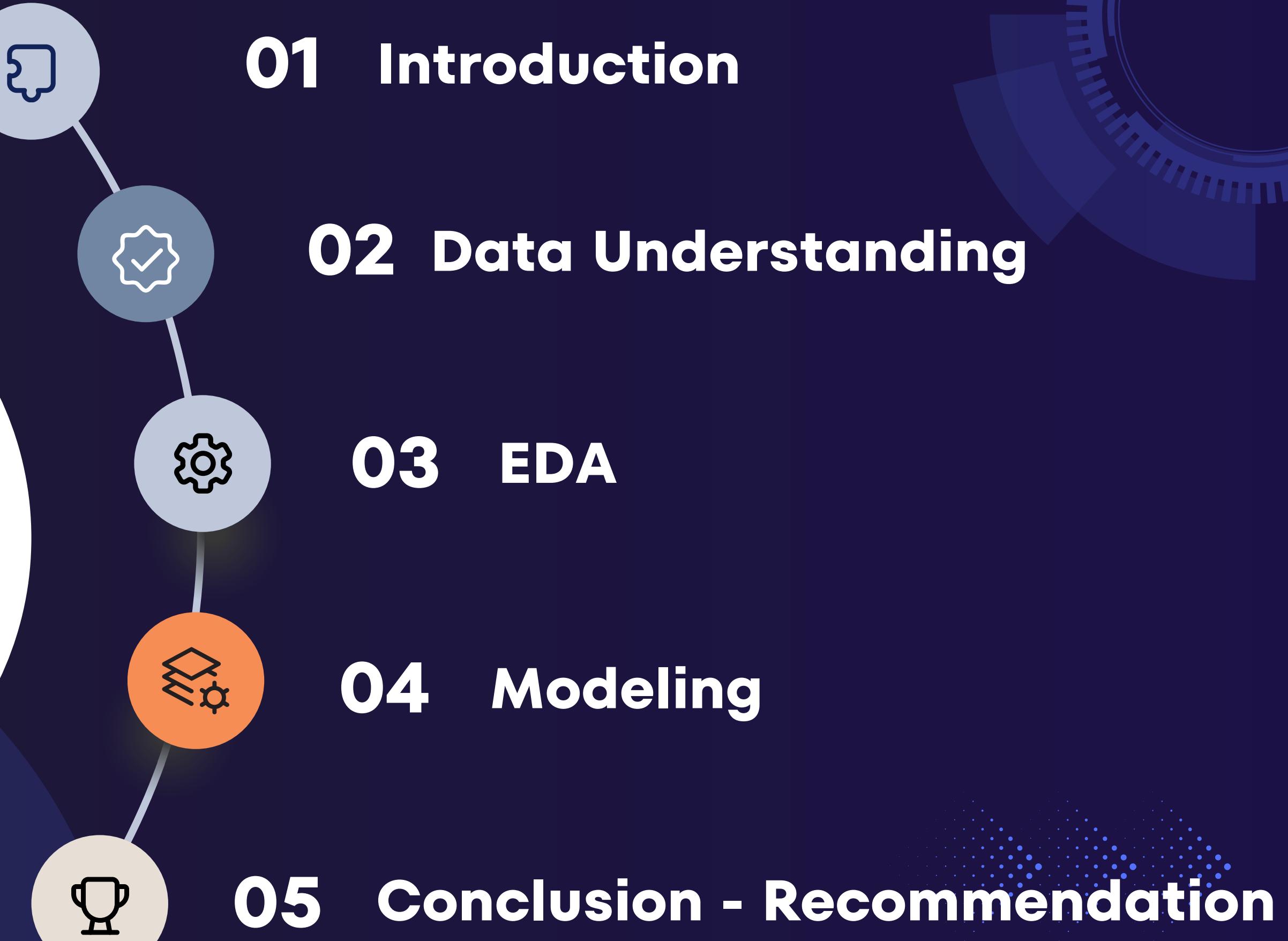


# Implementasi Machine Learning: Prediksi Churn pada Perusahaan Telekomunikasi

Presented by **Fatimah Azzahra**



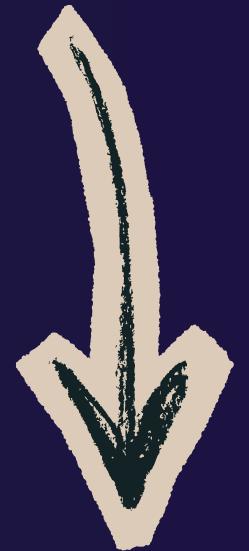
# Presentation Flow



# Executive Summary

- perusahaan menghadapi tantangan di mana pelanggan memiliki banyak pilihan alternatif
- customer retention > mengakuisisi pelanggan baru.
- Fenomena \*churn\* di mana pelanggan berhenti menggunakan layanan atau produk dapat menurunkan profit perusahaan secara signifikan dan menurunkan \*Customer Lifetime Value\* (CLV).\\
- pola perilaku pelanggan dan mendekripsi sinyal ketidakpuasan sejak dulu kebutuhan strategis untuk keberlanjutan bisnis jangka panjang.

**Predict Churn Menggunakan Machine Learning dengan Model Logistic Regression with Random Under Sampler after Hyperparameter Tuning**



Hasil predict dijadikan acuan untuk pengambilan keputusan agar semakin tepat sasaran

# Introduction



# Latar Belakang

- Perusahaan menghadapi tantangan di mana pelanggan memiliki banyak pilihan alternatif
- Customer retention > Akuisisi pelanggan baru
- Fenomena *churn* di mana pelanggan berhenti menggunakan layanan atau produk dapat menurunkan profit perusahaan secara signifikan dan menurunkan *Customer Lifetime Value*
- Pola perilaku pelanggan dan mendeteksi sinyal ketidakpuasan sejak dini kebutuhan strategis untuk keberlanjutan bisnis jangka panjang.

## Bussiness Problem

- Identifikasi reaktif
- Revenue Leakage
- Ketidakefektifan Strategi Retensi

## Tujuan

- Membangun model prediktif
- Identifikasi faktor utama Churn
- Rekomendasi Strategis kepada Perusahaan

# Pendekatan Analitis

1. Data Understanding
2. EDA
3. Data Preparation
4. Modeling
5. Interpretation Model

# Stakeholder

1. Customer Retention Team → berhubungan dengan pelanggan yang churn
2. Customer Relationship Team → berhubungan dengan pelanggan yang tidak churn
3. Marketing Team → berhubungan dengan strategi pemberian campaign yang tepat sasaran
4. Chief Marketing Officer → berhubungan dengan profit dan loss yang didapatkan perusahaan
5. Investor / Pemilik perusahaan → berhubungan dengan arah perusahaan kedepannya.

# Metrik Evaluasi

- **False Negative (FN)**: Model memprediksi pelanggan "Setia", padahal aslinya "Churn". Perusahaan kehilangan pelanggan berharga tanpa sempat melakukan pencegahan. Ini adalah kerugian finansial terbesar (High Cost) berupa hilangnya Customer Lifetime Value (CLV).
- **False Positive (FP)**: Model memprediksi pelanggan "Churn", padahal aslinya "Setia". Perusahaan memberikan promosi retensi kepada pelanggan yang sebenarnya tidak bermatiat pergi. Biayanya relatif kecil, hanya sebatas biaya promosi/diskon (Low Cost).
- Berdasarkan pertimbangan biaya di atas, prioritas Stakeholder adalah meminimalisir kehilangan pelanggan (FN). Maka, metrik utama yang digunakan adalah:
- 
- **PR AUC (Precision–Recall Area Under Curve)** : Digunakan sebagai metrik utama karena mampu mengevaluasi trade-off antara precision dan recall secara menyeluruh pada berbagai nilai threshold, serta lebih representatif untuk imbalanced data
- **Recall (Sensitivity)**: Fokus utama untuk menangkap sebanyak mungkin pelanggan yang berpotensi churn. Semakin tinggi Recall, semakin sedikit pelanggan churn yang lolos dari deteksi.
- **Precision**: Untuk mengontrol tingkat kesalahan False Positive agar biaya promosi tetap efisien.

