



# **CUSTOMER CHURN PREDICTION**

## **HYPERPARAMETER TUNNING**

Much Hafizh Izzulhaq  
Batch 33



-

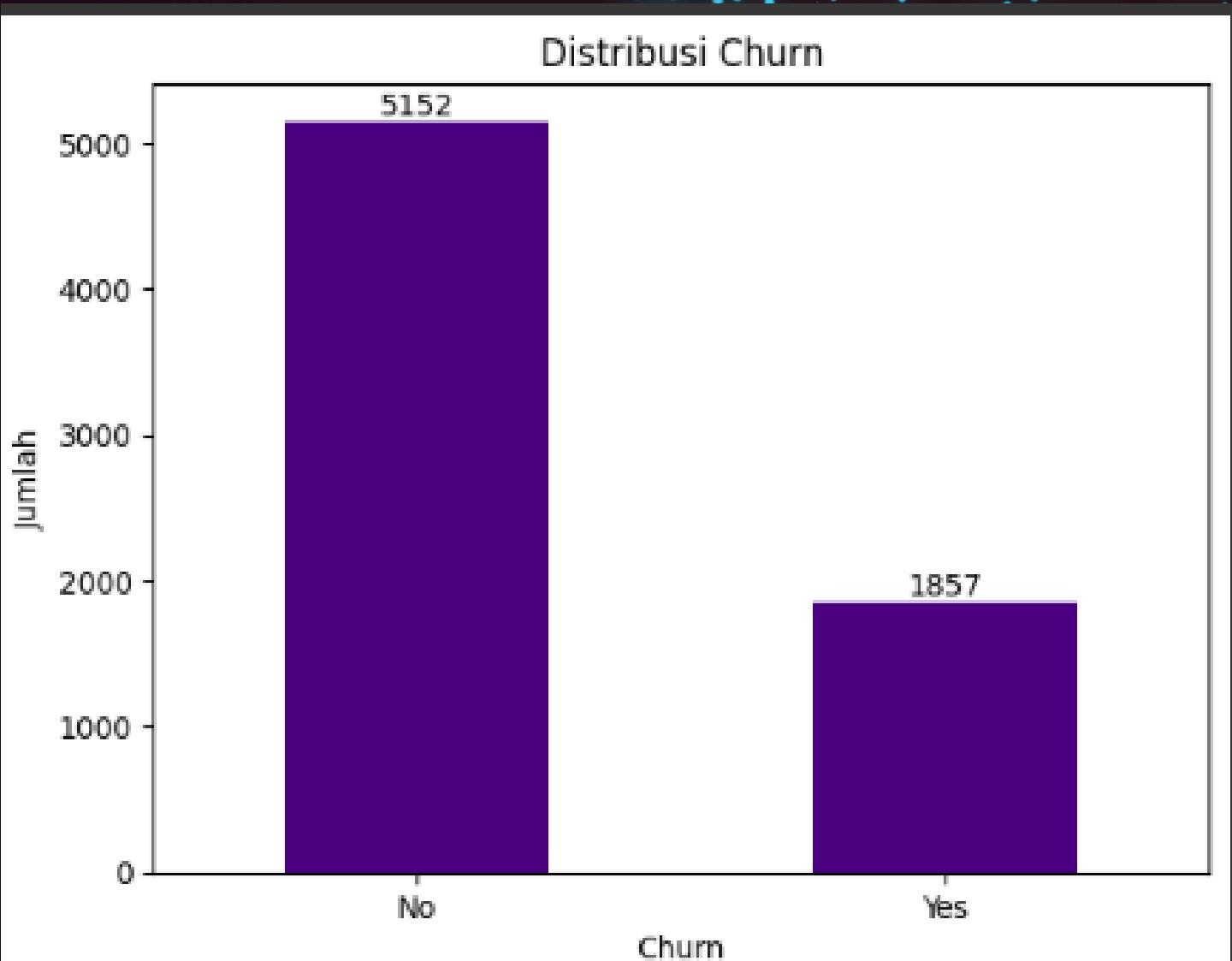
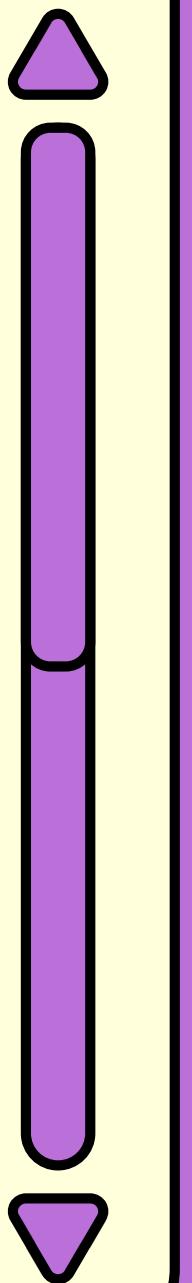
# Latar Belakang

Dalam industri telekomunikasi, tingkat perpindahan pelanggan (customer churn) menjadi salah satu tantangan besar yang dapat memengaruhi profitabilitas perusahaan. Dengan menganalisis data dapat ditemukan potensi kerugian, mengingat biaya untuk mendapatkan pelanggan baru jauh lebih tinggi dibanding mempertahankan pelanggan lama. Oleh karena itu, diperlukan analisis yang lebih mendalam untuk memahami pola dan faktor-faktor yang memengaruhi churn pelanggan.



# Objective

Pada data Telkom tercatat peningkatan churn pelanggan sekitar 26%, yang berpotensi menimbulkan kerugian signifikan bagi perusahaan. Untuk mengantisipasi hal tersebut, diperlukan analisis mendalam guna memahami faktor penyebab churn dan menjaga profitabilitas perusahaan.



