

Reza Dwi Putra

OUTLINE

02

Solving Churn

Business Understanding

Exploratory Data Analysis

03

04

Data Preprocessing & Modeling

05

Key Takeaways

01

Solving Churn



Churn

Churn modelling adalah sebuah pemodelan yang bertujuan untuk memprediksi pelanggan yang akan berhenti menggunakan jasa/produk di masa mendatang.

Dengan analisa churn perusahaan mampu membuat strategi bisnis untuk mempertahankan pelanggan yang akan 'churn'. Serta mengurangi biaya akuisisi pelanggan baru

Bagaimana mengurangi Tingkat Churn?

PREDIKSI

&

MENCARI FAKTOR KUNCI

02

Business Understanding





Data set customer churn diambil dari <u>kaggle</u> merupakan kumpulan data historis peminjam kredit bank (customer) yang memiliki beberapa detail dari masing-masing individu.

Data set tersebut juga memiliki keterangan tentang customer yang masih menggunakan layanan pinjaman dan yang sudah tidak menggunakan layanan pinjaman tersebut.

Dengan data set tersebut, dapat dibuat suatu churn modeling untuk memprediksi apakah customer akan berhenti menggunakan jasa pinjaman tersebut berdasarkan data historis pinjaman serta beberapa variabel lainnya yang harus diidentifikasi lebih lanjut

Identifikasi Variabel

Pada dataset churn predictive modelling tersebut terdapat beberapa variabel yang dapat menentukan apakah customer atau pegawai yang memiliki kredit pinjaman akan churn atau tidak churn melalui:

- 1. Row Number
- 2. Identitas (Customerld & Surname)
- 3. Customer Credit scores
- **4.** Letak Geografis Negara
- **5.** Gender
- **6.** Umur
- **7.** Tenor Pinjaman
- **8.** Saldo
- 9. Jumlah Produk Pinjaman yang dimiliki
- **10.** Kepemilikan kartu kredit
- **11.** Status aktif/tidak aktif sebagai peminjam
- **12.** Perkiraan gaji
- **13.** Status exit sebagai customer dari bank

Understanding

Problem: Bagaimana cara memprediksi suatu customer (individu yang melakukan pinjaman)

akan berhenti menggunakan jasa pinjaman tersebut atau tidak

Goals : 1. Mengetahui seorang customer akan berhenti atau tidak melalui data historis

Pinjaman.

2. Mengetahui Variabel yang mempengaruhi pelanggan yang akan churn

Objective: Membuat sistem untuk membantu tim pemasaran agar dapat mengambil

keputusan untuk mengetahui pelanggan yang melakukan pinjaman ke perusahaan

dan diharapkan dapat mempertahankan customer agar tidak churn

KPI / Metrik Bisnis

1. Mempertahankan Usia Pelanggan (Retention)

Tujuan utama dari analisa churn dalam kasus ini adalah mempertahankan customer agar dapat mengurangi biaya akuisisi customer baru.

2. Margin operasional

Biaya operasional dapat ditekan dengan menggunakan ML untuk analisa tingkat churn, dengan ini bisnis dapat mengurangi jumlah karyawan dan tentunya mengurangi biaya operasional yang dapat meningkat ketika pekerjaan dilakukan dengan cara manual

3. Trend based marketing

bisnis dapat menyesuaikan pendekatan marketing, misal seperti kondisi pandemi covid-19 ini, bisnis dapat meraih customer dengan membuat konten terkait #stayathome dan lain - lain. Bisnis dapat menggunakan variabel customer berupa demografi atau informasi lain sesuai features yang memberikan pengaruh besar terhadap tingkat churn

03

Exploratory Data Analysis



Statistika Deskriptif

	count	mean	std	min	25%	50%	75%	max
RowNumber	10000.0	5.000500e+03	2886.895680	1.00	2500.75	5.000500e+03	7.500250e+03	10000.00
CustomerId	10000.0	1.569094e+07	71936.186123	15565701.00	15628528.25	1.569074e+07	1.575323e+07	15815690.00
CreditScore	10000.0	6.505288e+02	96.653299	350.00	584.00	6.520000e+02	7.180000e+02	850.00
Age	10000.0	3.892180e+01	10.487806	18.00	32.00	3.700000e+01	4.400000e+01	92.00
Tenure	10000.0	5.012800e+00	2.892174	0.00	3.00	5.000000e+00	7.000000e+00	10.00
Balance	10000.0	7.648589e+04	62397.405202	0.00	0.00	9.719854e+04	1.276442e+05	250898.09
NumOfProducts	10000.0	1.530200e+00	0.581654	1.00	1.00	1.000000e+00	2.000000e+00	4.00
HasCrCard	10000.0	7.055000e-01	0.455840	0.00	0.00	1.000000e+00	1.000000e+00	1.00
IsActiveMember	10000.0	5.151000e-01	0.499797	0.00	0.00	1.000000e+00	1.000000e+00	1.00
EstimatedSalary	10000.0	1.000902e+05	57510.492818	11.58	51002.11	1.001939e+05	1.493882e+05	199992.48
Exited	10000.0	2.037000e-01	0.402769	0.00	0.00	0.000000e+00	0.000000e+00	1.00

