pelanggan memiliki saldo yang besar dibandingkan dengan pendapatannya, yang bisa menjadi indikasi kestabilan finansial atau potensi untuk menjadi nasabah yang berharga bagi bank.
- **Age\_to\_Tenure\_Ratio:** Rasio antara usia pelanggan dan masa jabatan (tenure) mereka di bank. Fitur ini dapat memberikan informasi tentang seberapa lama pelanggan telah menjadi nasabah bank relatif terhadap usia mereka. Hal ini dapat membantu dalam memahami tingkat loyalitas pelanggan atau seberapa lama mereka telah terlibat dalam hubungan dengan bank.



## 2. Feature Engineering

### C. Feature Tambahan

- **Total\_Transaction\_Amount:** Menambahkan fitur yang menggambarkan total jumlah transaksi keuangan yang dilakukan oleh pelanggan. Fitur ini dapat memberikan wawasan tentang seberapa aktif pelanggan dalam menggunakan layanan perbankan, yang dapat menjadi indikator kegiatan finansial dan keterlibatan dengan bank.
- **Credit\_Utilization\_Ratio:** Rasio antara total kredit yang dimiliki oleh pelanggan dengan batas kredit yang tersedia. Fitur ini dapat memberikan informasi tentang seberapa baik pelanggan mengelola penggunaan kreditnya. Jika rasio ini tinggi, itu bisa menjadi pertanda potensial bahwa pelanggan berisiko tinggi untuk mengalami kesulitan keuangan.
- **Average\_Daily\_Balance:** Rata-rata saldo harian pelanggan selama periode tertentu. Fitur ini dapat memberikan gambaran tentang seberapa stabil keuangan pelanggan selama periode waktu tertentu. Semakin tinggi rata-rata saldo harian, semakin stabil keuangan pelanggan tersebut.
- **Customer\_Satisfaction\_Score:** Skor kepuasan pelanggan yang diperoleh dari survei atau umpan balik pelanggan. Fitur ini dapat memberikan informasi tentang seberapa puas pelanggan terhadap layanan atau produk bank. Skor kepuasan yang tinggi dapat menunjukkan loyalitas pelanggan yang tinggi dan potensi untuk tetap menjadi pelanggan setia.

# 3. Git





# Thank You

