

RESEARCH REPORT : PREDIKSI PELANGGAN CHURN

PRESENTED BY SYAHDAN NAUFAL

DAFTAR ISI

Gambaran Umum

Pemahaman Data

EXPLORATORY DATA
ANALYTICS

1

2

3

4

5

6

ANALISIS MODEL

BISNIS
CONCLUSION

COST AND BENEFIT

ABOUT COMPANY

Lazada adalah platform e-commerce terkemuka di Asia Tenggara yang beroperasi di Indonesia, Malaysia, Filipina, Singapura, Thailand, dan Vietnam. Didirikan pada tahun 2012 dan diakuisisi mayoritas oleh Alibaba Group pada 2016, Lazada menjadi pionir dalam menghubungkan jutaan konsumen dengan berbagai penjual lokal dan internasional.

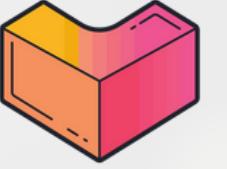
❖ Fokus Utama:

- Produk: elektronik, fashion, kecantikan, rumah tangga, kebutuhan sehari-hari.
- Layanan: pengiriman cepat, pembayaran digital, dan campaign marketing berbasis data.

🎯 Misi:

“To accelerate progress in Southeast Asia through commerce and technology.”





PERMASALAHAN BISNIS

Perusahaan LAZADA memiliki **masalah dalam mempertahankan pelanggan agar tetap aktif dan tidak berpindah ke kompetitor**. Untuk itu, diperlukan model prediksi pelanggan churn agar tim marketing dapat melakukan aksi preventif.

POINT PERMASALAHAN

- Banyak pelanggan tidak melakukan transaksi ulang → churn.
- Cukup Sulit memprediksi siapa saja yang akan churn tanpa bantuan Machine Learning



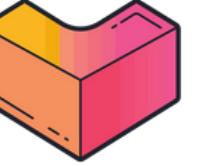
TUJUAN PROYEK

- Membangun model klasifikasi untuk memprediksi pelanggan churn.
- Mengidentifikasi fitur penting penyebab churn.
- Memberikan rekomendasi strategi retensi berbasis hasil prediksi