# Data Understanding



# DATA DICTIONARY

## Fitur

### Informasi Customer

<b>tenure</b>	Durasi (bulan) berlangganan
<b>Contract</b>	Jenis kontrak yang digunakan pelanggan
<b>PaperlessBilling</b>	Jenis Tagihan yang diterima pelanggan
<b>MonthlyCharges</b>	Biaya bulanan yang dikeluarkan pelanggan
<b>Dependants</b>	Tanggungan yang ikut menggunakan layanan

 Object

 Float

 Int

### Layanan yang digunakan

<b>InternetService</b>	Jenis koneksi internet yang digunakan pelanggan
<b>OnlineSecurity</b>	Pelanggan berlangganan layanan keamanan online
<b>OnlineBackup</b>	Pelanggan berlangganan layanan penyimpanan cadangan data
<b>DeviceProtection</b>	Pelanggan berlangganan layanan asuransi perangkat yang sedang digunakan
<b>TechSupport</b>	Pelanggan berlangganan layanan prioritas bantuan teknis

**4930 Baris Data**  
**11 Kolom Data**

### Target

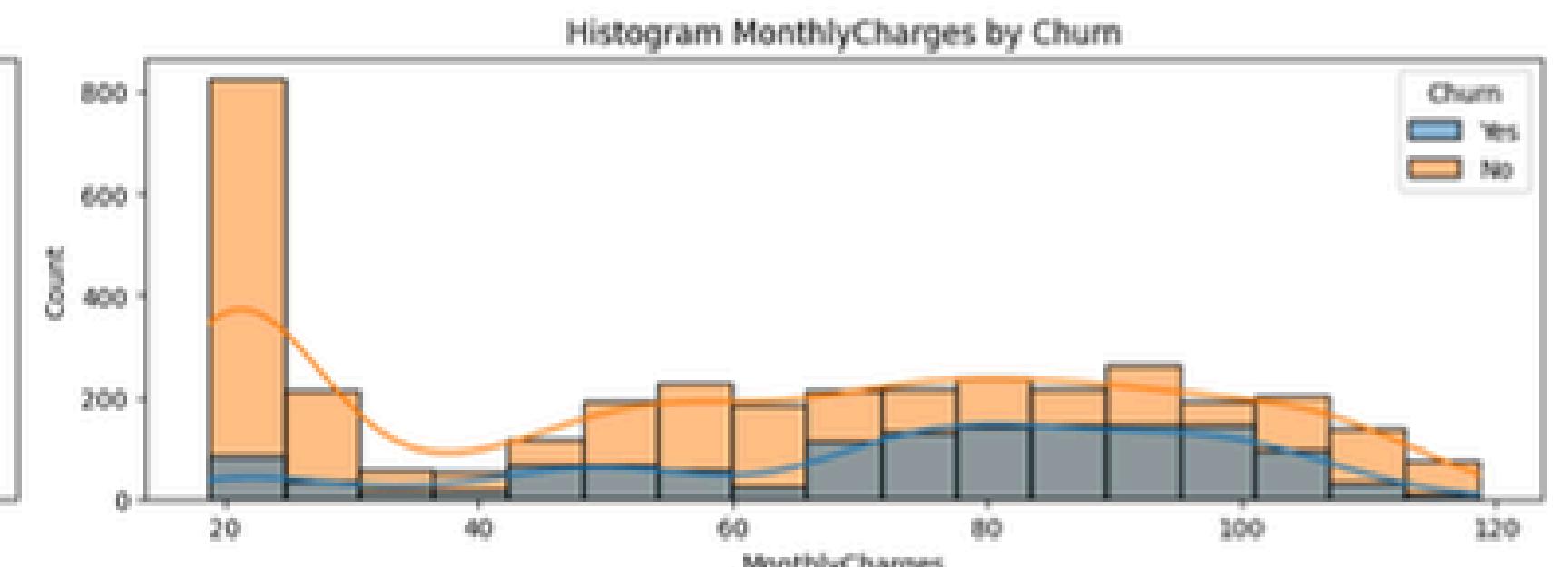
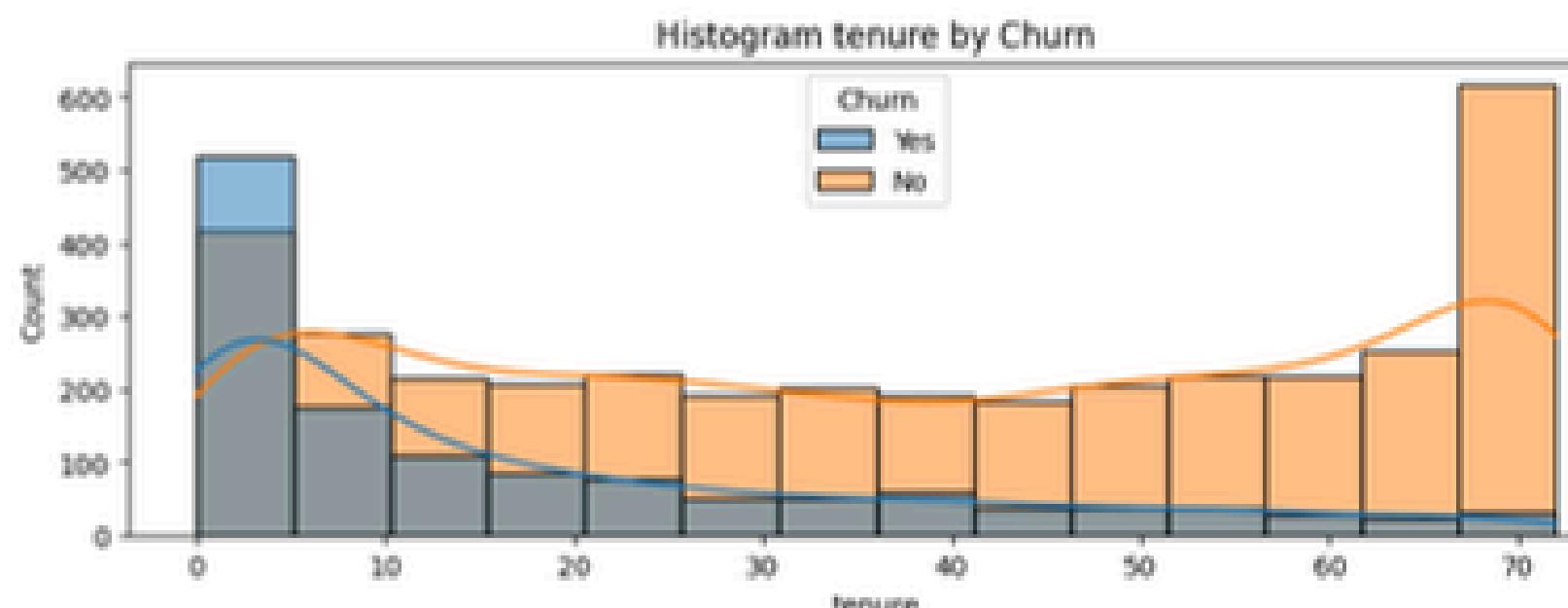
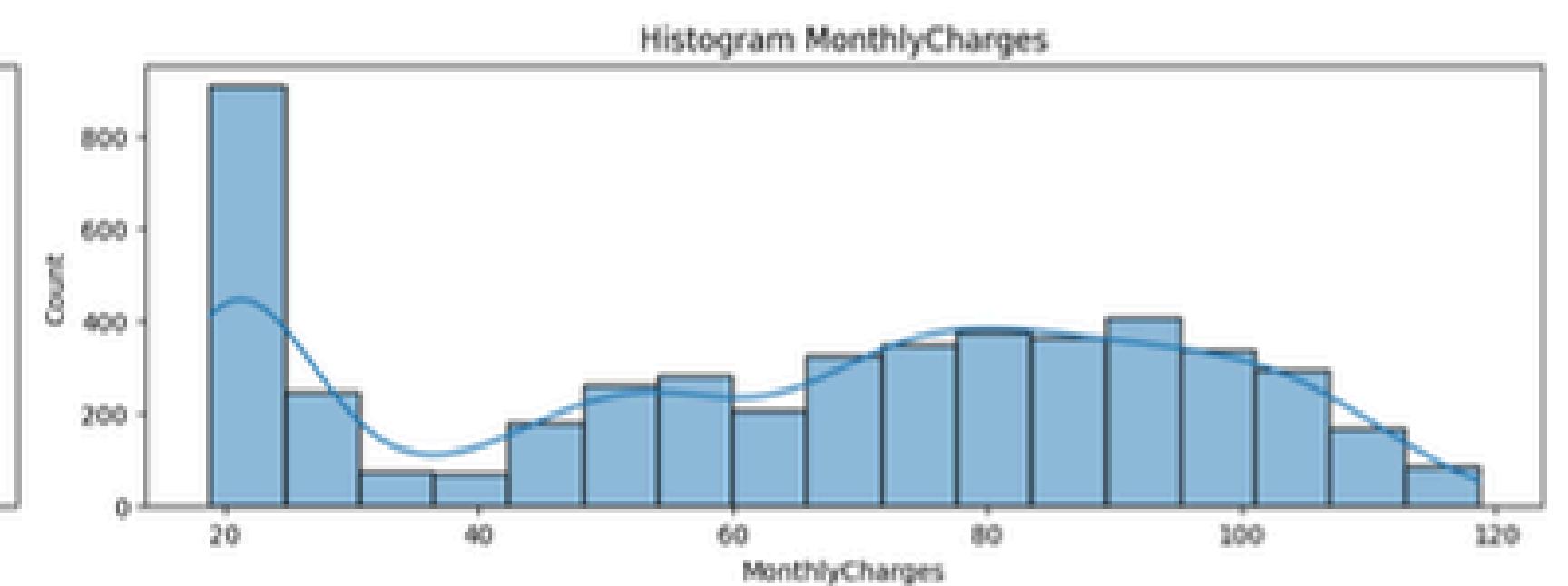
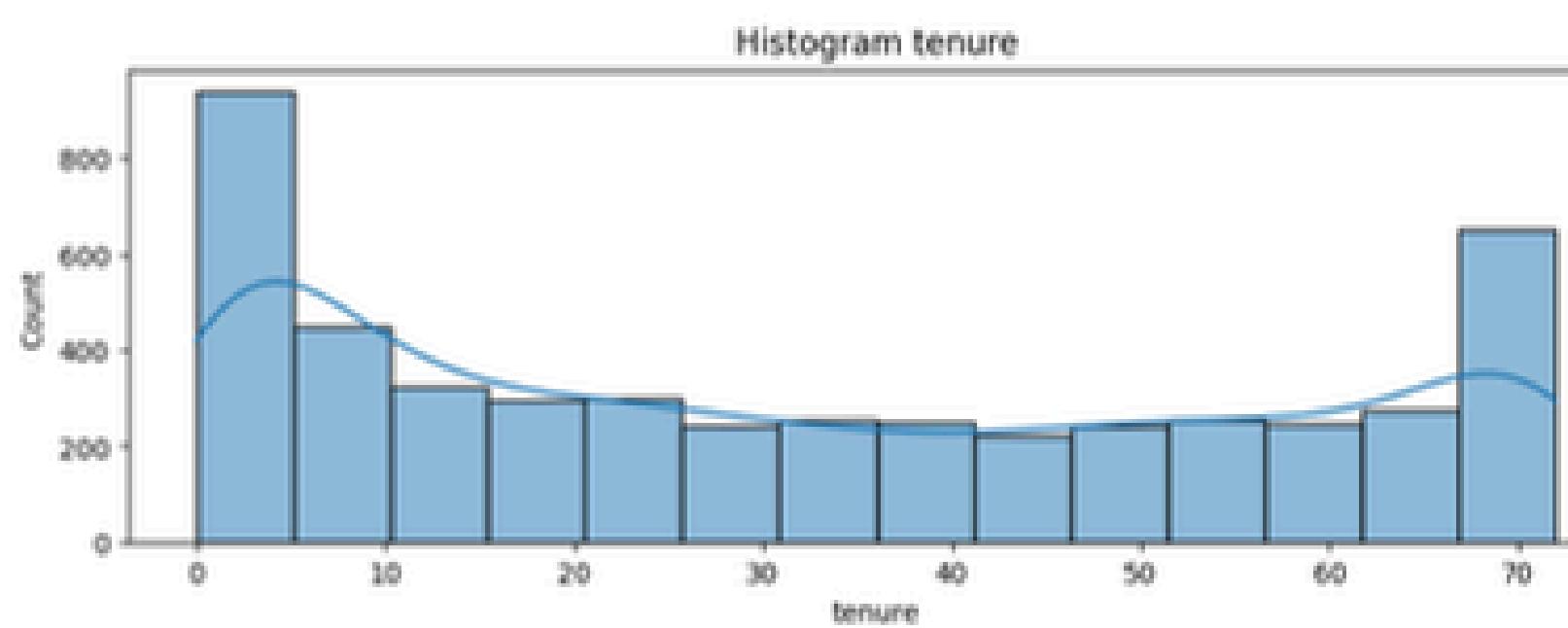
Churn	Status pelanggan (berhenti berlangganan atau tidak)
-------	---

