McKidney & Company

HSBC

Churn Prediction
For HSBC Holdings Bank

By ONE TEAM from McKidney&Company

One Team - Member

- Ismail Raji

- Muhammad Bian Suryoprakoso

- Sulthan Aulia Muhammad

McKidney&Company

Latar Belakang: Apa itu Churn Rate dan mengapa Churn Rate itu penting?

Deskripsi

- Churn = Customer yang berhenti menggunakan produk (**Tutup akun**)
- Churn Rate = Rasio customer yang berhenti menggunakan produk dari keseluruhan customer aktif

---

Latar Belakang General

- **Krisis finansial 2008**
- Akuisisi sebanyak-banyaknya customer → akuisisi customer baru memiliki biaya **7 kali lebih tinggi** dibanding untuk menjaga customer aktif

---

Latar Belakang HSBC

- Dataset: 2020, Customer churn HSBC sebanyak 2,000 dari 10,000 (20%)
- Menurut SopraBanking, **CAC (Customer Acquisition Cost) Retail Bank sekitar $300**
- Hidden loss Customer Churn 2020 = 2,000 x $300 = $600,000

Source: https://medium.com/@noah.fintech/creating-a-banking-customer-churn-model-1a2d0850f071
https://www.soprabanking.com/insights/how-to-reduce-customer-acquisition-costs/

McKidney&Company

Masalah yang ada dan bagaimana kita menghadapinya

Before

Business

Tanpa prediksi

Churn

20%

Hidden Loss

$600k

After

Business

Churn Prediction

Machine Learning

Marketing Effort

Churn

<=10%

Reduce Hidden Loss

McKidney&Company

Dataset

Feature Name	Description
CustomerId	Customer ID
Surname	
CreditScore	Customer's credit score
Geography	France, Germany, Spain
Gender	Female, Male
Age	
Tenure	Time from join (Month)
Balance	
NumOfProducts	Number of products that customer use
HasCrCard	Does the customer have a credit card through the bank? (Yes=1, No=0)
IsActiveMember	Is the customer an active member? (Yes=1, No=0)
EstimatedSalary	Estimated salary of the customer
Exited	Did the customer leave the bank within the last 12 month? (Yes=1, No=0)

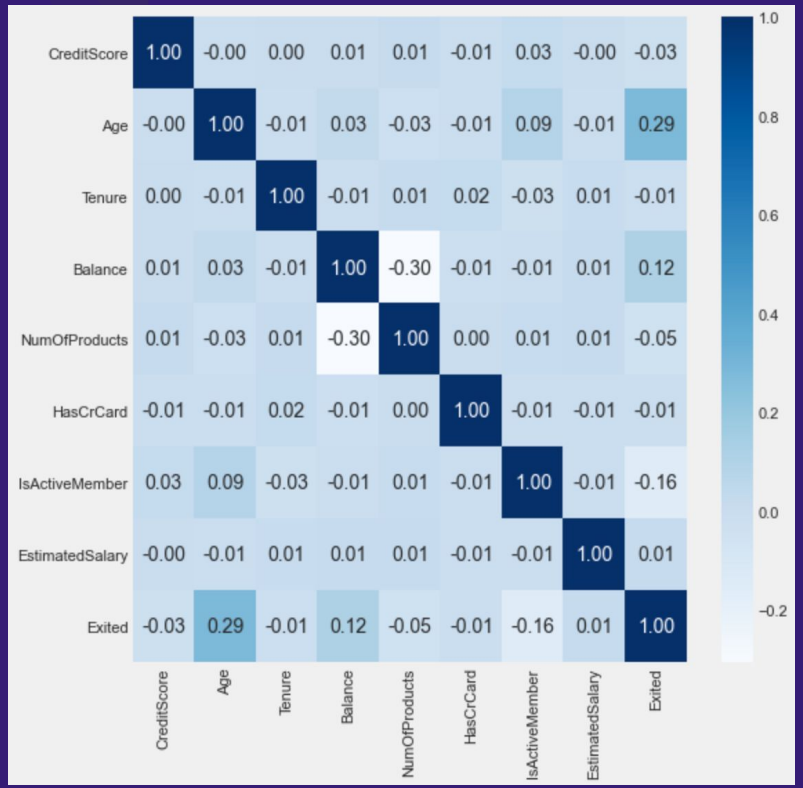
Keseluruhan Data

1. Total data memiliki 13 kolom/feature dengan jumlah 10,000 row
2. Tidak ada feature yang memiliki data null
3. Terdapat 3 feature categorical dan 10 feature numerical
4. Exited adalah target