# Tujuan

1

Mengurangi potensi  
peningkatan jumlah  
customer yang beresiko  
melakukan churn

2

Memberikan strategi  
retensi pelanggan  
berdasarkan hasil  
prediksi churn

3

Membantu perusahaan  
menurunkan churn rate  
dan meningkatkan  
profitabilitas

# Workflow



[LINK GOOGLE COLAB](#)

# Data Understanding

0	customerID	ID unik pelanggan
1	Gender	Jenis kelamin pelanggan
2	SeniorCitizen	Apakah pelanggan adalah warga senior (0 = tidak, 1 = ya)
3	Partner	Apakah pelanggan memiliki pasangan
4	Dependents	Apakah pelanggan memiliki tanggungan
5	Tenure	Lama berlangganan (dalam bulan)
6	Contract	Jenis kontrak langganan (Bulanan, Tahunan, dll.)
7	PaperlessBilling	Apakah menggunakan penagihan tanpa kertas
8	PaymentMethod	Metode pembayaran
9	MonthlyCharges	Biaya langganan bulanan
10	TotalCharges	Total biaya yang sudah dibayarkan
11	Churn	Apakah pelanggan berhenti berlangganan (Yes/No) — kolom target

## Dataset :

- Data ini berisi transaksi dan background pelanggan telkom dengan 12 Kolom dan 7046 data transaksi

# Data Preprocessing

```
# Check Missing Values  
df.isnull().sum()
```

	0
Gender	16
SeniorCitizen	12
Partner	16
Dependents	19
Tenure	17
Contract	13
PaperlessBilling	13
PaymentMethod	11
MonthlyCharges	10
TotalCharges	0
Churn	0

dtype: int64

## Missing Values

Menangani Missing Values dengan mengisi nilai kosong pada kolom Numerik dan Kategorik

## Duplicate Data

```
[159] # Cek Data Duplicate  
df.duplicated().sum()
```

→ np.int64(37)

```
[160] df = df.drop_duplicates()
```

Melakukan drop pada data dupliait agar menghindari Overfitting saat melakukan modeling

## Data Encoding

▼ Nominal

```
[175] from sklearn.preprocessing import LabelEncoder  
  
nominal_cols = ['Gender', 'Partner', 'Dependents', 'PaperlessBilling', 'PaymentMethod', 'Churn']  
  
encoder = LabelEncoder()  
  
for col in nominal_cols:  
    if col in df.columns:  
        df[col] = encoder.fit_transform(df[col].astype(str)).astype(int)
```

▼ Ordinal

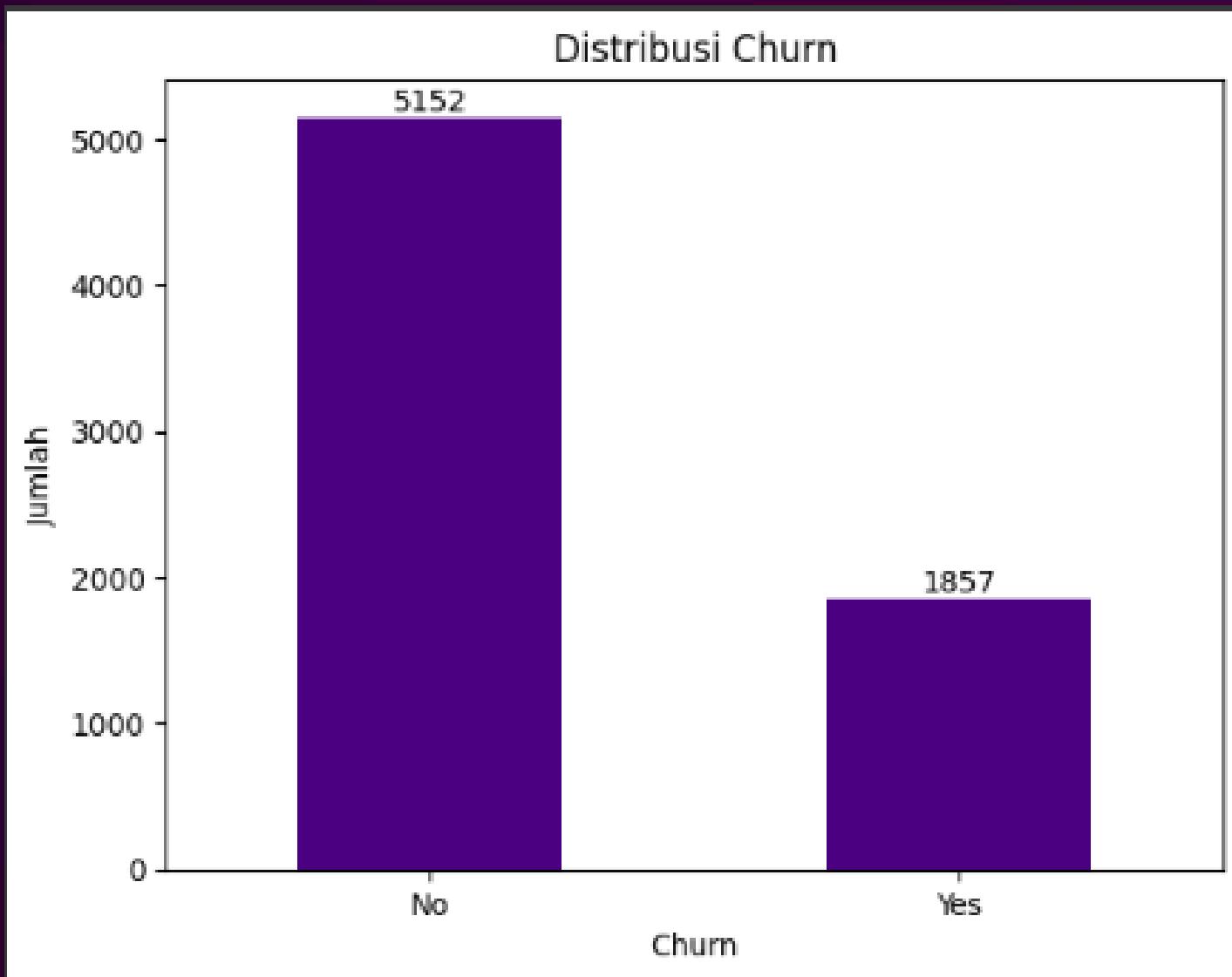
```
[176] encoder = OrdinalEncoder(  
    categories=[['Month-to-month', 'One year', 'Two year']],  
    handle_unknown='error',  
)  
  
df['ContractEncode'] = encoder.fit_transform(df[['Contract']])  
df = df.drop(columns='Contract')  
  
encoder.categories_
```

→ [array(['Month-to-month', 'One year', 'Two year'], dtype=object)]

Melakukan encode pada data kategoreikal nominal dan ordinal untuk proses selanjutnya