1

2

3



PEMAHAMAN DATA

Jumlah data

- Data : 3941 rows × 11 columns

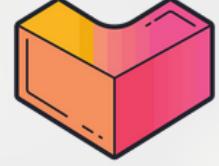
- Data Kategorikal : 2

- Data Numerik : 9

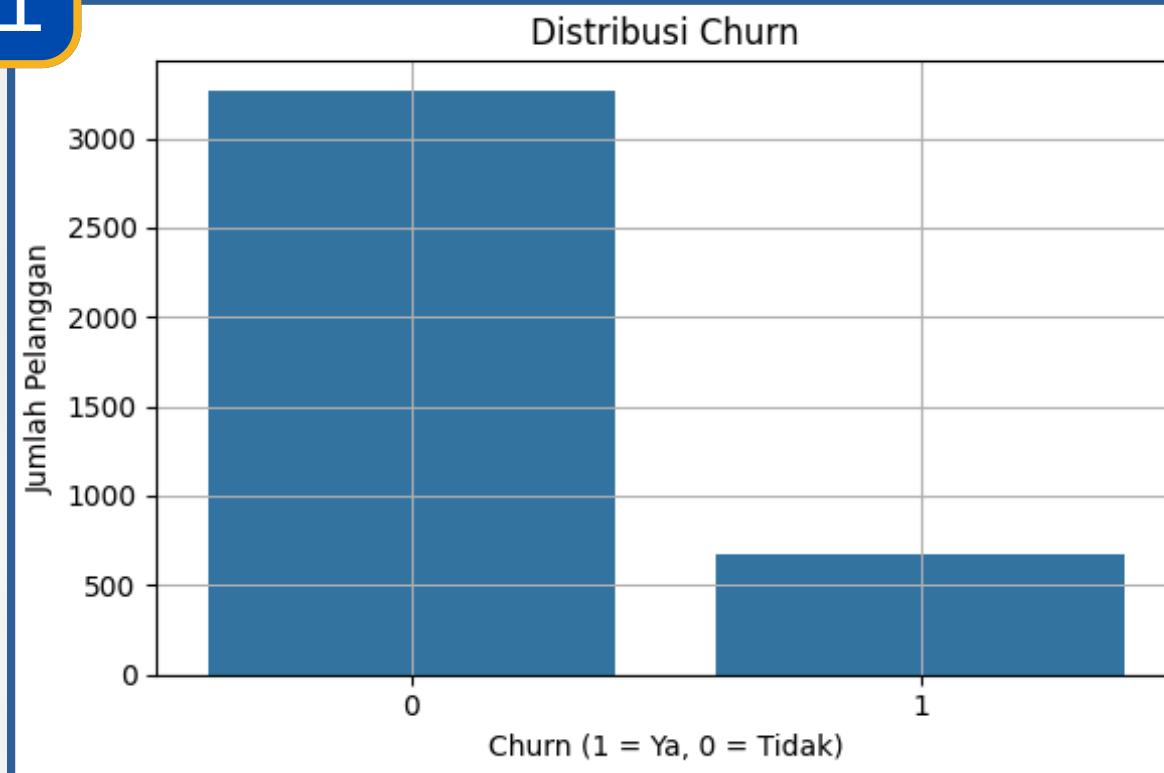
Variable Utama

- **Tenure:** Lama waktu (dalam bulan) pelanggan sudah bergabung di platform.
- WarehouseToHome: Jarak (dalam km) dari gudang ke alamat rumah pelanggan.
- NumberOfDeviceRegistered: Jumlah perangkat yang terdaftar pada akun pelanggan tersebut.
- **PreferredOrderCat:** Kategori produk yang paling sering dibeli pelanggan.
- SatisfactionScore: Skor kepuasan pelanggan terhadap layanan.
- MaritalStatus: Status pernikahan pelanggan (Single, Married, Divorced).
- NumberOfAddress: Jumlah alamat pengiriman yang tersimpan di akun pelanggan.
- **Complain:** Apakah pelanggan pernah mengajukan komplain dalam 1 bulan terakhir.
- **DaySinceLastOrder:** Jumlah hari sejak terakhir kali pelanggan melakukan pemesanan.
- **CashbackAmount:** Jumlah cashback yang diterima pelanggan pada bulan terakhir.
- **Churn:** data klasifikasi dengan metode manual