Dominansi

1. Geography terdapat 3 negara, dengan mayoritas France
2. Gender didominasi Male dengan 54,5%

Exploratory Data Analysis: Korelasi



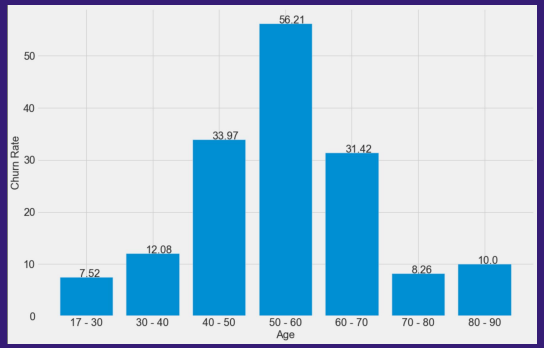
Age, Balance, dan IsActiveMember merupakan feature dengan korelasi yang cukup tinggi terhadap target dibanding feature lainnya.

Tidak ada pasangan fitur yang memiliki korelasi di atas 0.70.

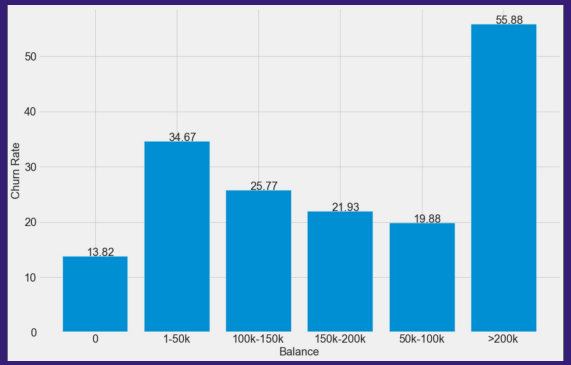
McKidney&Company

Exploratory Data Analysis: Hubungan Fitur (dengan korelasi) terhadap Target

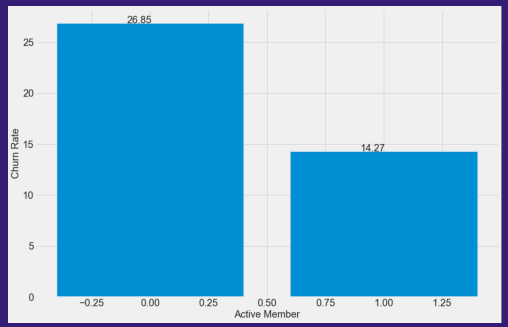
Age



Balance



IsActiveMember



1. Churn Rate tertinggi pada Age berkisar dari usia 40-70
2. Untuk Balance, Churn Rate cukup beragam, tetapi naik drastis untuk balance >200k
3. Non-Aktif member memiliki Churn Rate lebih tinggi dibanding yang aktif

McKidney&Company

Cleansing

- **Data Duplication and Null Checking -** Mengecek data duplicate dan null
  hasil yang didapat adalah **tidak ada data duplicate dan null**
- **Handling Outliers**
  Filtering outliers menggunakan **IQR untuk fitur yang berdistribusi skewed** dan z**-score yang berdistribusi normal,** data customer yang terhapus 367
- **Standardization**
  Melakukan **standardisasi untuk fitur numerik**
- **One hot encoding**
  Melakukan **one hot encoding untuk fitur Categorical karena fitur tidak punya urutan**
- **Pengurangan Fitur**
  Hapus fitur original yang belum di standarisasi dan one hot encoding agar tidak redundant
- **Class Imbalance**
  Menggunakan **SMOTE untuk mengatasi Class Imbalance** tetapi hanya dilakukan pada data train setelah split data, agar data test tidak menjadi bias

Modelling

- Metric yang diperhatikan
  Recall -> Memaksimalkan Jumlah prediksi Churn (memperkecil nasabah yang actualnya churn namun di prediksi tidak churn)