# Exploratory Data Analysis

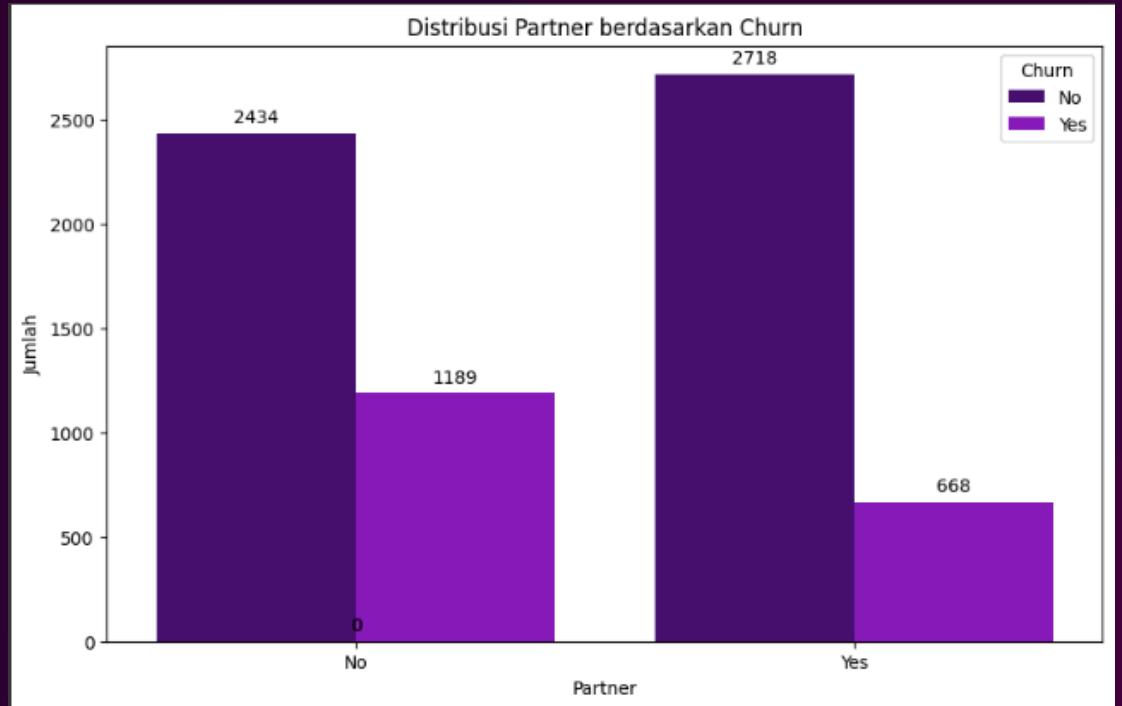
## Churn



**No** : 5152 (Tidak Churn)  
**Yes** : 1857 (Churn)

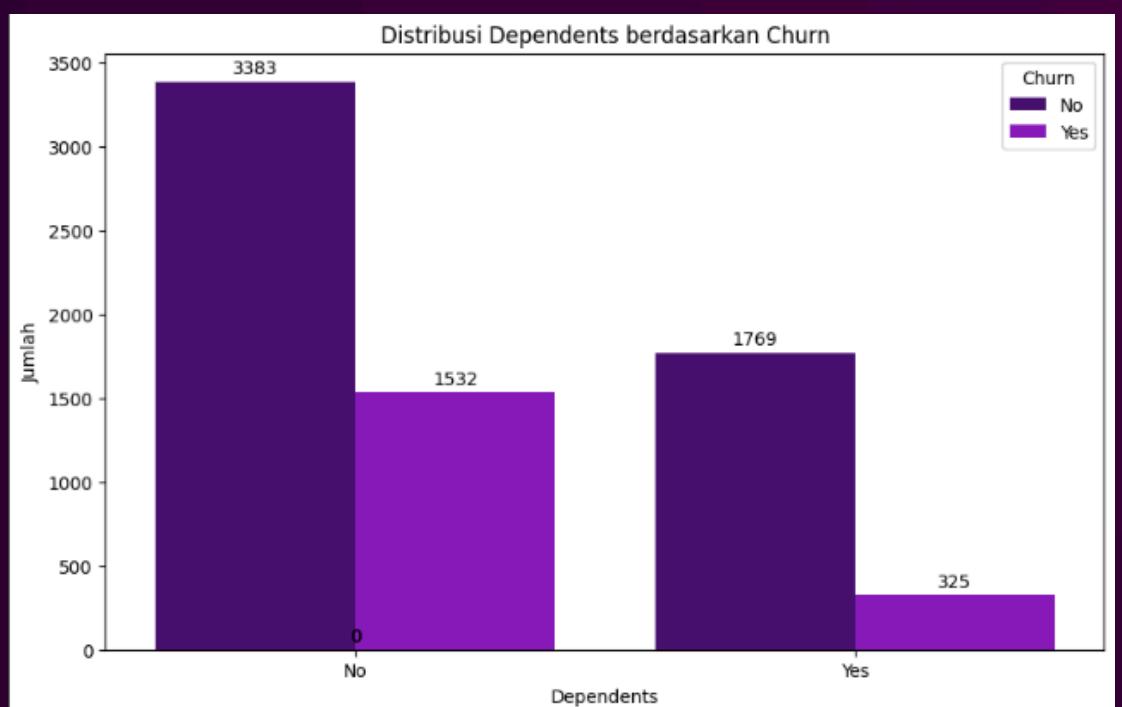
Perbandingan antara customer yang melakukan churn sekitar 26% dari customer yang tidak melakukan churn