**Missing Value**  
**77 Data Duplikat**

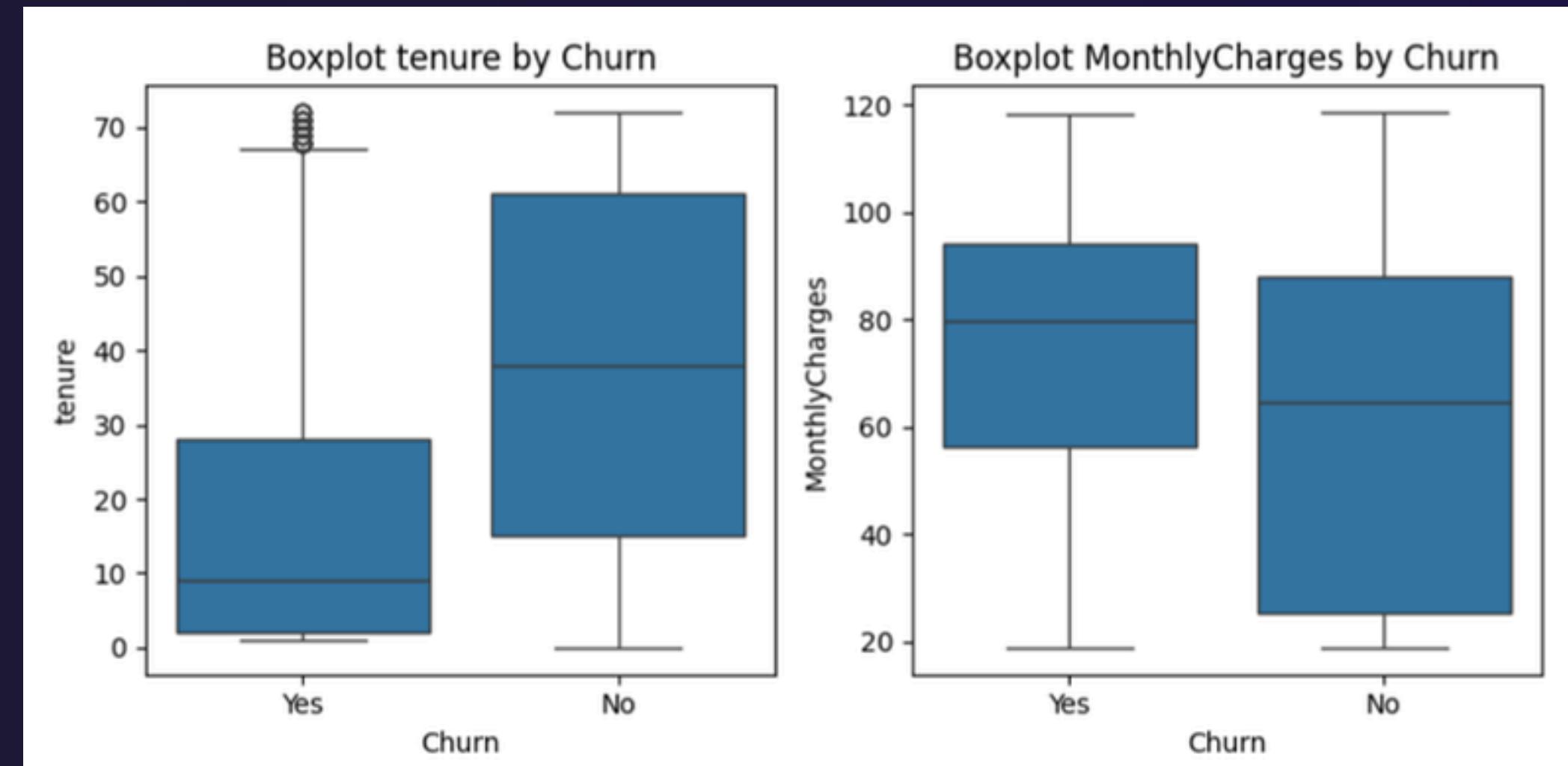
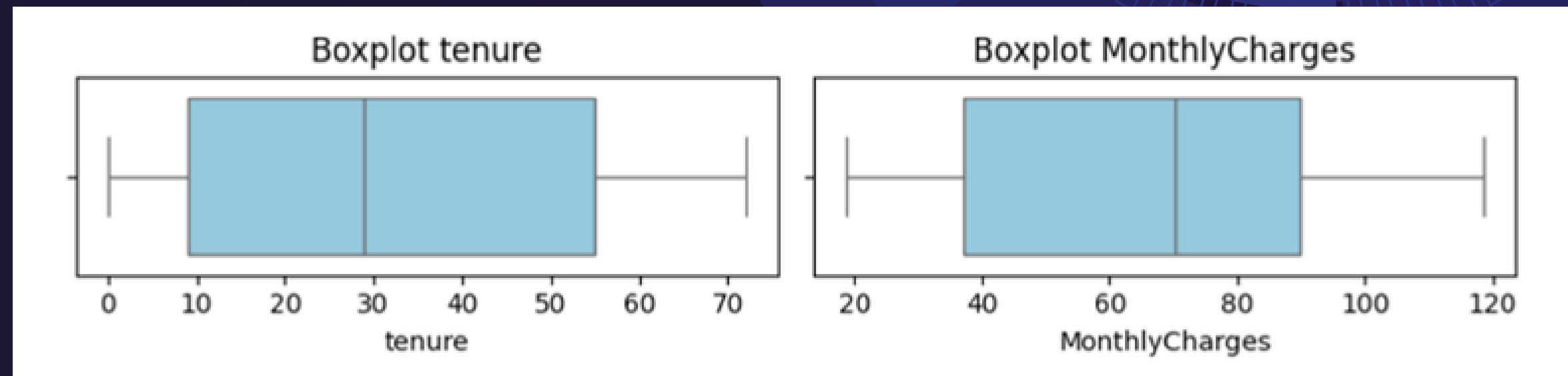
# Exploratory Data Analysis



