



AHA Consulting Group

# BUSINESS ANALYTICS AND CHURN PREDICTION TO ENHANCE TELCO CUSTOMER RETENTION



# Meet Our Team



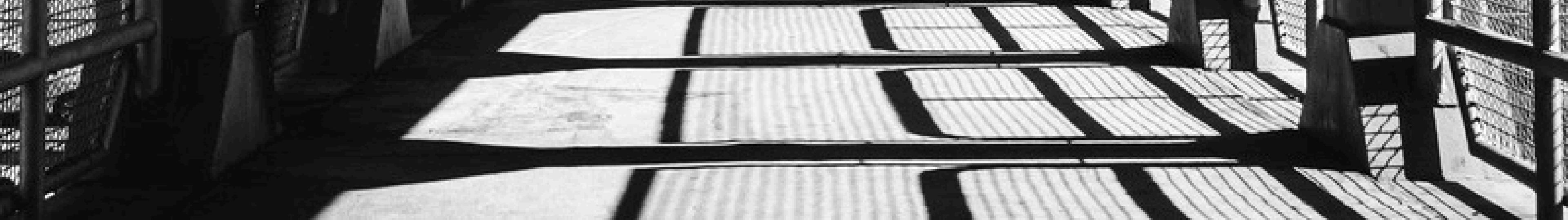
**Abednego Andries**  
Business Analyst



**Hans Darmawan**  
Data Analyst



**Alfidhiya Amany Ramli**  
Data Scientist



# AGENDA

1 Business Overview

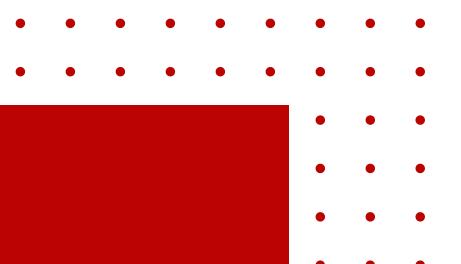
2 Data Understanding

3 Exploratory Data Analytics

4 Model Benchmarking & Hyperparameter Tuning

5 Model Interpretation

6 Conclusion & Recommendation



# Executive Summary

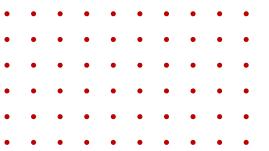
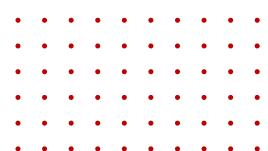


- **Solusi Prediktif:** Menggunakan Business Analytics dan Machine Learning untuk memprediksi pelanggan telekomunikasi yang berisiko churn, mengidentifikasi faktor kunci seperti kontrak bulanan dan metode pembayaran.
- **Dampak Bisnis:** Mengurangi potensi kerugian pendapatan dan memastikan strategi retensi lebih tepat sasaran.
- **Pengambilan Keputusan:** Solusi ini dirancang untuk mendukung pengambilan keputusan retensi yang dilakukan oleh tim Customer Retention (CRM) secara cepat dan berbasis data.

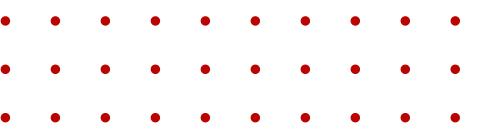




# Business Overview



# Business Overview



Business Background	Problem Statement	Risk Cost	Goals
<ul style="list-style-type: none"><li>Perusahaan XYZ adalah <b>penyedia layanan telekomunikasi</b> yang menawarkan berbagai layanan.</li><li>Mereka menghadapi <b>tantangan churn pelanggan</b> yang berdampak langsung pada pendapatan.</li><li><b>Churn</b> adalah suatu kondisi di mana pelanggan tidak melanjutkan atau menghentikan langganan layanan dalam jangka waktu tertentu.</li></ul>	<ul style="list-style-type: none"><li><b>Tidak ada sistem prediktif</b> untuk <b>mengenali pelanggan berisiko churn secara dini</b>.</li><li><b>Strategi retensi masih reaktif</b> dan <b>tidak efisien (penawaran massal)</b>.</li><li><b>Tidak diketahui secara pasti faktor-faktor utama penyebab churn</b>.</li></ul>	<ul style="list-style-type: none"><li><b>False Negative:</b> Pelanggan diprediksi <b>tidak churn</b>, ternyata <b>churn</b>. <b>Biaya akuisisi pelanggan baru (CAC)</b> adalah <b>\$694</b> per pelanggan.</li><li><b>False Positive:</b> Pelanggan diprediksi <b>churn</b>, ternyata <b>tidak churn</b>. <b>Biaya retensi pelanggan (CRC)</b> adalah <b>\$139</b>.</li></ul>	<ul style="list-style-type: none"><li>Mengembangkan model prediksi churn dengan <b>F2-score</b> untuk <b>mengurangi kerugian</b> finansial akibat <b>churn yang tak terduga</b>.</li><li><b>Mengidentifikasi faktor utama churn</b> dari sisi <b>demografi</b> dan <b>perilaku pelanggan</b>.</li><li><b>Mengurangi kerugian bisnis</b> melalui <b>strategi retensi</b> yang lebih tepat sasaran.</li></ul>