# Exploratory Data Analysis



## Partner (Memiliki Pasangan)

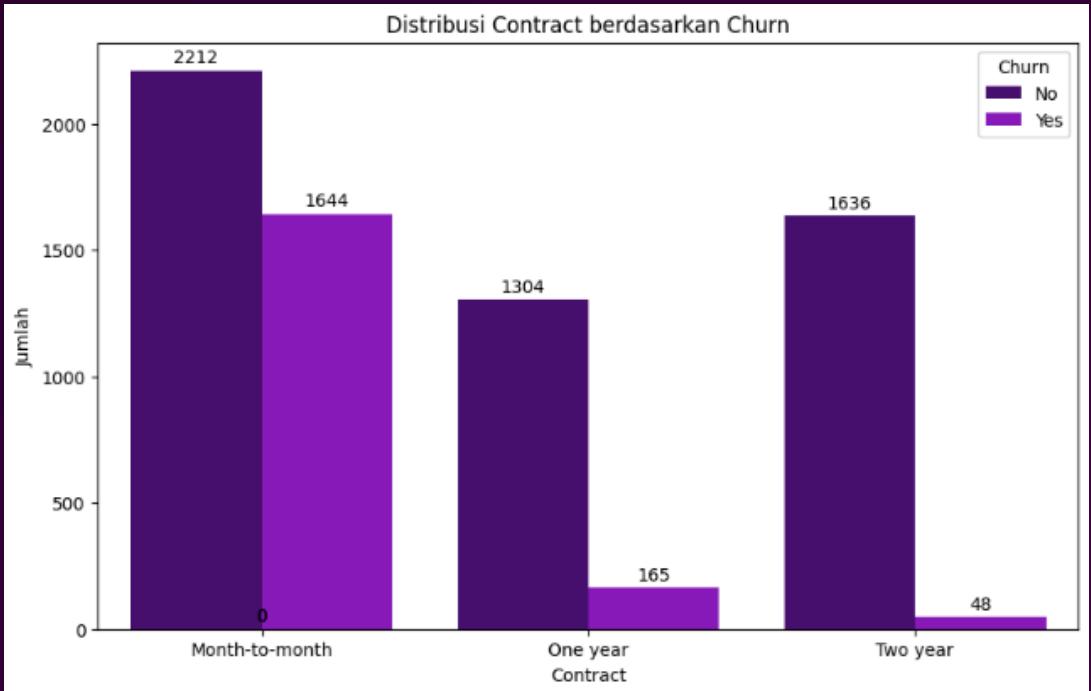
Customer yang memiliki pasangan paling banyak melakukan churn sekitar 1189 customer, mungkin di karenakan biaya yang mahal yang harus dibayarkan di bandingkan customer yang memiliki pasangan.



## Dependents (Memiliki tanggungan)

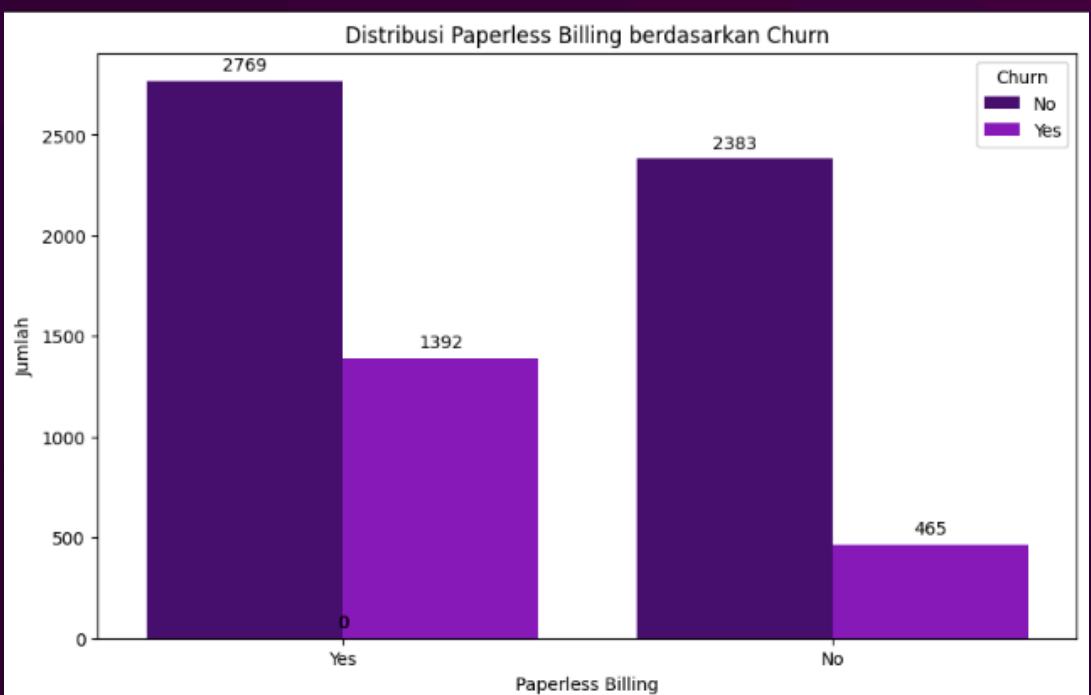
Customer yang tidak memiliki tanggungan cenderung melakukan churn, terdapat 1532 customer yang melakukan churn. Mungkin customer lebih memilih churn karena lebih fleksibel untuk berhenti di bandingkan yang memiliki tanggungan.