Statistika Deskriptif

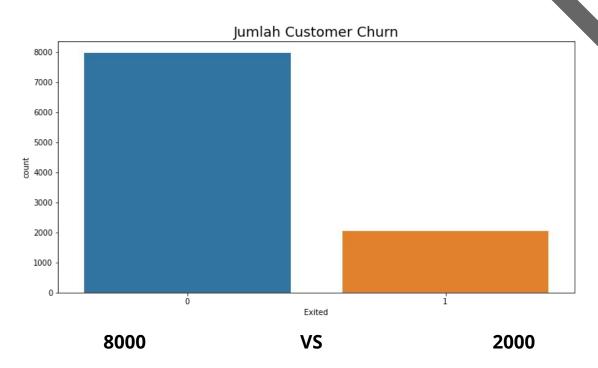
Data set memiliki **14 Features**, **10000 Row Data pelanggan**. Untuk melanjutkan analisa, beberapa features akan dibuang, yaitu RowNumber, CustomerID, Surname karena tidak memiliki informasi tambahan pada model.

Tidak terdapat missing value pada dataset

Selanjutnya Features lainnya akan dilakukan EDA untuk melihat hubungan Variabel prediktor dengan variabel target (Exited)

Target Variabel

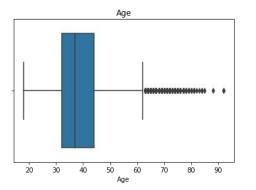
Terdapat ketimpangan pada target variabel (exited), Maka perlu dilakukan Smote untuk menambahkan data sintetik agar model klasifikasi tidak bias

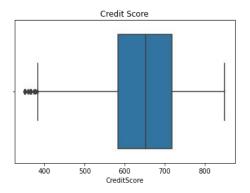


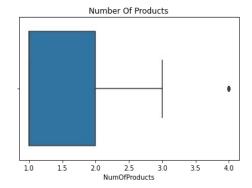
Terdapat outlier pada beberapa features khususnya pada features Age

Namun outlier dirasa masih normal secara angka

Outlier





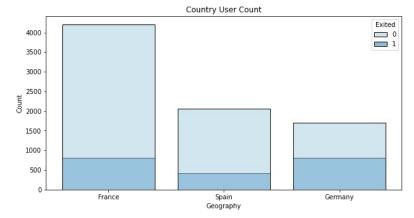


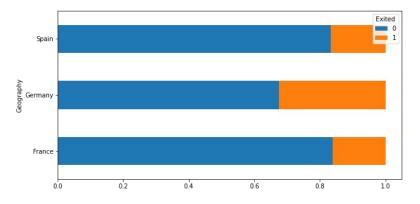
By Country Analysis

Negara dengan pengguna layanan tertinggi secara jumlah adalah Perancis.

Negara dengan pengguna layanan terendah secara jumlah adalah Jerman.

Negara dengan tingkat churn tertinggi secara rasio adalah Jerman.

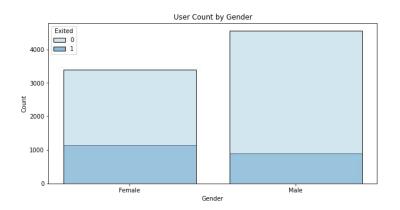


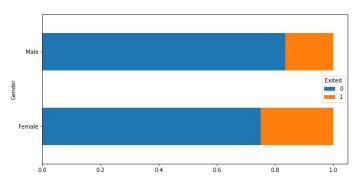


By Gender Analysis

Jumlah pengguna layanan t**erbanyak** adalah laki-laki

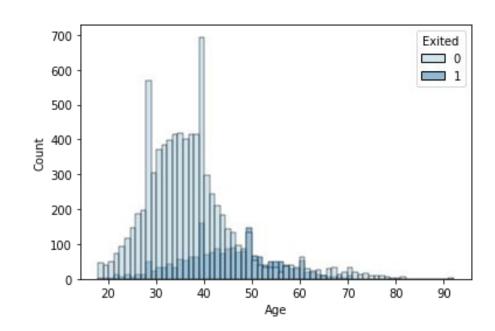
Secara rasio, **perempuan lebih banyak churn** daripada laki-laki





