

E-COMMERCE CHURN PREDICTION

Bootcamp Data Science Rakamin Academy Batch 22







Adi Herianto R. (Leader)



Ryana Tammi Putri



Dian Maulida



Susi Ninik Mahudi



Virginia Wulansari



Andini Febriati



TABLE OF CONTENTS



Project Background



EDA (Exploratory Data Analysis)



Data Pre-Processing



Modeling & Evaluation



Business Insight & Recommendation





ROLE



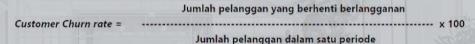
Avengers merupakan tim **Data Scientist** di sebuah e-commerce bernama Marvel Shop yang bertugas untuk mencari **solusi** dari masalah yang ada dan memberikan **rekomendasi bisnis** menggunakan dataset yang tersedia. Dataset: <u>Churn Prediction</u>



MARVEL SHOP??



946 Customer Churn 16,8% Churn Rate





Dilansir dari artikel gorgias.com, menurut Recurly Research, pada perusahaan dengan bisnis Consumer good angka churn rate yang standar itu ada di angka 8,56%

WHY??



CHURN RATE IS AN IMPORTANT PART OF GROWING A SUSTAINABLE ECOMMERCE BUSINESS.

Dilansir dari artikel gorgias.com, menurut VWO, 43%

keuntungan untuk e-commerce berasal dari pelanggan yang dipertahankan (tidak churn).

• Terlebih lagi, menurut Outbound Engine, "memperoleh pelanggan baru dapat menghabiskan biaya lima kali lebih banyak daripada mempertahankan pelanggan yang sudah ada."



Membuat model machine learning yang dapat memprediksi customer yang akan churn.

OBJECTIVE S



- Mengurangi tingkat churn customer dengan target churn rate < 8%

 Memberikan solusi dan rekomendasi bisnis
- berdasarkan insight data naik

BUSINESS METRICS



- Churn Rate
- CRC (Customer Retention Cost)



DATA OVERVIEW



5630 Baris

20 Kolom

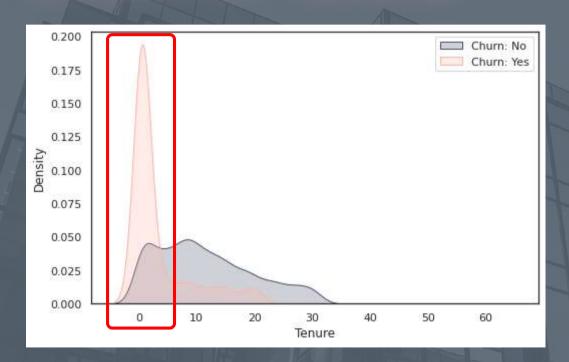


Kolom mengandung missing values

Data duplikat

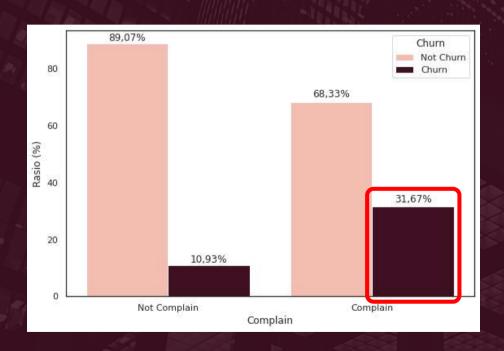
Tipe Data	Nama Feature	Deskripsi		
Target Feature	Churn	Churn flag		
Numerical	CustomerID	ID unik dari pelanggan		
	Tenure	Berapa lama customer menggunakan layanan (loyalitas) (bulan)		
	CityTier	Tingkatan kota		
	WarehouseToHome	Jarak gudang (waregouse) e commercenya ke rumah pelanggan		
	HourSpendOnApp	Waktu yang dihabiskan untuk menjelajah di aplikasi (dalam jam)		
	NumberOfDeviceRegistered	Jumlah total perangkat yang terdaftar		
	SatisfactionScore	Tingkat kepuasan pelanggan		
	DaySinceLastOrder	Jarak hari ini dengan hari terakhir/rentang waktu customer order pesan		
	NumberOfAddress	Jumlah alamat yang ditambahkan customer		
	OrderAmountHikeFromlastYear	Kenaikan Jumlah Pesanan Dari Tahun Lalu		
	CouponUsed	Jumlah total Kupon yang telah digunakan bulan lalu		
	OrderCount	Jumlah orderan bulan lalu		
	CashbackAmount	Jumlah cashback bulan lalu		
Categorical	PreferredLoginDevice	Perangkat login yang digunakan pelanggan		
	PreferredPaymentMode	Metode pembayaran yang digunakan oleh pelanggan		
	Gender	Jenis kelamin pelanggan		
	PreferedOrderCat	Kategori produk yang dibeli bulan lalu		
	MaritalStatus	Status pernikahan pelanggan		
	Complain	Complain atau tidaknya pelanggan bulan lalu		