# Exploratory Data Analysis



## Contract

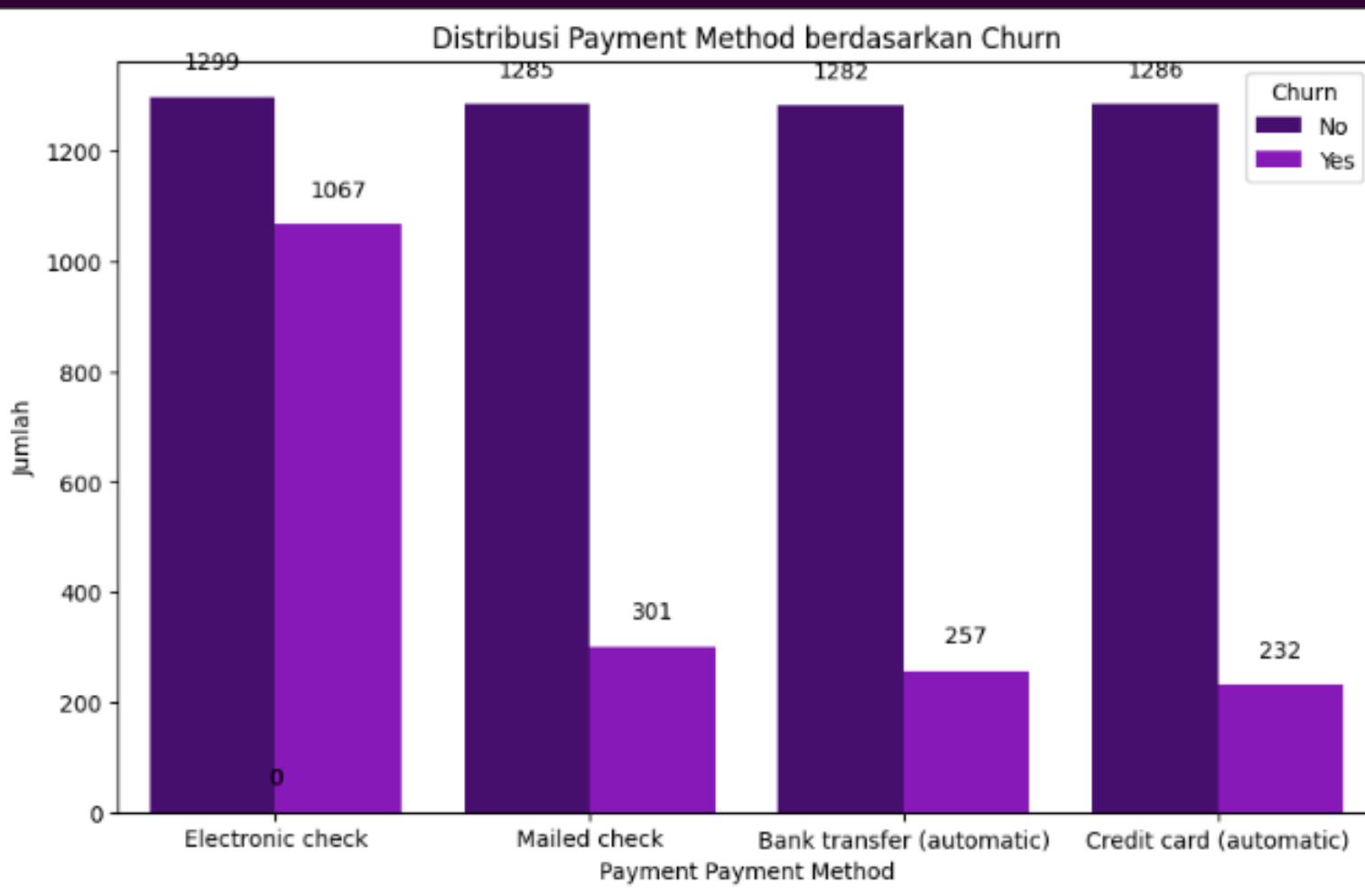
Dari ke 3 katergori contract yang paling signifikan adalah Month To Month, pelanggan yang paling banyak melakukan churn adalah pelanggan baru. Mungkin di karenakan biaya yang tinggi di awal yang customer rasakan tidak sebanding dengan apa yang di dapatkan



## Paperless Billing

Customer yang menerima tagihan melalui e-bill paling banyak melakukan churn yang sekitar 1392 customer. karena pelanggan yang baru mungkin melakukan pembayaran e-bill.