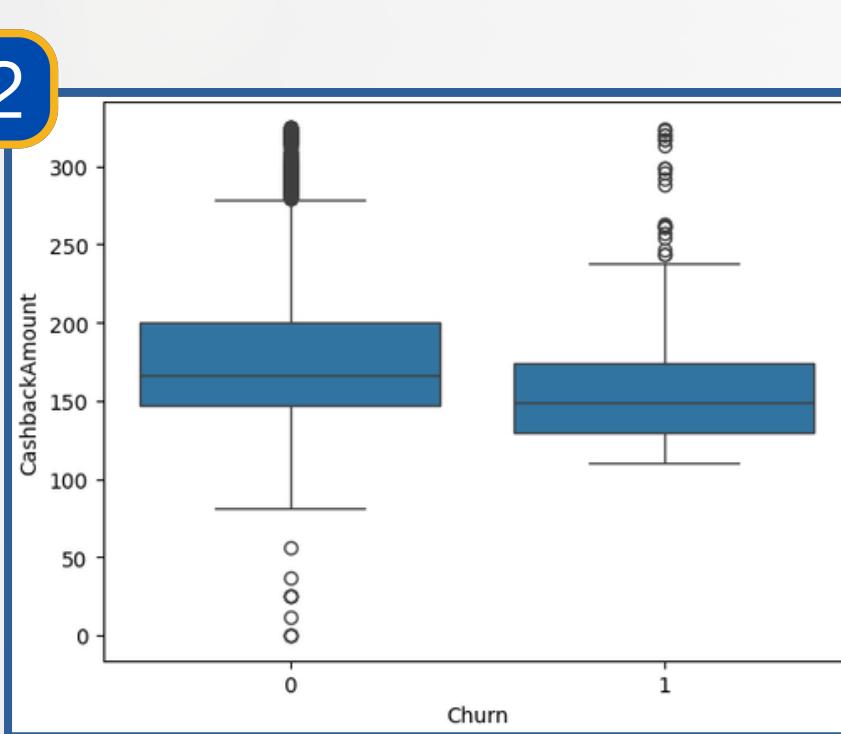
EXPLORATORY DATA ANALYST



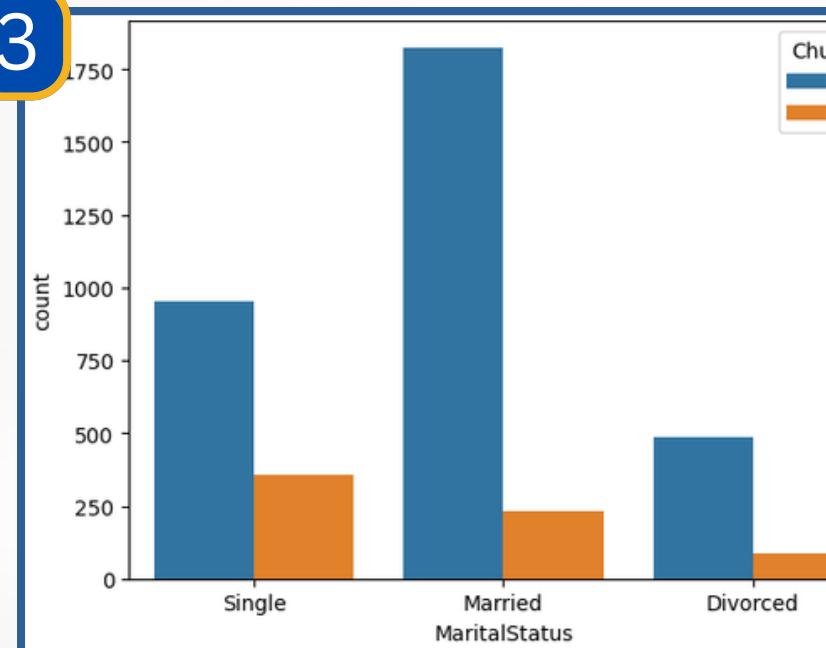
1



2



3



1. Distribusi Churn Pelanggan

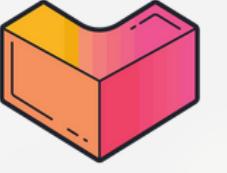
- Total pelanggan aktif (Churn = 0): 3.267 ($\approx 83\%$)
- Total pelanggan churn (Churn = 1): 674 ($\approx 17\%$)
- Terlihat data sangat imbalanced: mayoritas tetap aktif

2. Distribusi CashbackAmount vs Churn (Boxplot)

- Median cashback pelanggan aktif ≈ 165 k, pelanggan churn ≈ 145 k
- Pelanggan churn cenderung menerima cashback lebih rendah
- Outlier tinggi di kedua grup, tapi jarak IQR lebih sempit pada churn

3. Analisis MaritalStatus vs Churn (Countplot ber-hue)

- Married terbanyak secara absolut, tetapi proporsi churn relatif rendah
- Single: angka churn absolut (≈ 350) lebih tinggi dari Married (≈ 240)
- Divorced: meski jumlah lebih sedikit, persentase churn serupa



DATA TRANSFORMATION

Data Transformation information

• SimpleImputer

- Mengisi nilai hilang → mencegah error dan bias akibat missing data

• OneHotEncoder

- Ubah kategori teks → representasi numerik agar model ML dapat memprosesnya

• Pipeline

- Rangkaian transformasi → mudah reproduksi, meminimalkan kesalahan manual

• ColumnTransformer

- Terapkan pipeline berbeda di kolom yang tepat → kode lebih bersih & terstruktur