By Age Analysis

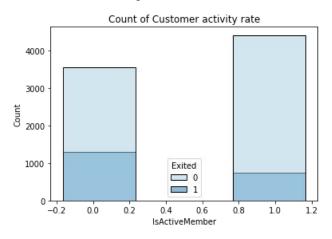
Range **usia 40-50 Tahun lebih banyak churn** dibandingkan range usia 40 kebawah

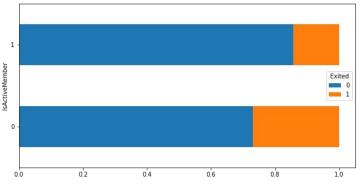


By Customer Activity Rate

Berdasarkan keaktifan dari customer, terlihat bahwa jumlah customer yang aktif lebih banyak daripada yang tidak aktif.

Selain itu, secara rasio, customer yang paling banyak churn adalah customer yang berada pada kategori tidak aktif

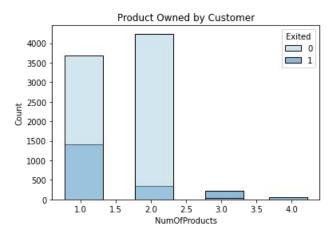


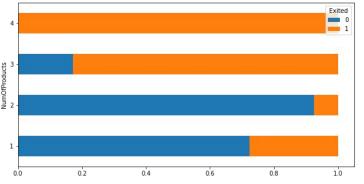


By Owning Products

Bersarkan jumlah produk yang dimiliki oleh customer, **mayoritas dari customer hanya memiliki 2 atau kurang**,

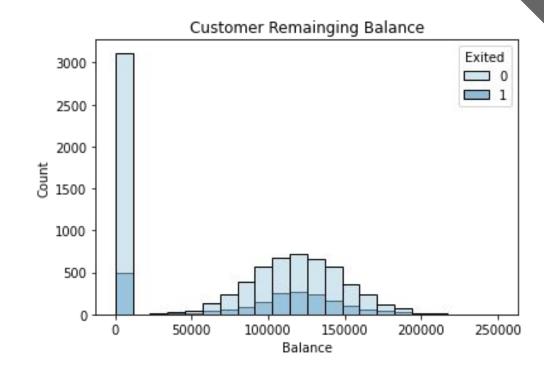
Secara rasio, customer yang paling banyak churn adalah kelompok customer dengan kepemilikan 3 atau 4 produk





By Customer Balance

Terlihat pada distribusi data, mayoritas dari customer yang churn memiliki saldo 0



Correlation Heatmap (spearman)

Features Age (0.32), Balance (0.11), memiliki nilai positif (walaupun lemah). Num Of Product (-0.13) dan Active Member (-0.16) memiliki nilai negatif (walaupun lemah)



-08

- 0.6

0.4

- 0.2

04

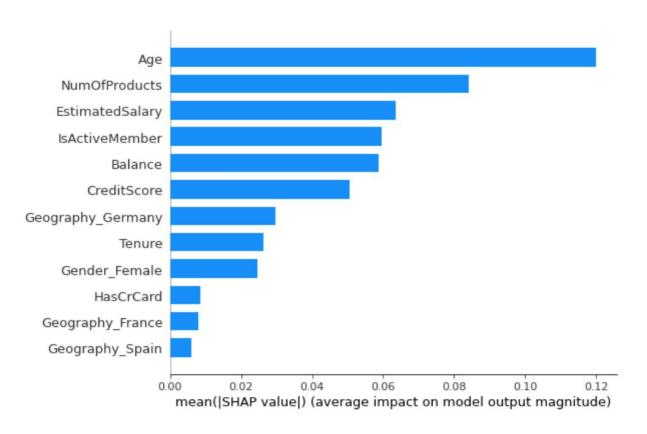
Preprocessing & Modeling



Preprocessing

- Handling categorical data (Geography & Gender)
- Feature Importance (SHAP Method)
- Feature Scaling (Standard Scaler)
- Balancing (SMOTE)

Feature Importance



Balancing Data

Rasio Train Test (80:20)

Sebel	um	Sesudah			
Not churn	6377	Not churn	6377		
Churn	1623	Churn	6377		

Jumlah Data sintetik : 4377 (174%)

Balancing Data

Rasio Train Test (80:20)

Sebel	um	Sesudah			
Not churn	6377	Not churn	6377		
Churn	1623	Churn	6377		

Jumlah Data sintetik : 4377 (174%)

Model & Evaluation Matrix

	KNN	Ranfor	Dec Tree	XG Boost	Stacking	
Akurasi	75.35%	82.05%	81.65%	83.95%	84.30%	
Presisi	43.78%	55.12%	54.75%	60.13%	66.03%	
Recall	67.15%	71.50%	65.46%	66.67%	49.76%	71.50% Ranfor
F1-Score	53.00%	62.25%	59.63%	63.23%	56.75%	63.23% XGBoost
-						