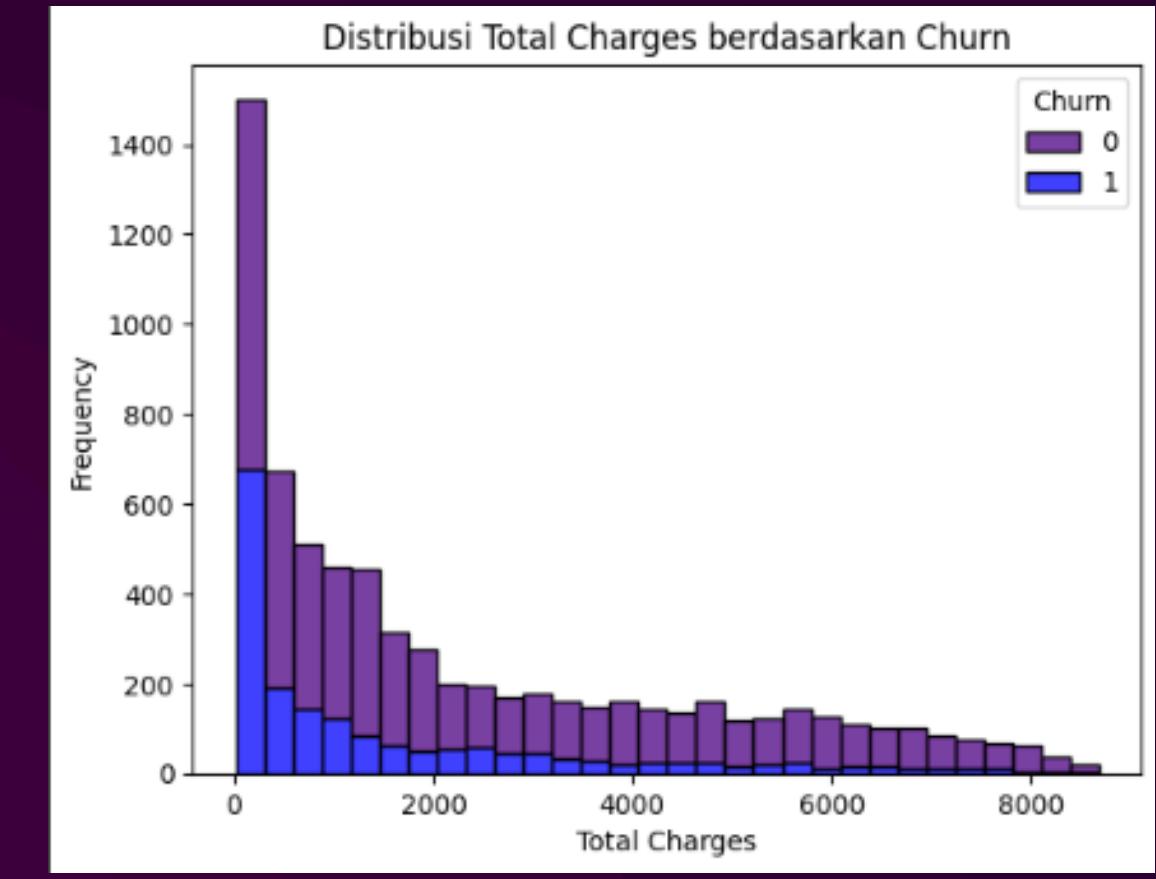
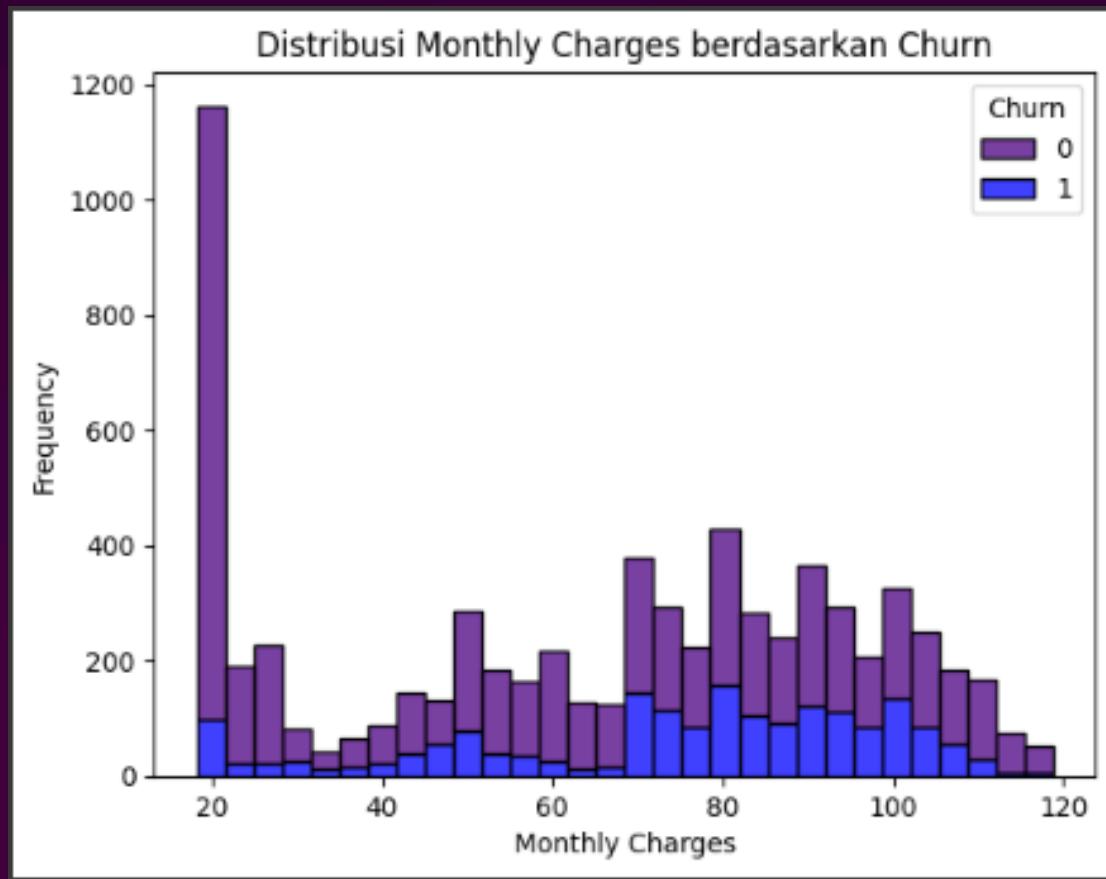
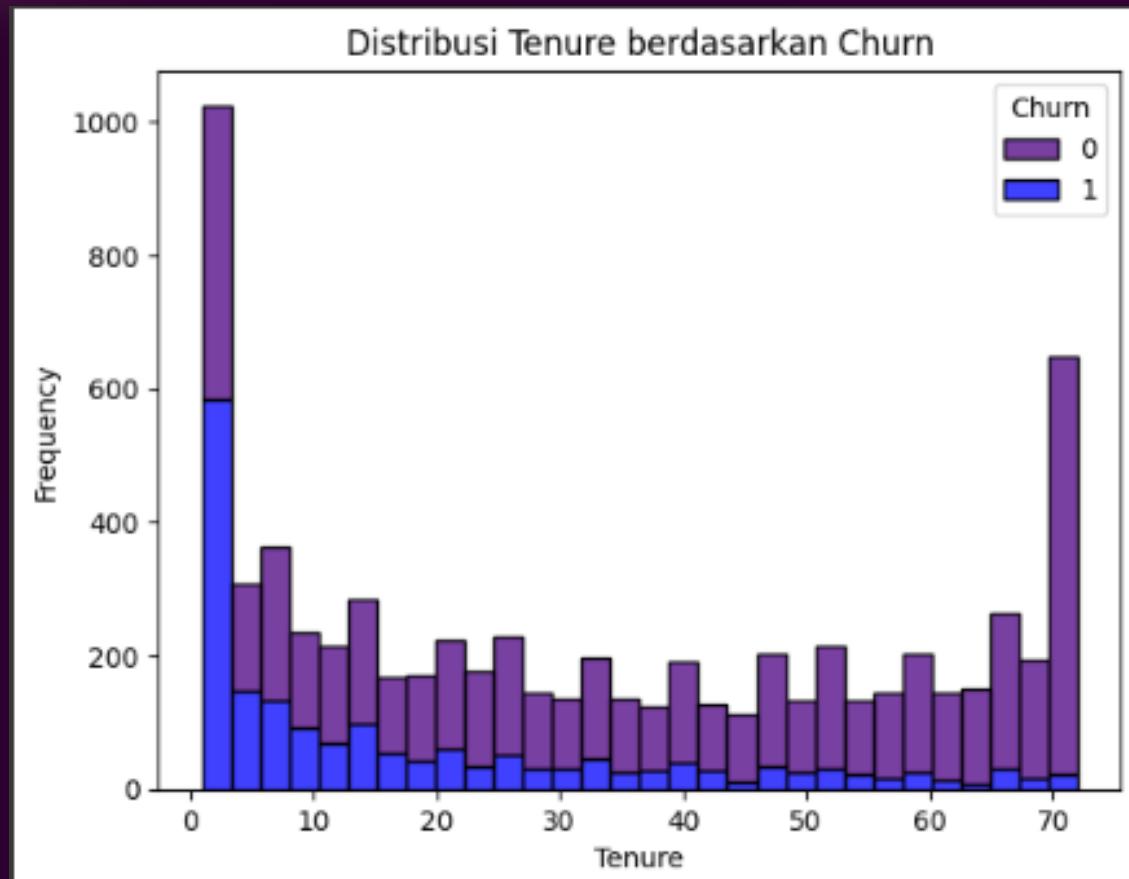
# Exploratory Data Analysis



## Electronic Check

Customer yang menerima tagihan melalui electronic check paling banyak melakukan churn sekitar 1067 customer dan memiliki gap yang sangat tinggi di bandikan metode pembayaran lainnya.