• get_feature_names_out()

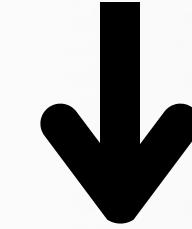
- Dapatkan nama kolom baru setelah encoding → memudahkan interpretasi & debugging

Data Sebelum Transformasi

	Tenure	WarehouseToHome	NumberOfDeviceRegistered	PreferredOrderCat	SatisfactionScore	MaritalStatus	NumberOfAddress	Complain	DaySinceLastOrder	CashbackAmount	Churn
0	15.0	29.0	4	Laptop & Accessory	3	Single	2	0	7.0	143.32	0
1	7.0	25.0	4	Mobile	1	Married	2	0	7.0	129.29	0
2	27.0	13.0	3	Laptop & Accessory	1	Married	5	0	7.0	168.54	0
3	20.0	25.0	4	Fashion	3	Divorced	7	0	Nan	230.27	0
4	30.0	15.0	4	Others	4	Single	8	0	8.0	322.17	0
...
3936	28.0	9.0	5	Fashion	3	Married	8	0	1.0	231.86	0
3937	8.0	7.0	2	Mobile Phone	2	Single	4	0	4.0	157.80	0
3938	30.0	6.0	5	Laptop & Accessory	3	Married	3	1	2.0	156.60	0
3939	6.0	Nan	4	Mobile	3	Married	10	1	0.0	124.37	1
3940	2.0	7.0	3	Laptop & Accessory	5	Married	1	0	2.0	153.73	0

3941 rows × 11 columns

3941 rows × 11 columns

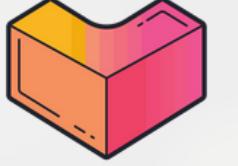


Data Setelah Transformasi

	Tenure	WarehouseToHome	NumberOfDeviceRegistered	SatisfactionScore	NumberOfAddress	Complain	DaySinceLastOrder	CashbackAmount	PreferredOrderCat_Fashion	PreferredOrderCat_Grocery
0	15.0	29.0	4.0	3.0	2.0	0.0	7.0	143.32	0.0	0.0
1	7.0	25.0	4.0	1.0	2.0	0.0	7.0	129.29	0.0	0.0
2	27.0	13.0	3.0	1.0	5.0	0.0	7.0	168.54	0.0	0.0
3	20.0	25.0	4.0	3.0	7.0	0.0	3.0	230.27	1.0	0.0
4	30.0	15.0	4.0	4.0	8.0	0.0	8.0	322.17	0.0	0.0
...
3936	28.0	9.0	5.0	3.0	8.0	0.0	1.0	231.86	1.0	0.0
3937	8.0	7.0	2.0	2.0	4.0	0.0	4.0	157.80	0.0	0.0
3938	30.0	6.0	5.0	3.0	3.0	1.0	2.0	156.60	0.0	0.0
3939	6.0	14.0	4.0	3.0	10.0	1.0	0.0	124.37	0.0	0.0
3940	2.0	7.0	3.0	5.0	1.0	0.0	2.0	153.73	0.0	0.0