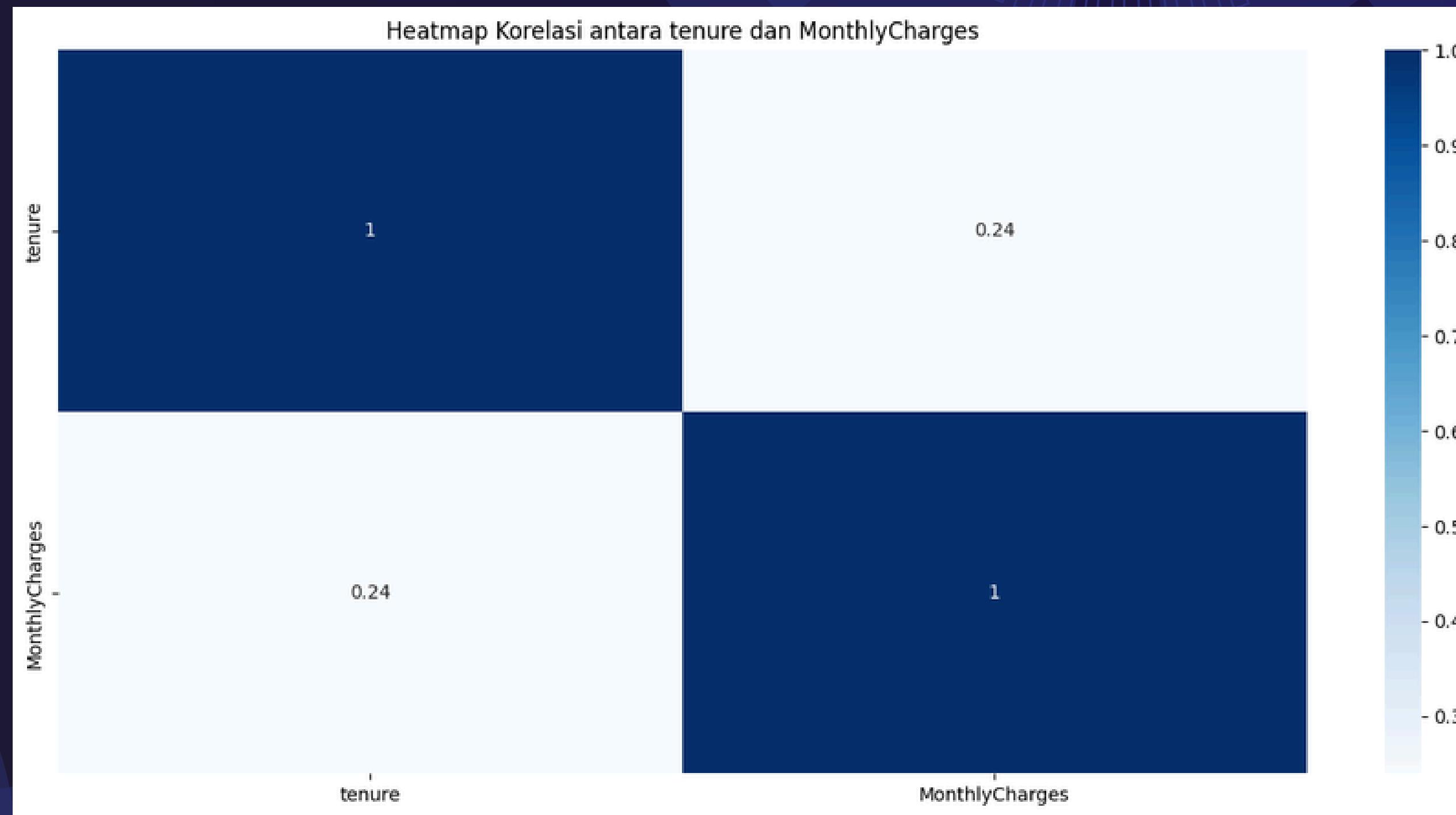
# Histogram Data Numerik



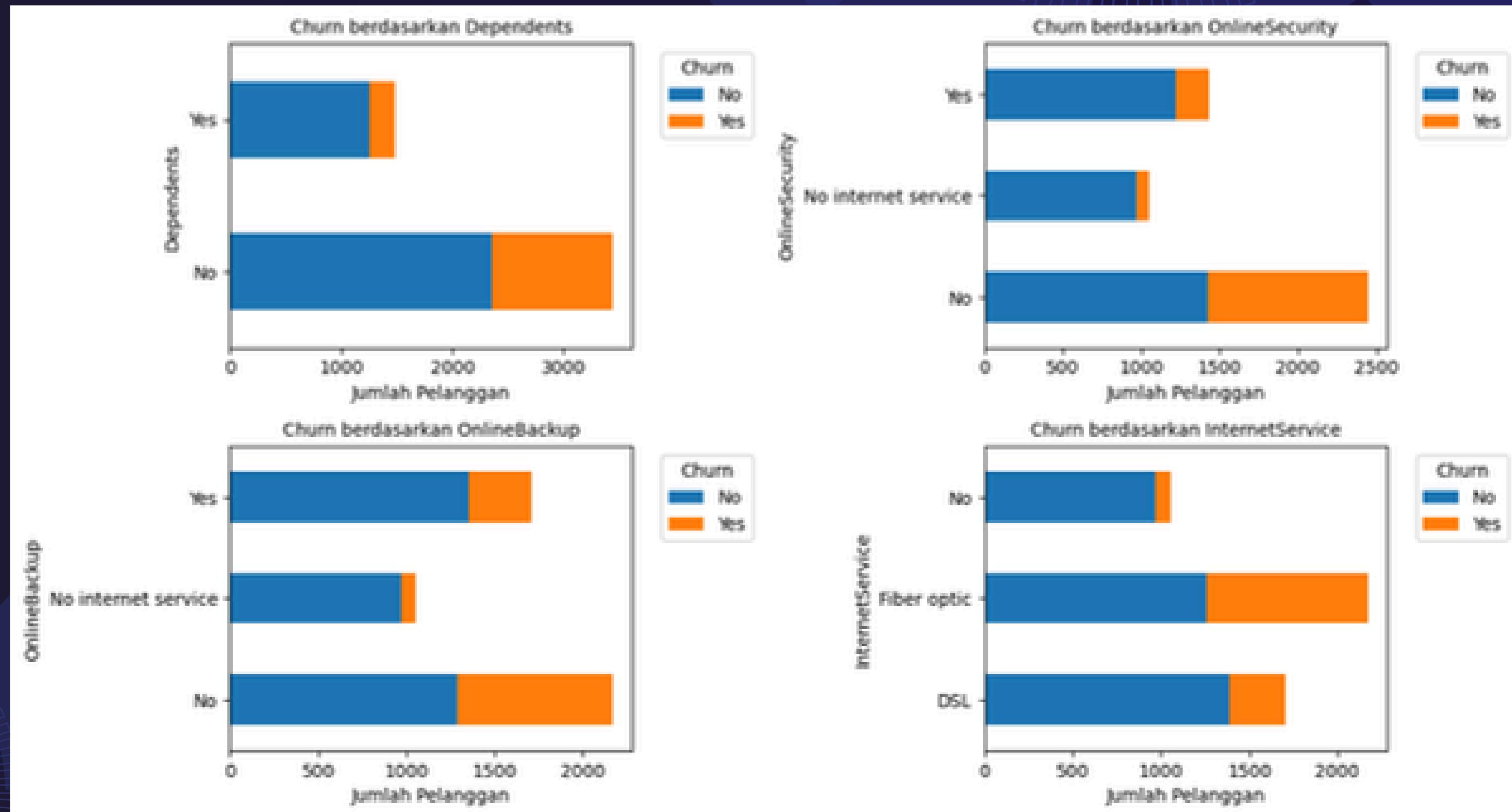
# Outlier Data Numerik



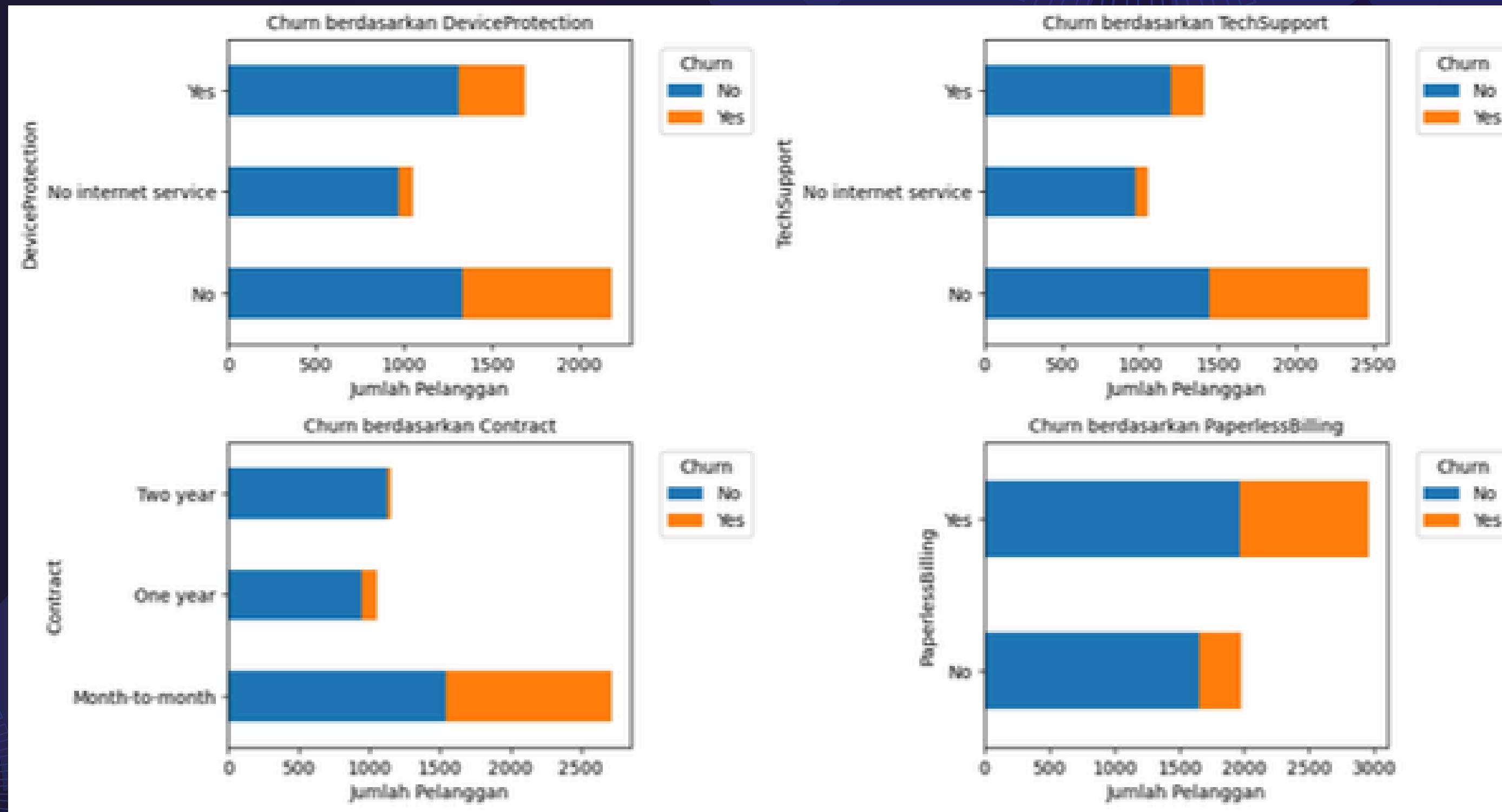
# Correlation Data Numerik



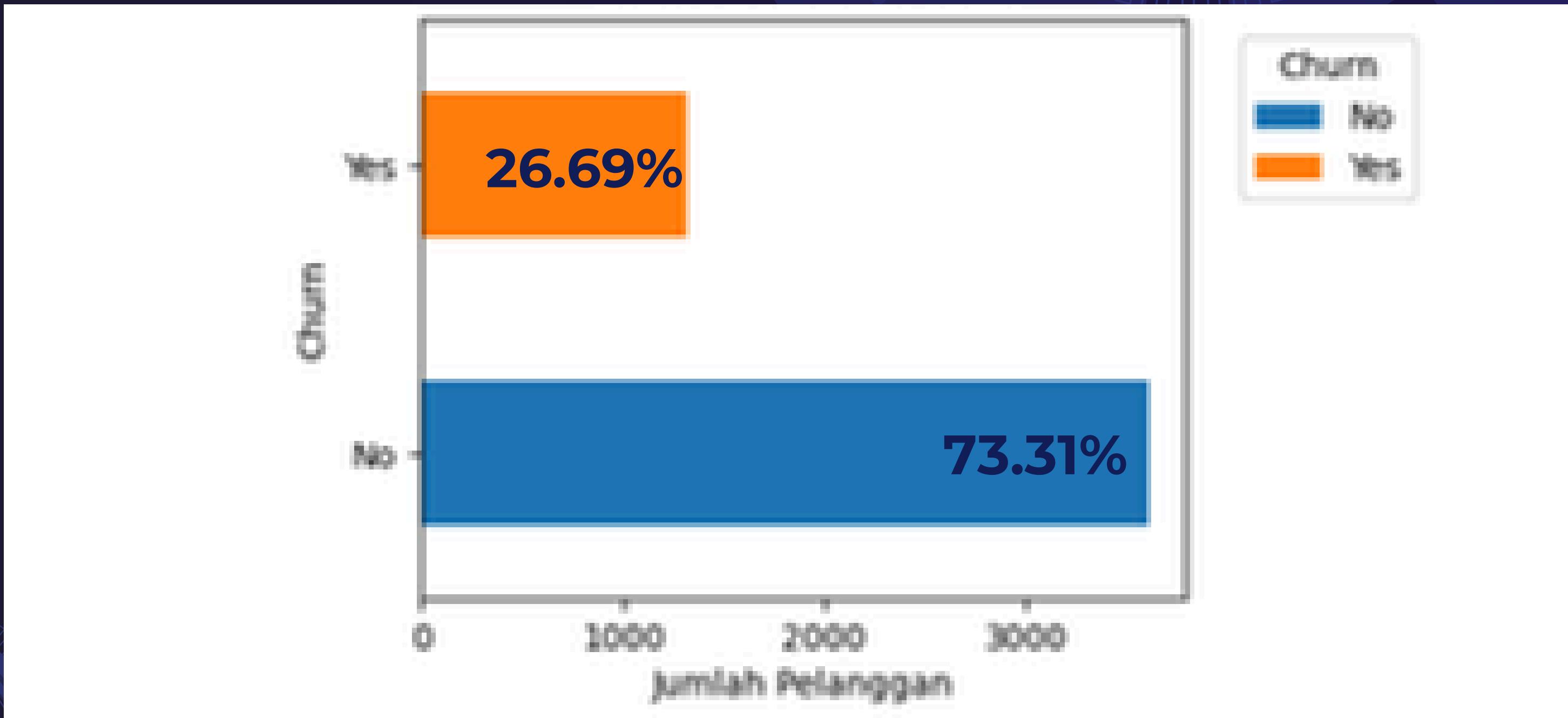
# Bar Chart Data Kategorik



# Bar Chart Data Kategorik



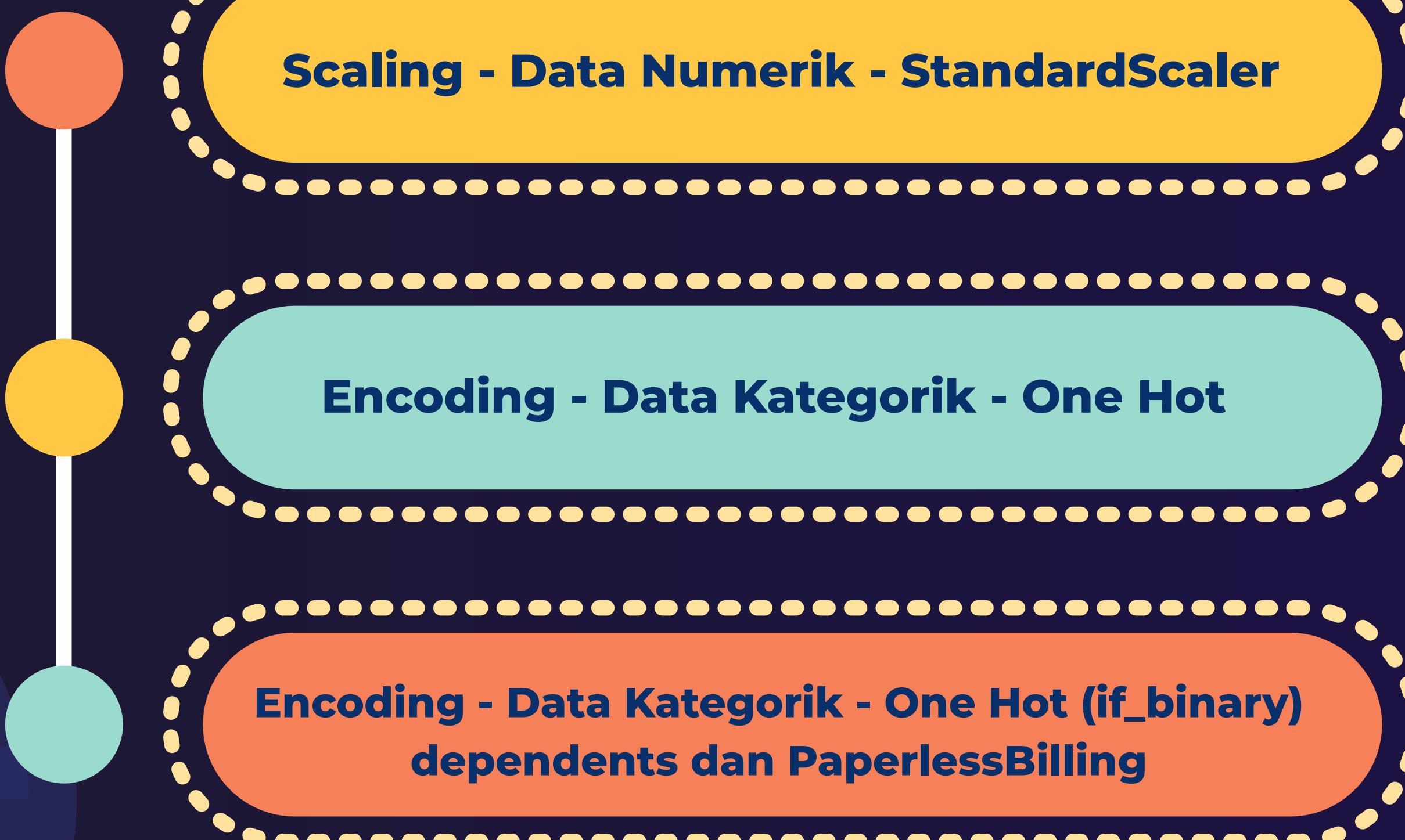
# Imbalanced Data



# Data Preparation



# Feature Engineering



Scaling - Data Numerik - StandardScaler

Encoding - Data Kategorik - One Hot

Encoding - Data Kategorik - One Hot (if\_binary)  
dependents dan PaperlessBilling

# Modeling



# Base Model Summary

	Model	PR_AUC (cv)	PR_AUC_STD	Training Time (s)