# Exploratory Data Analysis



Rata - rata customer yang melakukan churn yang memiliki tenure kurang dari 20 bulan, maka dari ini untuk rekomendasi bisnis akan berfokus pada customer baru

Pelanggan dengan biaya bulanan antara 70 hingga 110 bulan cenderung banyak melakukan churn hal ini bisa mengidentifikasi bahwa pelanggan biaya bulanan lebih tinggi merasa layanan kurang sepadan dengan biaya yang mereka keluarkan

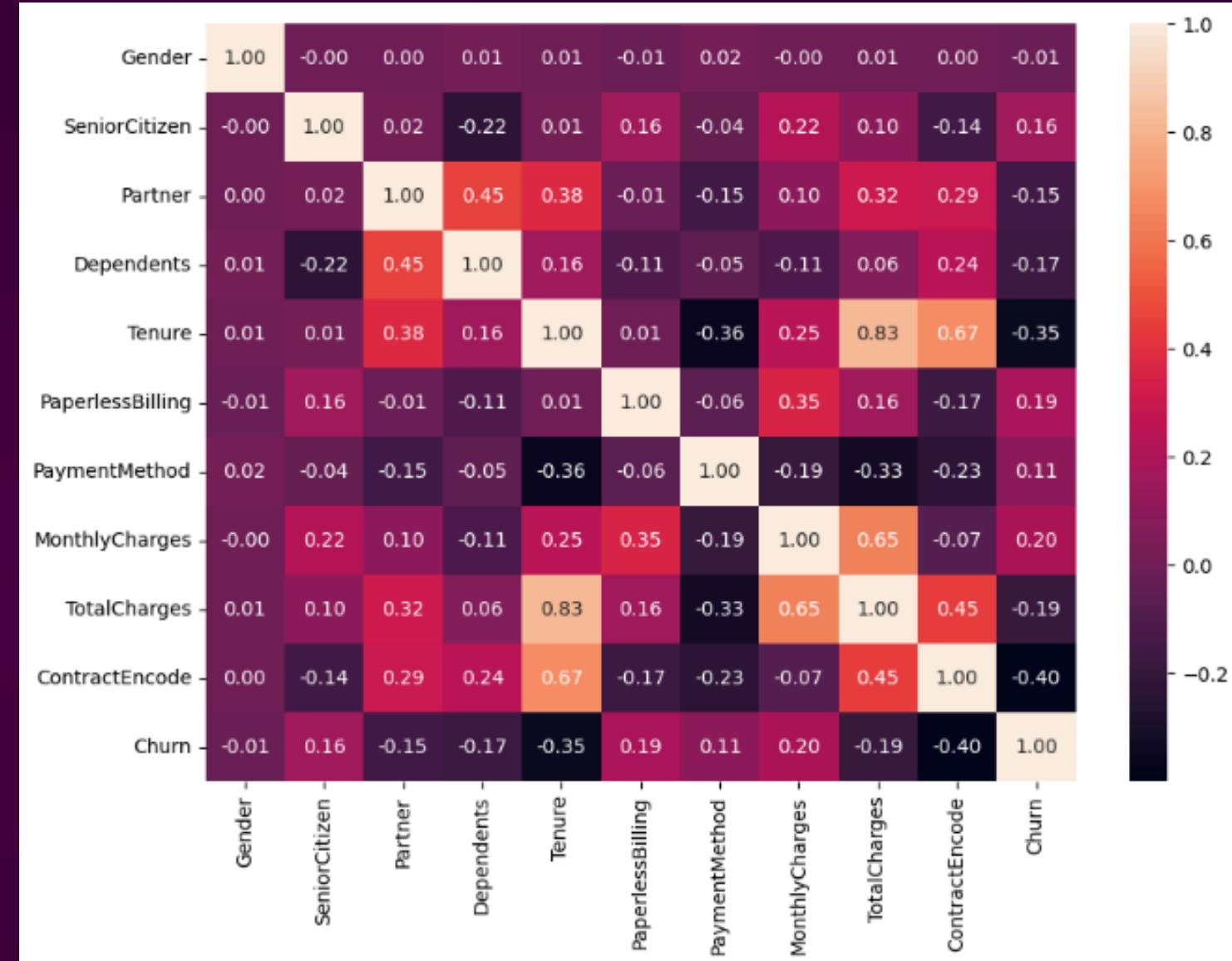
Sebagian besar yang churn memiliki total langganan kurang dari 1800 USD. ini menguatkan pelanggan baru yang paling banyak melakukan churn

# Analysis Correlation

## VIF Score

	feature	vif_score
1	Gender	1.000980
2	SeniorCitizen	1.132052
3	Partner	1.458637
4	Dependents	1.378430
5	Tenure	7.226695
6	PaperlessBilling	1.180540
7	PaymentMethod	1.175248
8	MonthlyCharges	3.604543
9	TotalCharges	9.508796
10	ContractEncode	2.162289

Melakukan Analysis correlation dengan menggunakan metode VIF Score, dan terlihat nilai VIF > 0.4 yaitu Tenure dan Total Charges dengan itu kita melakukan drop salah satu kolom .



Melakukan drop pada feature Total Charge karena nilai korelasi antar Target rendah sekitar 0.19 di bandingkan Feature Tenure yang memiliki nilai korelasi 0.35

# Model Evaluation

Model	Train	Test
kNN	0.5	0.47
Random Forest	0.46	0.48
Decision Tree	0.51	0.62
XGBoost	0.79	0.83

Dalam evaluasi model, metrik yang digunakan adalah Recall karena bertujuan meminimalkan resiko kehilangan pelanggan yang benar-benar melakukan churn.