3941 rows × 17 columns

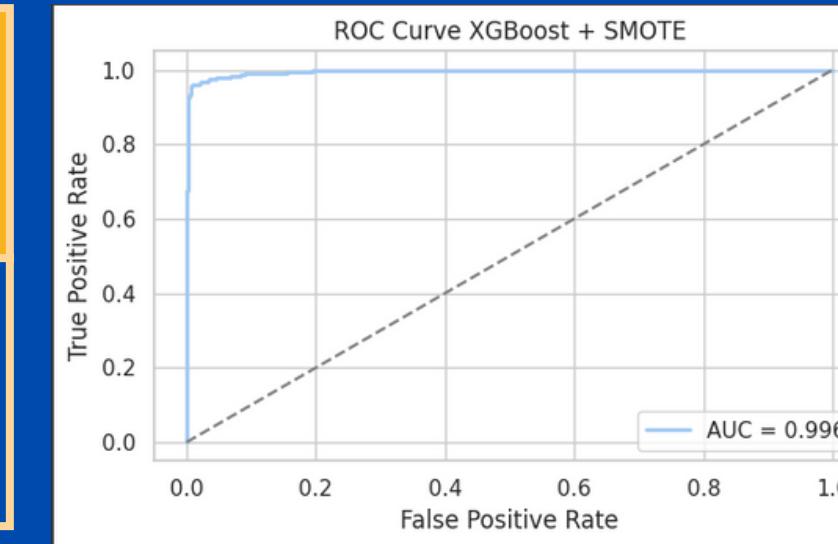
3941 rows × 17 columns



PEMODELAN & EVALUASI MATRIX

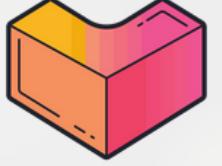
- Algoritma Utama:
 - XGBoost + SMOTE
- Data Split:
 - 80% train
 - 20% test

	accuracy	precision	recall	f1 score	ROC
1	0.97	0.97	0.98	0.97	0.996
0	0.97	0.98	0.97	0.97	