0	Logistic Regression	0.647814	0.025108	3.449919
1	AdaBoost	0.646129	0.034327	1.000018
2	Gradient Boosting	0.643732	0.030583	2.173739
3	LightGBM	0.622054	0.012070	3.682949
4	XGBoost	0.605906	0.023789	1.086210
5	Random Forest	0.602678	0.005808	2.181738
6	KNN	0.551059	0.017385	0.319065
7	Decision Tree	0.372715	0.013240	0.281545

# Base Model Summary

		Logistic Regression				
		precision	recall	f1-score	support	
KNN		0	0.85	0.88	0.86	904
		1	0.64	0.57	0.60	329
Decision Tree		precision	recall	f1-score	support	
		0	0.82	0.86	0.84	904
		1	0.57	0.49	0.53	329
Random Forest		precision	recall	f1-score	support	
		0	0.80	0.81	0.80	904
		1	0.46	0.46	0.46	329
		precision	recall	f1-score	support	
		0	0.82	0.88	0.85	904
		1	0.59	0.49	0.53	329

		Gradient Boosting				
		precision	recall	f1-score	support	
AdaBoost		0	0.84	0.89	0.87	904
		1	0.64	0.54	0.59	329
LightGBM		precision	recall	f1-score	support	
		0	0.85	0.89	0.87	904
		1	0.66	0.57	0.61	329
		precision	recall	f1-score	support	
		0	0.85	0.88	0.87	904
		1	0.64	0.57	0.60	329

# Logistic Regression Model Summary

	Model	PR_AUC (CV)	PR_AUC_STD	Training Time (s)
0	Random Under Sampling	0.648809	0.040552	0.272715
1	Logistic Regression (Penalized)	0.647816	0.026241	0.263020
2	Baseline (No Resampling)	0.646809	0.044358	0.285963
3	SMOTE	0.644983	0.043753	0.293461
4	Random Over Sampling	0.643672	0.039778	0.729877