- Supporting Metric:
  Accuracy -> Persen target yang berhasil diprediksi model
  AUC -> Seberapa yakin model memprediksi target

- Dataset di test dengan Algorithma:
  - Logistic Regression
  - K Nearest Neighbour
  - Decision Tree
  - Random Forest

Hasil test sebelum Hyperparameter Tuning

Metric	LR	KNN	DT	RF
Recall	0.65	0.64	0.55	0.57
Accuracy	0.77	0.77	0.78	0.84
AUC	0.85	0.79	0.69	0.84

Hasil test setelah Hyperparameter Tuning

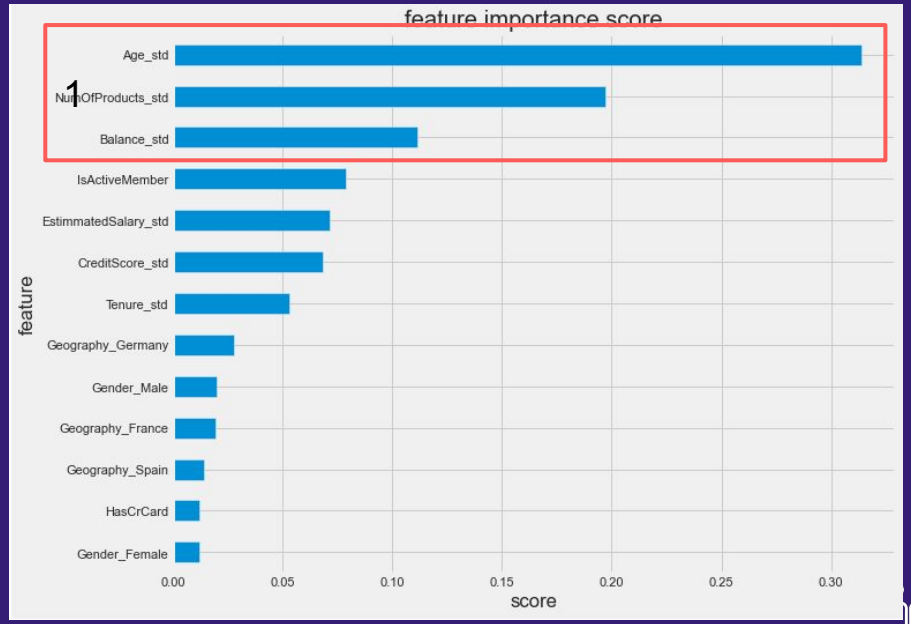
Metric	LR	KNN	DT	RF
Recall	0.70	0.64	0.82	0.66
Accuracy	0.76	0.62	0.67	0.83
AUC	0.80	0.78	0.67	0.85