CHURN BY TENURE

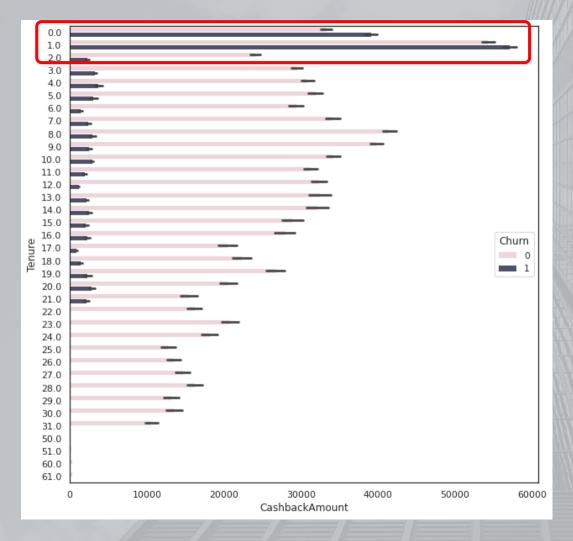


Churn cenderung terjadi pada customer dengan loyalitas yang rendah atau belum terlalu lama menggunakan layanan ecommerce.

CHURN BY COMPLAIN

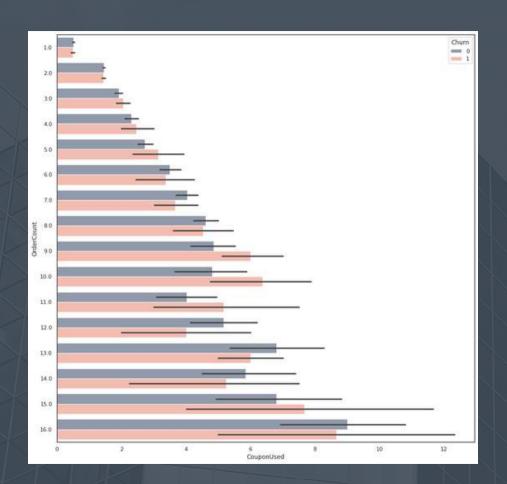


31.67% dari customer yang churn merupakan customer yang complain pada bulan lalu.



TENURE VS CASHBACK AMOUNT

Customer dengan Tenure yang rendah cenderung paling banyak churn sekaligus banyak menggunakan cashback.
Kemungkinan besar customer tersebut merupakan customer baru yang hanya ingin menikmati cashback yang disediakan oleh e-commerce saja.



ORDER COUNT VS COUPON USED

emakin banyak kupon yang diberikan, semakin banyak pula jumlah pesanan/order customer. Jadi, jumlah order produk itu sangat ditentukan oleh kupon yang kita berikan.