Model Prediksi 1

Dengan Model **Random Forest**, Prediksi terhadap pelanggan Churn dapat dilakukan dengan tingkat **Recall** 71%

Dengan Recall yang tinggi, dapat menargetkan lebih banyak orang untuk dilakukan retention marketing

Tradeoff: nilai presisi rendah (55%), dimana akan lebih banyak orang dan memakan biaya yg lebih tinggi

Model Prediksi 2

Dengan Model **XGboost**, Prediksi terhadap pelanggan Churn dapat dilakukan dengan tingkat **F1 Score 63.23**%

Dengan F1 Score yang lebih tinggi, dapat menargetkan strategi selanjutnya kepada target yang lebih tepat dengan biaya yang lebih efisien.

Tradeoff: nilai Recall sedikit lebih rendah dari model Ranfor, namun nilai Presisi lebih tinggi.

05

Key Takeaways

Pemilihan Model

Pemilihan salah satu model dapat disesuaikan dengan budget marketing dari perusahaan dengan rasio jumlah pelanggan yang akan ditargetkan untuk mendapat retention marketing

Faktor Kunci

Berdasarkan Feature Importance terdapat beberapa fitur yang harus diperhatikan yang mempengaruhi churn/tidaknya pelanggan secara general:

- 1. Umur
- 2. Jumlah Produk yang dimiliki pelanggan
- 3. Gaji
- 4. Keaktifan pelanggan
- 5. Saldo

THANKS!

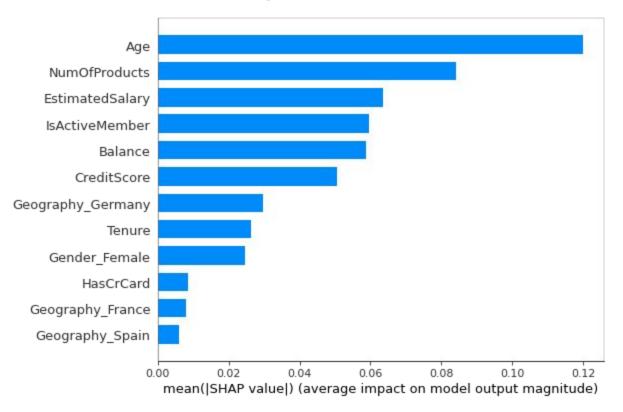
Appendix



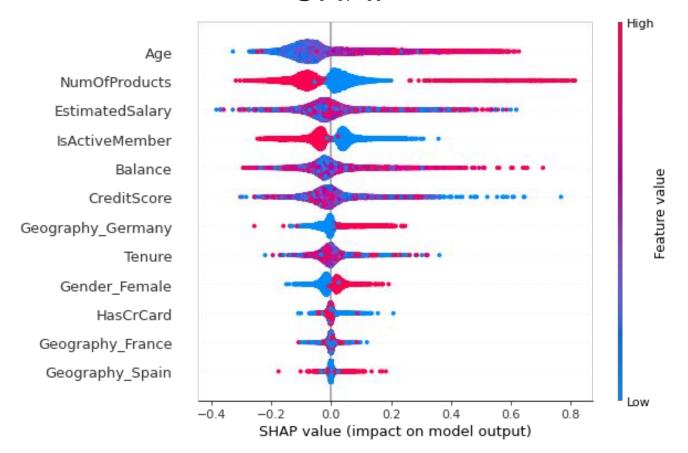
Eval Matrix

	w/o Smote					
	KNN	Ranfor	Dec Tree	XG Boost	Stacking	
Akurasi	83.35%	85%	84.85%	85.45%	85.55%	85.55% Stacking
Presisi	69.01%	79.38%	71.94%	77.58%	75.51%	79.38% Ranfor
Recall	35.51%	37.20%	43.96%	41.79%	44.69%	44.69% Stacking
F1-Score	46.89%	50.66%	54.57%	54.32%	56.15%	56.15% Stacking
	w/ Smote					
	KNN	Ranfor	Dec Tree	XG Boost	Stacking	
Akurasi	75.35%	82.05%	81.65%	83.95%	84.30%	84.30% Stacking
Presisi	43.78%	55.12%	54.75%	60.13%	66.03%	66.03% Stacking
Recall	67.15%	71.50%	65.46%	66.67%	49.76%	71.50% Ranfor
F1-Score	53.00%	62.25%	59.63%	63.23%	56.75%	63.23% XGBoost

SHAP



SHAP



Ranfor regressor Feature Importance