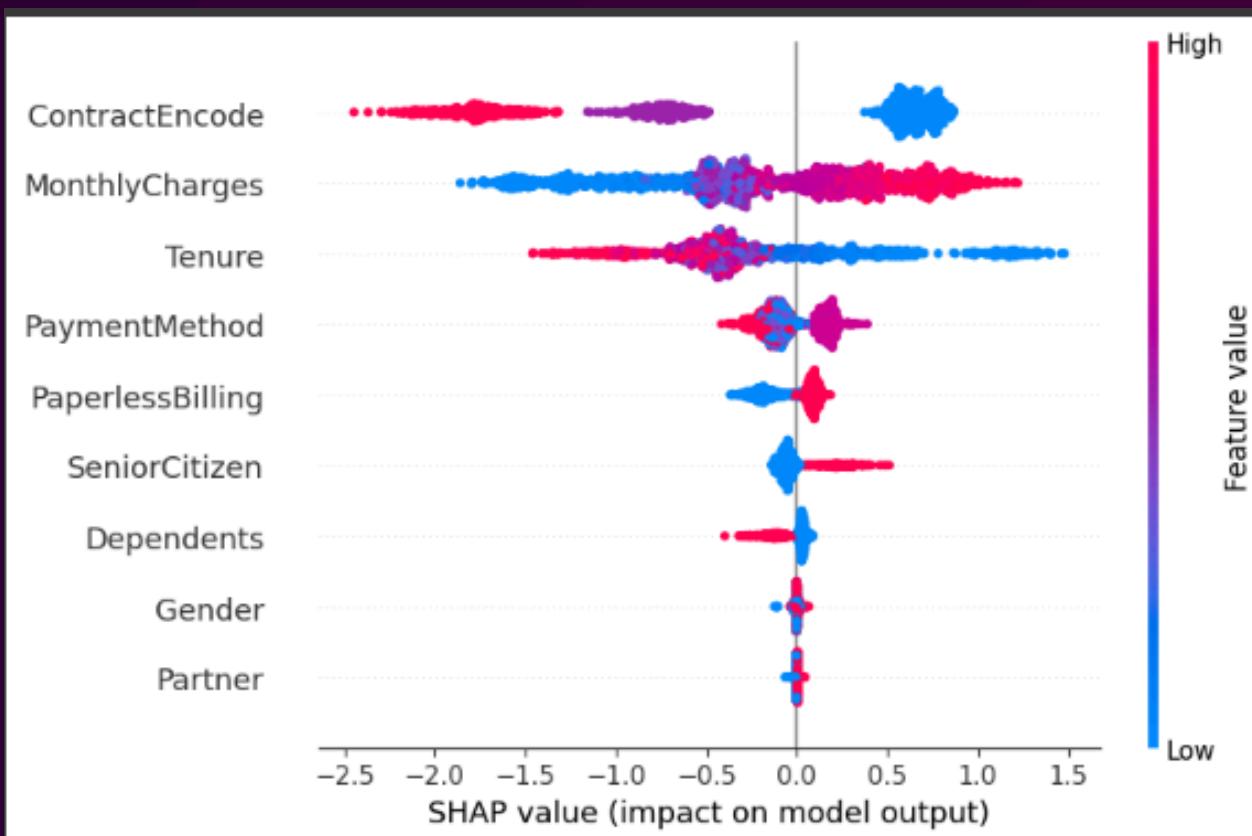
Berdasarkan hasil evaluasi dengan Hyperparameter Tuning pada keempat algoritma Machine Learning yang diuji, XGBoost memberikan nilai Recall terbaik dibandingkan model lainnya, sehingga dipilih sebagai model yang lebih optimal untuk memprediksi churn pelanggan.

# Feature Importance

**Contract** : Month to Month

**Tenure** : Kurang dari 2 tahun

**MonthlyCharges** : Semakin besar, prob churn makin tinggi



- **Contract** : Feature paling berpengaruh terhadap churn apabila kontrak semakin rendah maka kemungkinan churn makin tinggi
- **Monthly Charges** : semakin tinggi tagihan bulanan maka semakin tinggi kemungkinan melakukan churn
- **Tenure** : semakin rendah tenure semakin tinggi kemungkinan customer melakukan churn

```
[190] df_churn = df_test[df_test['probability_churn'] >= 0.8]
df_churn.sample(3)

   en Partner Dependents Tenure PaperlessBilling PaymentMethod MonthlyCharges TotalCharges ContractEncode probability_churn
0 0 0 0 14.0 1 2 100.55 1415.55 0.0 0.857277
0 0 0 1.0 1 2 81.95 81.95 0.0 0.911560
0 0 0 7.0 1 2 79.70 586.05 0.0 0.804960

[191] print(f'Jumlah Target Customer',len(df_churn))
Jumlah Target Customer 158

[194] df_target = df_churn[(df_churn['ContractEncode'] == 0) & (df_churn['Tenure'] <= 24) & (df_churn['MonthlyCharges'] >= 80)]
df_target.sample(3)

   Gender SeniorCitizen Partner Dependents Tenure PaperlessBilling PaymentMethod MonthlyCharges TotalCharges ContractEncode
customerID
5747-PMBSQ 1 1.0 1 0 10.0 1 3 92.50 934.1 0
4285-GYRQC 0 0.0 1 0 7.0 1 2 86.25 587.1 0
0334-GDDSO 1 1.0 0 0 3.0 1 2 81.35 233.7 0

[195] print(f'Jumlah Target Customer',len(df_target))
Jumlah Target Customer 86
```

Terdapat 61 customer yang kemungkinan melakukan churn dengan nilai prob 0.7, Tenure kurang dari 24 bulan atau 2 tahun, dan pembayaran bulanan lebih dari \$80

# Recommendation

1. Customer dengan kontrak bulanan (month-to\_month), perusahaan sebaiknya menawarkan potongan harga di awal berlangganan seperti potongan harga selama 3-6 bulan pertama untuk mengurangi churn pada customer baru.
2. Customer dengan tenure kurang dari 2thn :
  - a. Layanan Tambahan bagi tenure 12 bulan
  - b. Potongan khusus bagi tenure 24 bulan
  - c. Memberikan program loyalti bagi tenure lebih dari 24 bulan
3. Customer dengan pembayaran tinggi bulanan :
  - a. Membrikan tawaran downgrade layanan yang lebih affordable
  - b. Komunikasi Proaktif kepada pelanggan





**THANK  
YOU**

Any Questions?