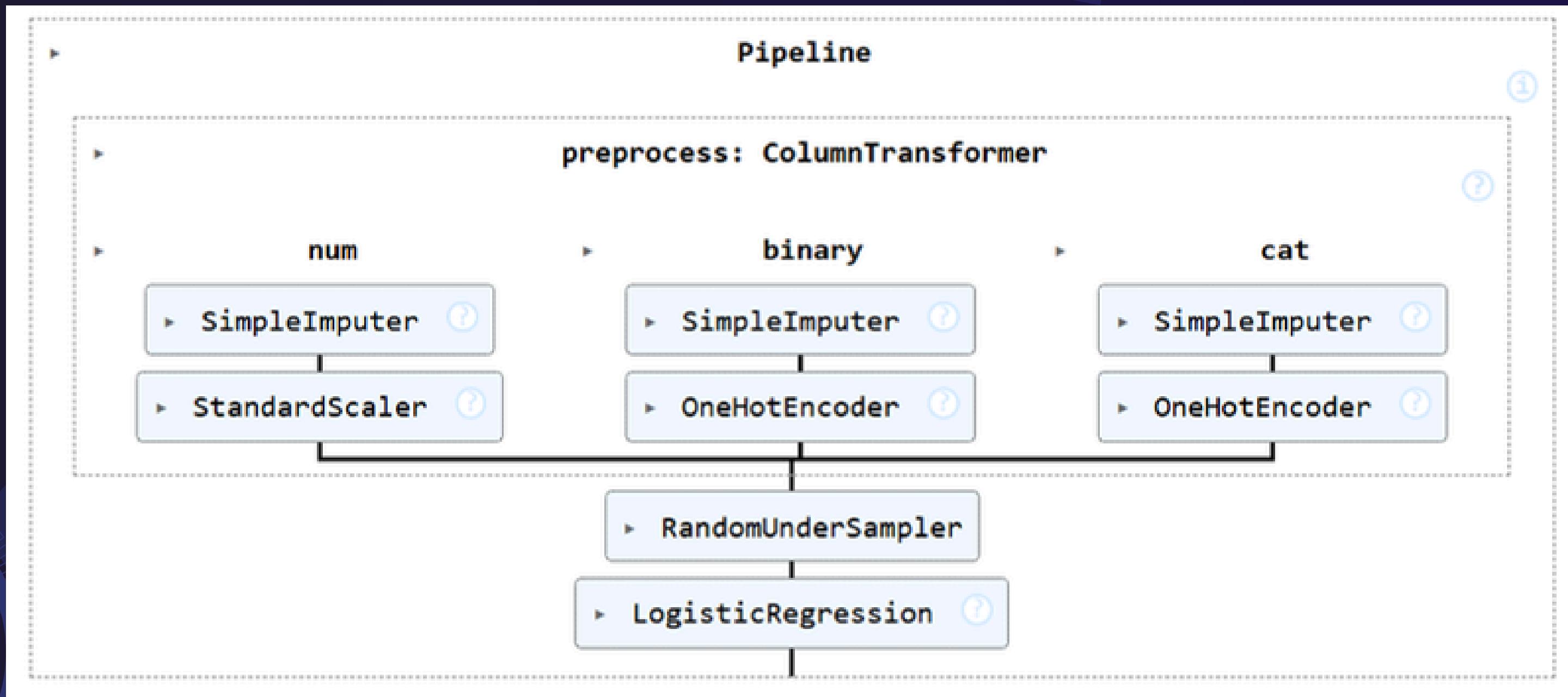
# Hyperparameter Tuning

Best Parameters:

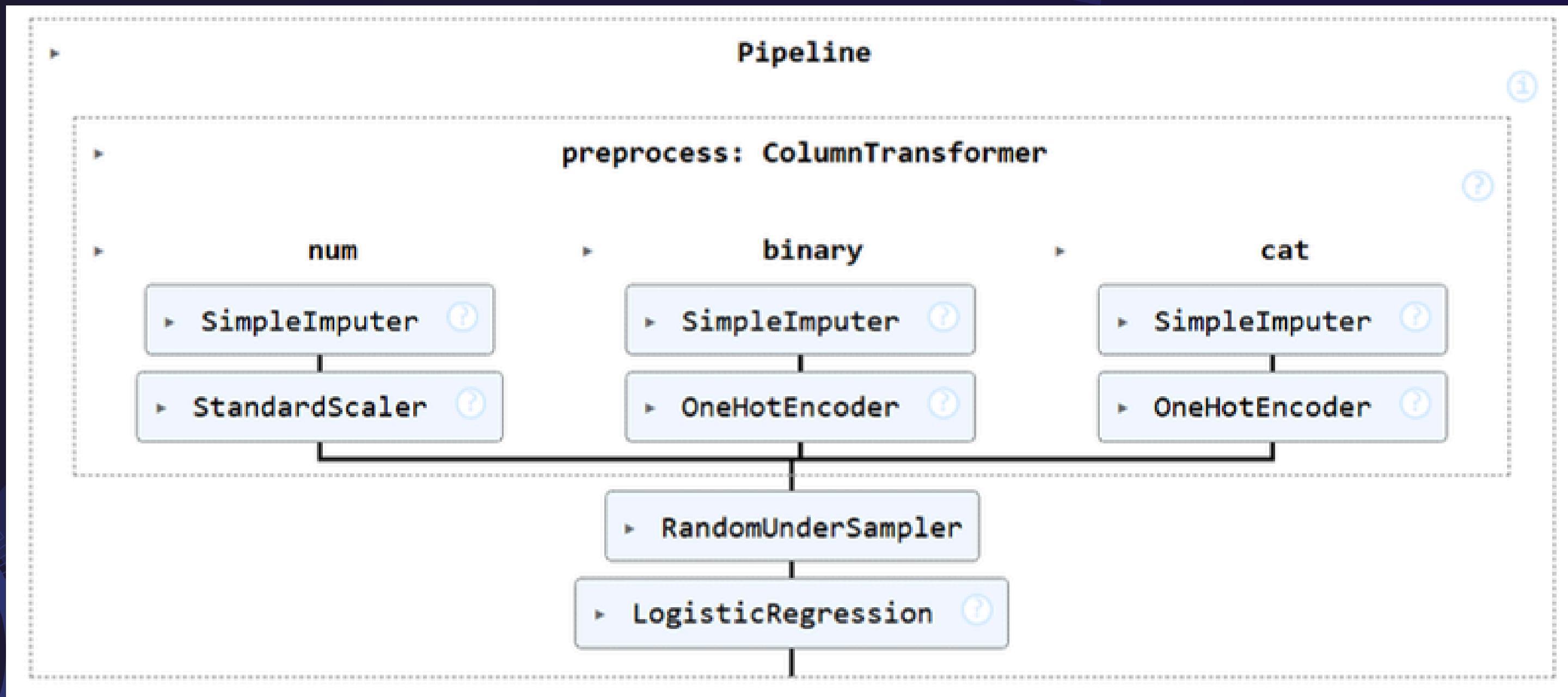
```
{'model__solver': 'lbfgs',
'model__penalty': 'l2',
'model__class_weight': None,
'model__C': 0.1}
```

	Model	PR_AUC (cv)
0	Before Tuning (RUS + LogReg)	0.648809
1	After Tuning (RUS + LogReg)	0.649079