Modeling - Pemilihan Algoritma

Alasan terpilih random forest

1. Recallnya naik secara signifikan setelah di tuning dan Fitur supporting yaitu Accurracy dan AUC stabil

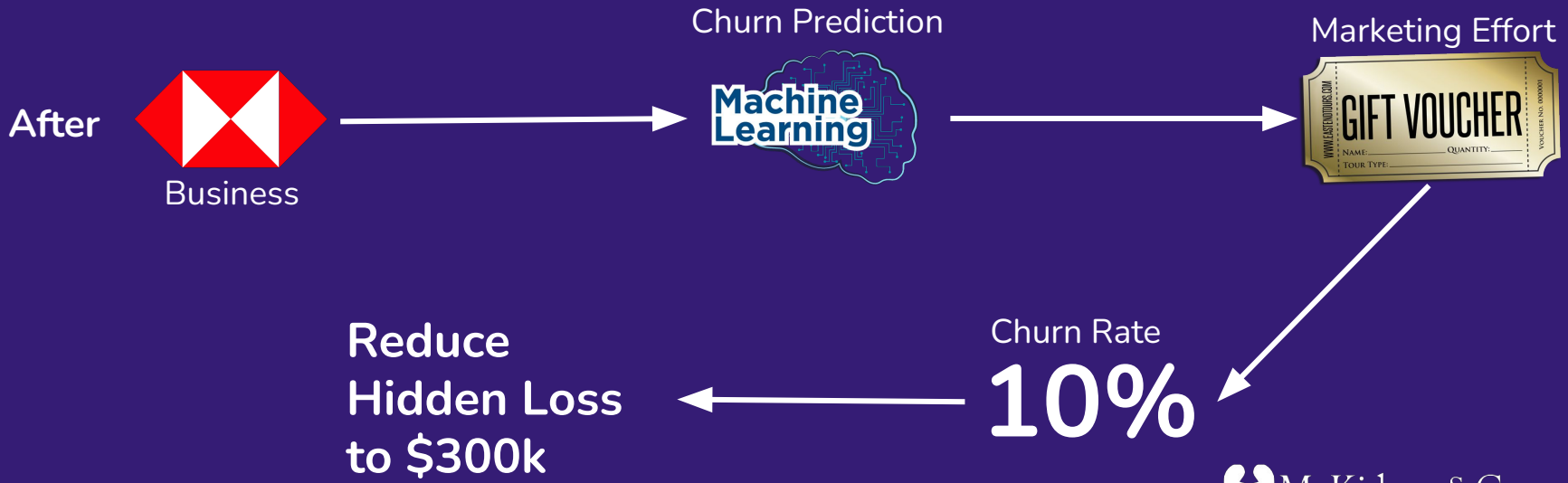
Feature Importance



McKidney&Company

Hasil Modelling

1.	Recall 0.66	-> True Predict 66% from 20%	= True Predict 13,2%
2.	If Marketing Effort success 75%	-> 75% from 13,2%	= Keep 10%
3.	Churn Rate	-> 20% - 10%	= 10%
4.	Reduce Hidden Loss	-> 1,000 x $300	= $300k

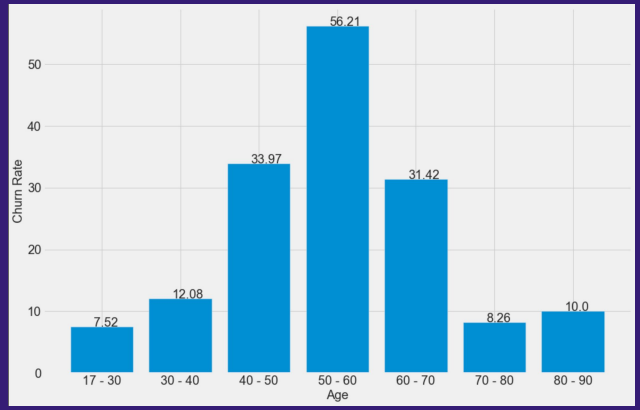


McKidney&Company

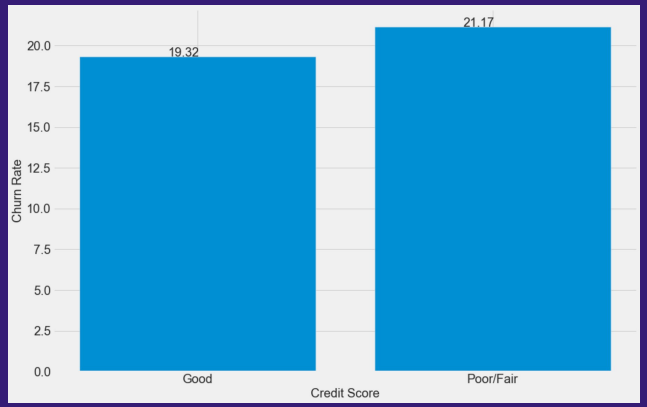
Business Insight #1

Probabilitas Churn yang lebih tinggi pada Age, kemungkinan disebabkan juga karena kombinasi dari fitur Age dan CreditScore