HANDLING MISSING VALUES

 imputasi data dengan mean & median

NO DUPLICATED DATA

HANDLING OUTLIERS

• metode Z-Score

FEATURE TRANSFORMATION

- normalization
- standardization

FEATURE ENCODING

- label encoding
- one hot encoding



FEATURE SELECTION

drop irrelevant feature

SPLIT DATA 70:30

HANDLING IMBALANCE DATA

• metode SMOTE





METRIC EVALUATION



Main Metrics: F1-Score Supporting Metrics: Recall, Precision, Accuracy, AUC

Tujuan:

- Menghindari False Negative (Customer yang dianggap tidak Churn padahal Churn).
- Menghindari False Positive (Customer yang dianggap Churn padahal tidak Churn)

Alasan:

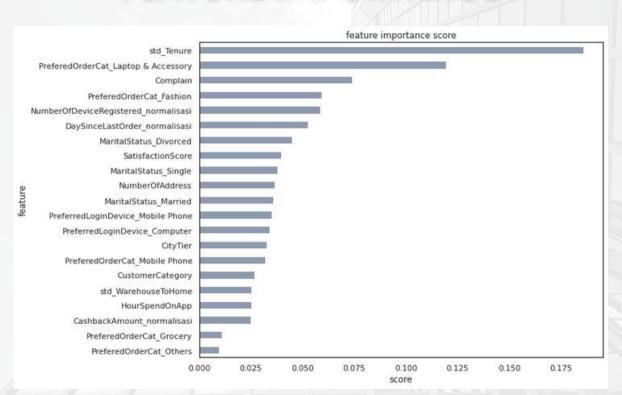
- FN besar -> terlalu banyak cost yang terbuang karena acquisition cost 5x lebih besar dari retention cost
- FP besar -> terjadi salah sasaran dalam menempatkan retention cost



MODEL EVALUATION

Model Evaluation	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	0.77	0.41	0.88	0.56	0.81
K-Nearest Neighbor	0.87	0.57	0.84	0.68	0.86
Decision Tree	0.88	0.64	0.65	0.64	0.79
Random Forest	0.95	0.87	0.84	0.86	0.91
AdaBoost	0.90	0.66	0.81	0.73	0.86
XGBoost	0.96	0.89	0.86	0.87	0.92

FEATURE IMPORTANCE



METRIC EVALUATION

Predicted to Not Churn

1310
(81,6%)

29
(1,8%)

228
(14,2%)

Actually Not Churn

Actually Churn



BEFORE MODELING

Data pelanggan Marvel Shop









Semua pelanggan diberi perlakuan yang sama yang bertujuan agar mereka tidak churn (retention)

BIAYA RETENSI SANGAT TINGGI

AFTER MODELING

Data pelanggan Marvel Shop



Melakukan prediksi untuk mengetahui pelanggan yang akan churn



Memberikan perlakuan khusus pada pelanggan yang diprediksi akan churn

BIAYA RETENSI MENJADI LEBIH RENDAH, SEHINGGA BISA DIALIHKAN **UNTUK KEPENTINGAN LAINNYA**



- Membuat fitur baru
- Memperbaiki dalam penanganan complain
- Memperbaiki strategi marketing



- Tenure naik
- Complain turun 15%
- HourSpendOnApp naik 10%
- Satisfaction Score naik 10%





HOUR SPEND ON APP

BEFORE

AFTER

15.656,8



28.496,8

Hours

NAIK 82%





TENURE

BEFORE

9,8

10,09

AFTER

NAIK 2%





TENURE

BEFORE

9,8

AFTER

>>>>

10,09

NAIK 2%



BEFORE

16,8%
Churn Rate

\$10M Retention Cost/Year **AFTER**

12,8%
Churn Rate

\$1.28M

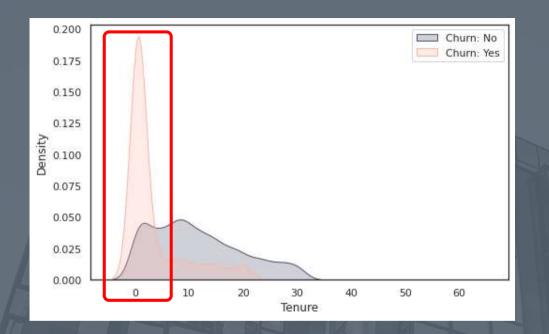
Retention Cost/Year

Assumption:

- total customer = 100.000
- Average CRC = \$100/year



RC/year= Total Customer*Average CRC*Churn rate



LOYALITAS RENDAH

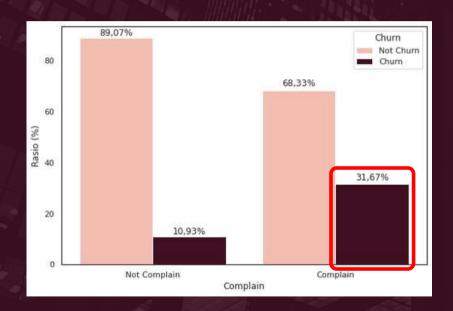
(Tenure < 9 bulan)





Membuat fitur baru dalam layanan e-commerce seperti game mengumpulkan coin/poin tersebut yang nantinya dapat ditukar dengan reward berupa **voucher diskon**, **gratis ongkir**, **dan lain sebagainya**.

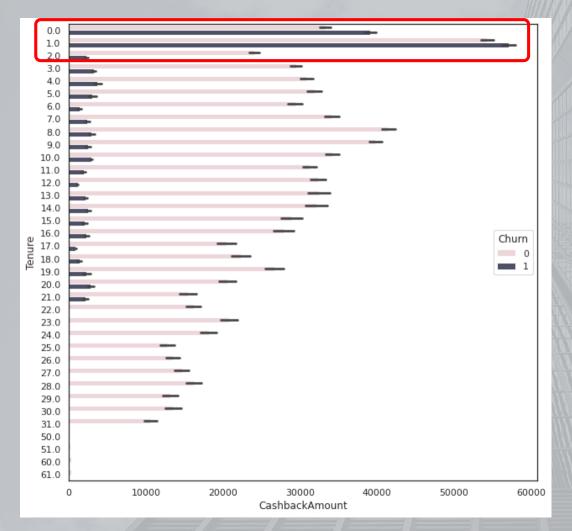
31,7% customer yang CHURN merupakan COMPLAIN





Membuat perbaikan dalam penanganan complain:

- membuat sistem bot untuk mengarahkan customer secara otomatis dan realtime
- membuat fast call service untuk penanganan masalah yang kompleks dengan cepat sekaligus sebagai upaya untuk membuat komunikasi dengan customer.



Pada tenure < 2 bulan, penggunaan

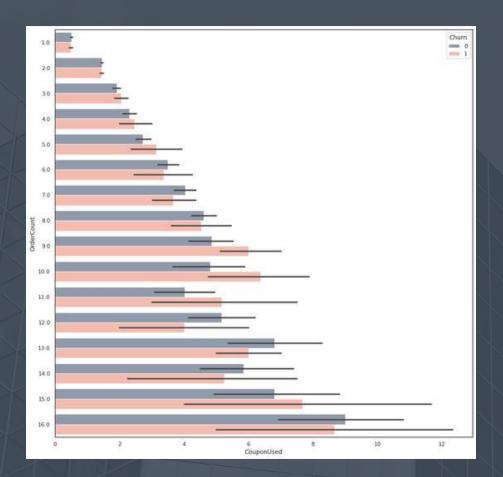
CASHBACK SANGAT TINGGI

dan customer cenderung





- Membuat klasifikasi user berdasarkan tenure agar lebih mudah menargetkan pemberian cashback terhadap customer
- Jika ingin melakukan akuisisi calon customer, bisa menggunakan strategi marketing lain yang memungkinkan new user benarbenar betah dengan layanan kita dan tidak mengincar promo saja.



Semakin banyak

KUPON yang digunakan, semakin banyak pula

JUMLAH ORDER



Jika ingin meningkatkan jumlah order, maka kita bisa lebih sering memberikan kupon kepada customer dengan mempertimbangkan ROI, BEP, dll. atau kupon diberikan kepada pelanggan berdasarkan cluster RFM.