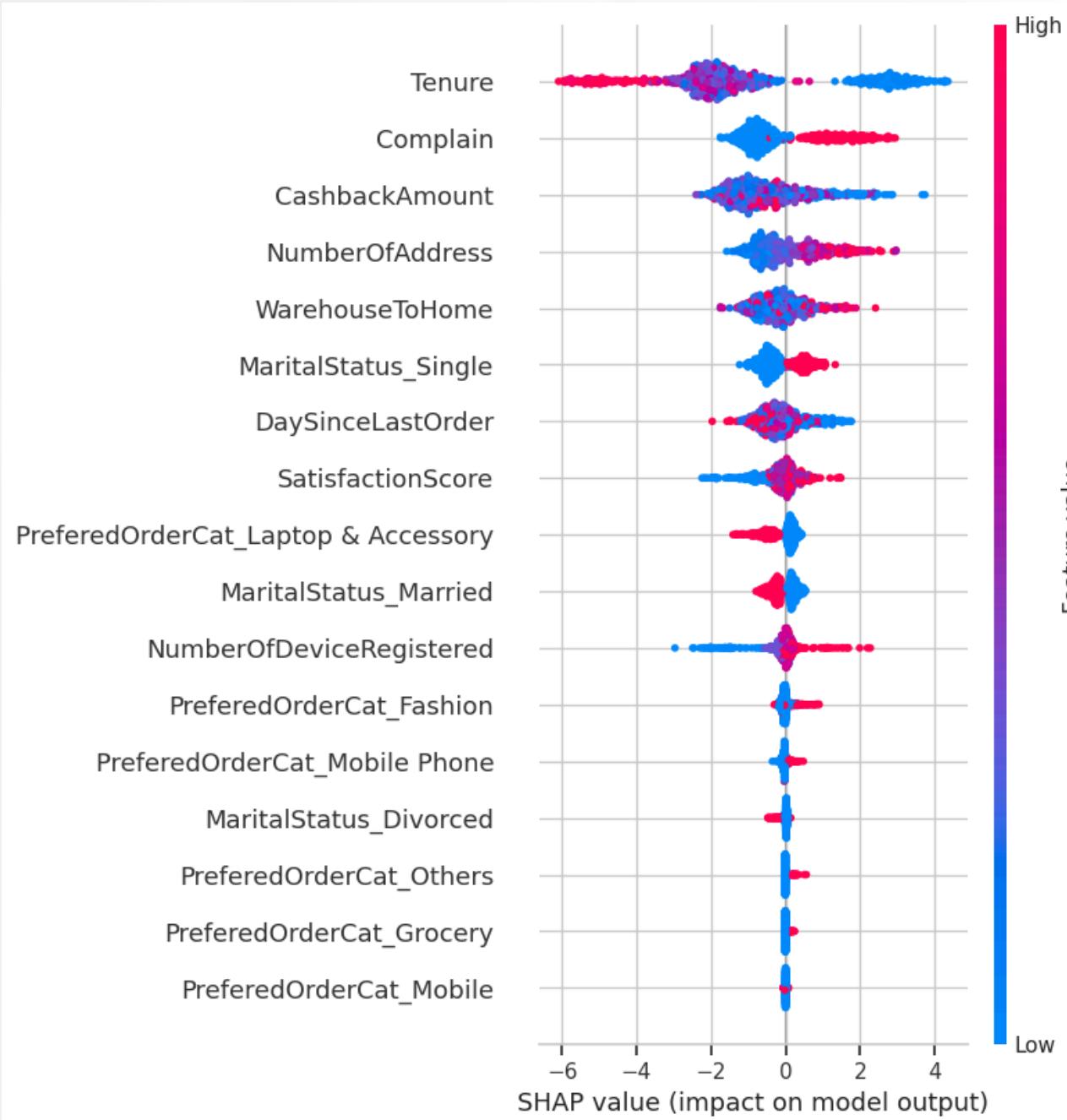


Confusion matrix

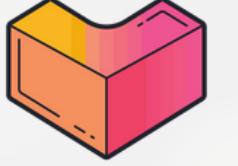
	Prediksi Loyal	Prediksi Churn
Actual Loyal	639 True Neg)	15 (False Pos)
Actual Churn	21 (False Neg)	639 (True Pos)



SHAPLEY ADDITIVE EXPLANATIONS (SHAP)



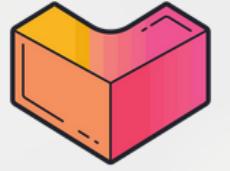
Fitur	Bagaimana Arah Pengaruhnya
Tenure (lama berlangganan)	Pelanggan baru (tenure rendah) → risiko churn lebih tinggi
Complain	Pernah komplain → risiko churn lebih tinggi
CashbackAmount	Cashback kecil → risiko churn lebih tinggi
DaySinceLastOrder	Semakin lama tidak order → risiko churn lebih tinggi
WarehouseToHome (jarak kirim)	Jarak lebih jauh → risiko churn lebih tinggi



RECOMMENDATIONS PLANNING

No	Recommendation	Target Segment	Expected Impact
1	Onboarding Booster	Pelanggan baru (Tenure < 3 bln)	+30% engagement bulan pertama → turunkan early churn
2	Proaktif Follow-Up Complain	Pelanggan yang komplain	50% pengurangan churn komplain → jaga CSAT tinggi
3	Dynamic Cashback Campaign	CashbackAmount rendah	+20% repeat order → tingkatkan loyalitas
4	Re-engagement Reminder	DaySinceLastOrder > 30 hari	+15% re-order → hentikan “silent churn”
5	Optimasi Logistik	WarehouseToHome jauh	Percepat delivery → +10% retention pelanggan

COST BENEFIT



COST :

Kategori	Komponen
Direct Costs	<ul style="list-style-type: none">Pengembangan & training model MLLisensi cloud + infra (pay-as-you-go)
Indirect Costs	<ul style="list-style-type: none">Waktu tim data & marketingPelatihan pengguna baru
Intangible Costs	<ul style="list-style-type: none">Risiko minoritas kesalahan prediksi (False Positives)Adaptasi proses marketing

BENEFIT :

Kategori	Komponen
Direct Benefits	<ul style="list-style-type: none">Pelanggan churn dicegah: $108 \text{ orang} \times 1.5 \text{ Jt/bln} = 162 \text{ Jt/bln}$Penghematan biaya akuisisi
Indirect Benefits	<ul style="list-style-type: none">Efisiensi manual analisis churn
Intangible Benefits	<ul style="list-style-type: none">Peningkatan CSAT & brand loyaltyCompetitive edge lewat data-driven marketing



THANK YOU