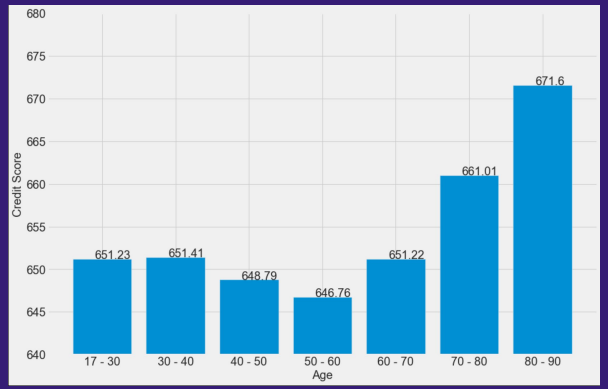
Age
vs
Churn Rate



Credit Score
(Based on FICO's Rating)
vs
Churn Rate



Age
vs
Credit Score

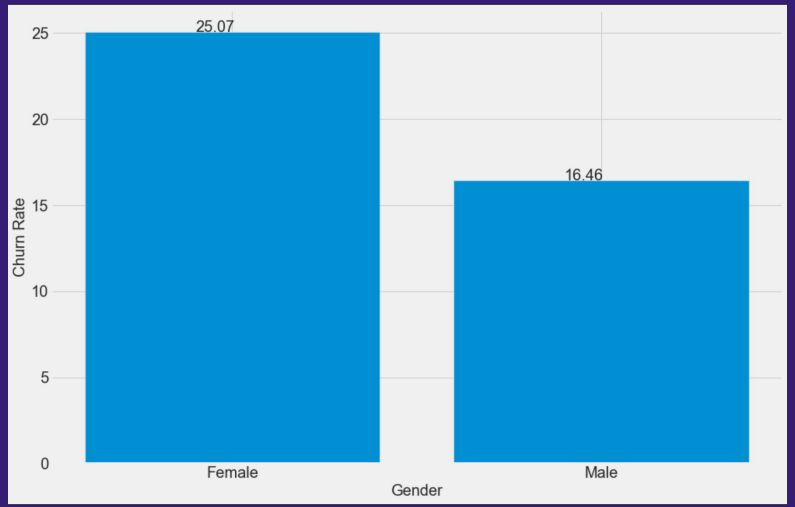


McKidney&Company

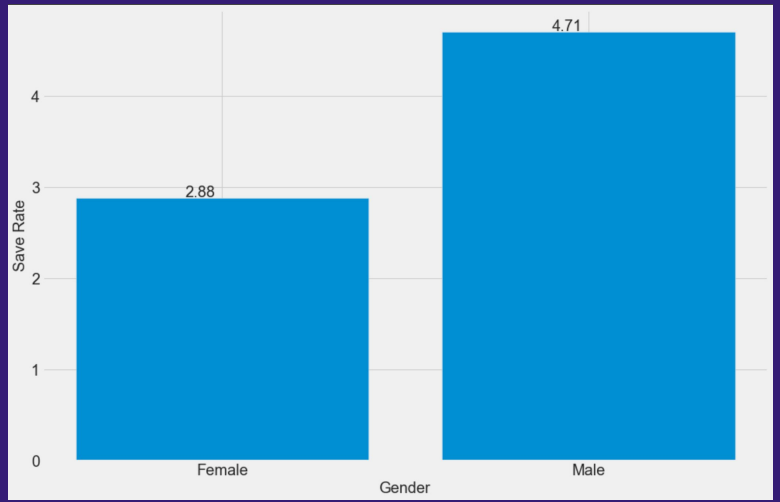
Business Insight #2

1. Wanita lebih berpotensi Churn
2. Wanita memiliki Save Rate (Balance/Salary) lebih rendah dibanding Pria -> Lebih suka membelanjakan uangnya

Gender vs Churn Rate



Gender vs Save Rate



McKidney&Company

Rekomendasi

Modelling

1. Untuk pengembangan model, menambahkan keterangan terkait jenis produk yang digunakan
2. Melakukan Clustering model dalam melakukan Marketing Effort agar konversi pencegahan Churn meningkat

Business Insight

1. Melakukan survey “Alasan Tutup Akun” untuk seluruh customer yang menutup akun agar mendapatkan alasan yang lebih valid sehingga hal tersebut dapat diperbaiki kedepannya
2. Mengutamakan Customer Service dan Real-Life Interaction untuk dapat menjaga hubungan dengan customer di rentang usia 40-60 dan customer Wanita
3. Membuat Loyalty Program untuk Customer, tetapi hanya untuk yang worth untuk diselamatkan. Usia terlalu tua (>60) dan Credit Score terlalu rendah (<475) akan berdampak pada Lifetime Customer Value yang kecil.
4. Membuat program Wanita-oriented, seperti Reward Point jika melakukan pembelanjaan.

Source: https://medium.com/@noah.fintech/creating-a-banking-customer-churn-model-1a2d0850f071

McKidney&Company