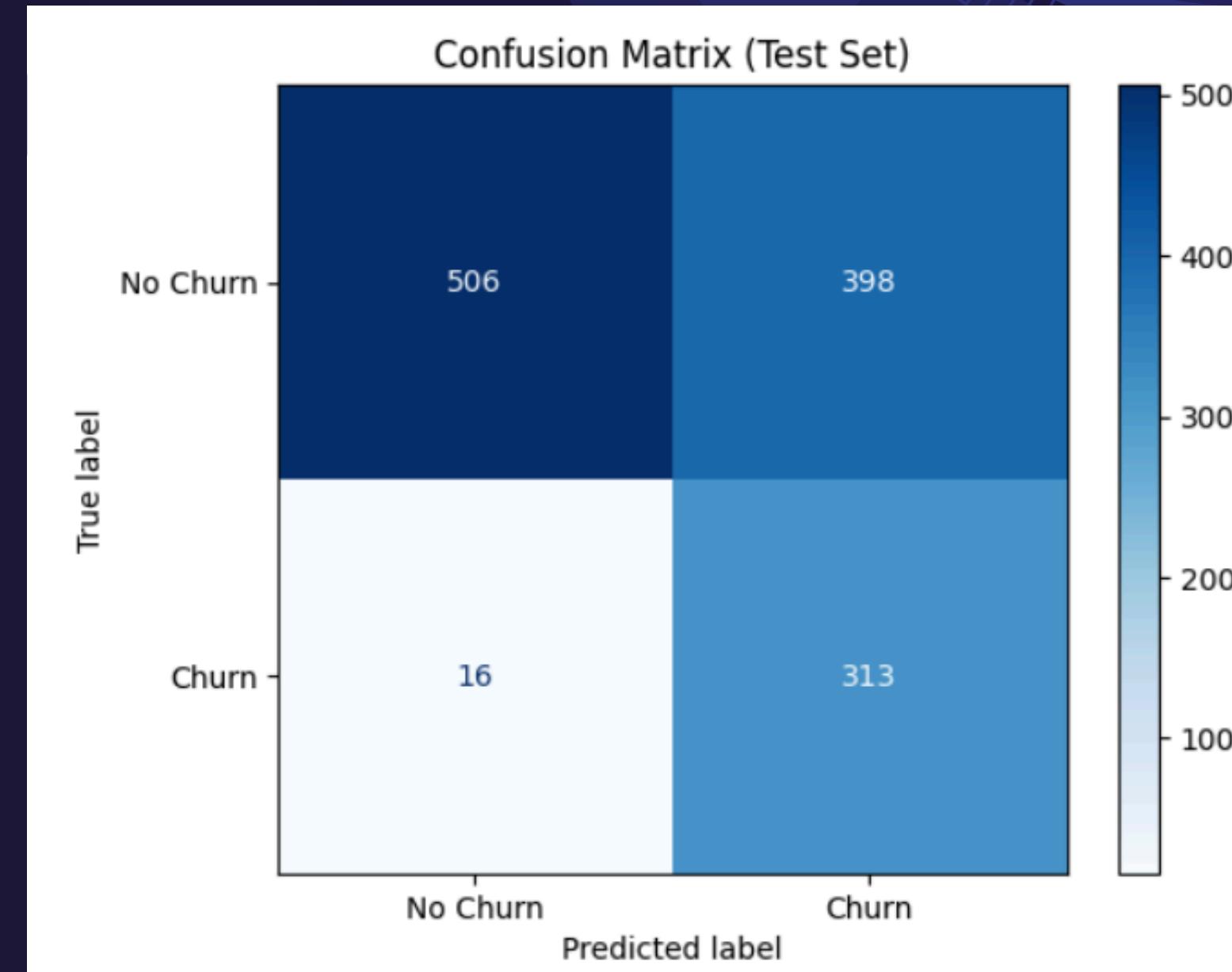
# Best Model after Hyperparameter Tuning



# Best Model after Hyperparameter Tuning

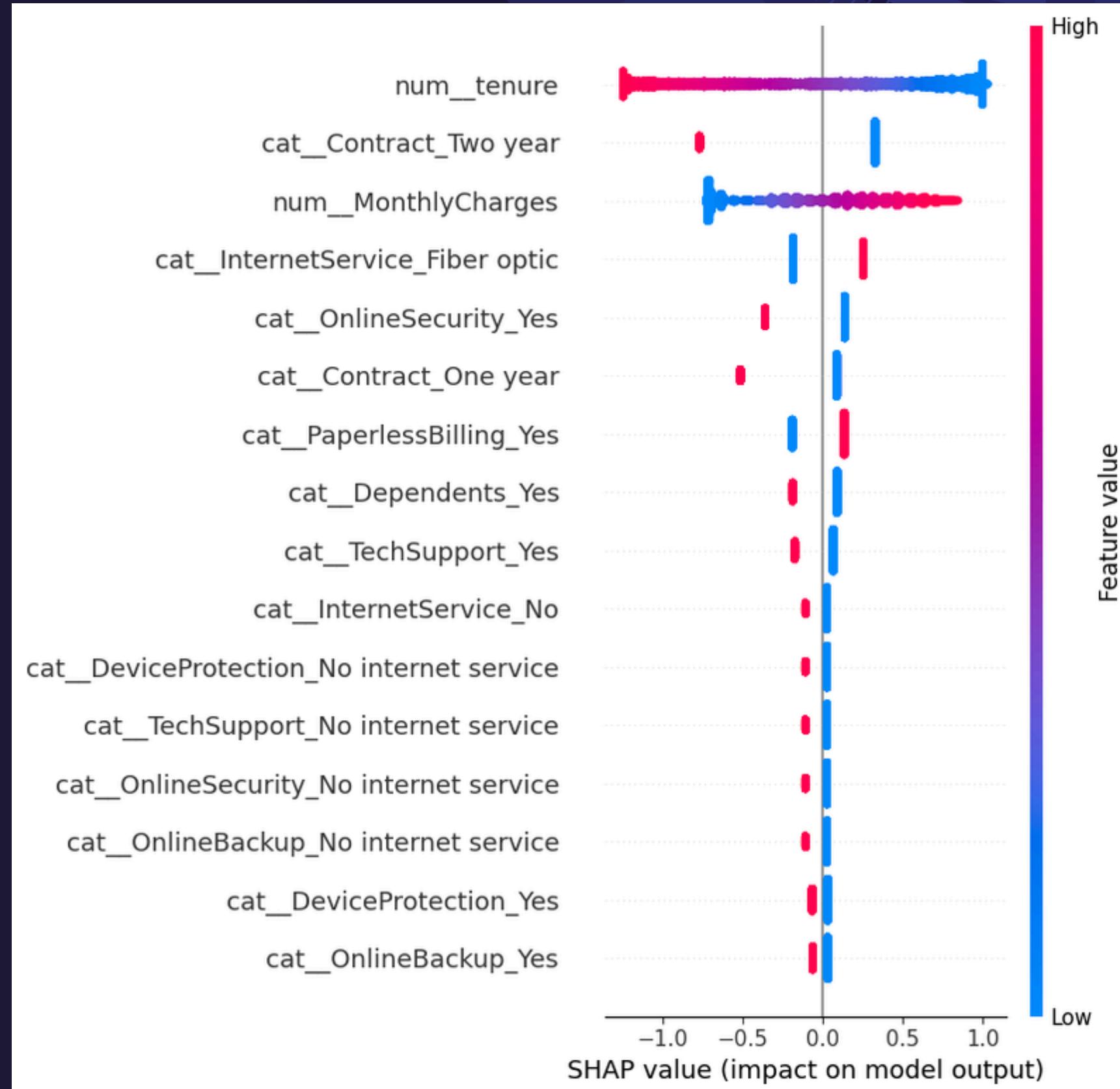


# Evaluation



	precision	recall	f1-score	support
No Churn	0.97	0.56	0.71	904
Churn	0.44	0.95	0.60	329

# Interpretasi Model - Shap



# Conclusion Recomendation



# Conclusion

- Dataset Imbalanced dengan dominasi kelas negatif (tidak Churn)
- Logistic Regression merupakan base model terbaik
- Random Under Sampling meningkatkan performa model
- Hyperparameter Tuning meningkatkan performa model
- Fitur Jenis kontrak, tenure, MonthlyCharges dan layanan tambahan berpengaruh signifikan terhadap probabilitas Churn

## Best Model

**Logistics Regression  
with RandomUnderSampler  
after Hyperparameter Tuning**

# Rekomendasi

- penambahan fitur ID\_Customer agar jika terdapat duplikasi data dapat dilihat apakah benar-benar duplikasi atau pelanggan yang berbeda dengan karakteristik yang sama
- dapat ditambahkan fitur rata-rata pemakaian layanan internet dalam sehari, riwayat keterlambatan pembayaran dan riwayat komplain untuk meningkatkan performa dan ketepatan prediksi.
- Model dapat di retraining secara berkala agar model up-to-dated
- Model dapat digunakan oleh tim CRM dan Marketing sebagai basis pengambilan keputusan agar relevan
- Pelanggan dengan risiko Churn yang tinggi dapat menjadi prioritas dalam pemberian promosi
- \*Performa model perlu di monitoring agar jika terjadi perubahan yang signifikan dapat segera diperbaiki dan tidak mempengaruhi atau bahkan merugikan bisnis yang sedang berjalan.
-

# Thank You