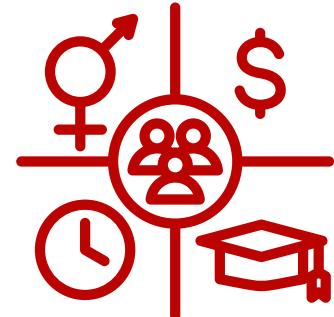
# Stakeholder Analysis

## Customer Relations Management Team (CRM)

- **Alasan:** CRM adalah **pengguna** langsung **hasil prediksi** churn.
- **Peran:** Eksekusi **kampanye penyelamatan** pelanggan.
- Tanpa mereka, model hanya jadi insight tanpa aksi.



# Analytical Approach



## Exploratory Data Analysis (EDA)

- Univariate
- Bivariate

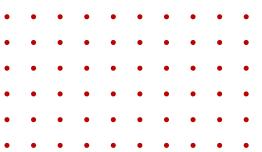
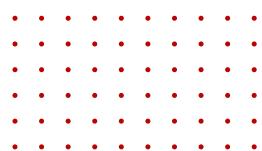


## Machine Learning Approach

- **Goal** → Identifikasi pelanggan berisiko *Churn*.
- **Model** → Menggunakan **algoritma klasifikasi** (ex: Logistic regression, stacking, bagging, boosting, dsb).
- **Metric** evaluation → **F2 Score** (Biaya akuisisi pelanggan baru > biaya retensi pelanggan)
- **Interpretasi** model → **Feature Importances / SHAP / LIME**



# Data Understanding



# Data Understanding: Demografi

Kolom	Definisi	Nilai dan Penjelasan
customerID	ID unik untuk setiap pelanggan.	Teks alfanumerik unik.
gender	Jenis kelamin pelanggan.	Male: Laki-laki, Female: Perempuan.
SeniorCitizen	Status pelanggan sebagai warga senior.	1: Ya, senior. 0: Bukan senior.
Partner	Apakah pelanggan memiliki pasangan.	Yes: Punya pasangan. No: Tidak.
Dependents	Apakah pelanggan memiliki tanggungan.	Yes: Ada tanggungan. No: Tidak.
tenure	Lama pelanggan berlangganan (dalam bulan).	Nilai numerik (misal: 1–72 bulan).
Contract	Jenis kontrak langganan.	Month-to-month, One year, Two year.

# Data Understanding: Fitur Tambahan (Layanan)

Kolom	Definisi	Nilai dan Penjelasan
PhoneService	Apakah pelanggan menggunakan layanan telepon.	Yes: Punya. No: Tidak.
MultipleLines	Apakah memiliki lebih dari satu saluran telepon.	Yes: Ya. No: Hanya 1 saluran. No phone service: Tidak berlangganan telepon.
InternetService	Jenis layanan internet yang digunakan.	DSL, Fiber optic, No (tidak berlangganan internet)
OnlineSecurity	Apakah menggunakan layanan keamanan online.	Yes, No, No internet service.
OnlineBackup	Apakah menggunakan layanan backup online.	Yes, No, No internet service.
DeviceProtection	Apakah menggunakan proteksi perangkat.	Yes, No, No internet service.
TechSupport	Apakah mendapatkan dukungan teknis.	Yes, No, No internet service.
StreamingTV	Apakah menggunakan layanan streaming TV dari Telco.	Yes, No, No internet service.
StreamingMovies	Apakah menggunakan layanan streaming film dari Telco.	Yes, No, No internet service.

# Data Understanding: Pembayaran dan Status

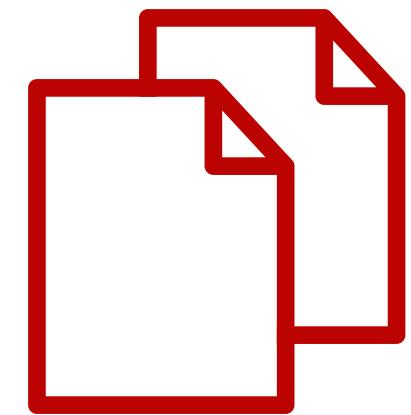
Kolom	Definisi	Nilai dan Penjelasan
PaperlessBilling	Apakah menggunakan tagihan digital.	Yes: Tagihan elektronik (email). No: Tagihan fisik.
PaymentMethod	Metode pembayaran yang digunakan.	Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic).
MonthlyCharges	Biaya bulanan pelanggan.	Nilai desimal (misal: 70.35). Total semua layanan yang digunakan per bulan.
TotalCharges	Akumulasi total biaya selama berlangganan.	Nilai desimal. Biasanya = MonthlyCharges × tenure, jika tidak ada data kosong.
Churn	Status pelanggan berhenti atau tidak (target prediksi).	Yes: Berhenti berlangganan. No: Masih aktif.

# EDA Data Preparation



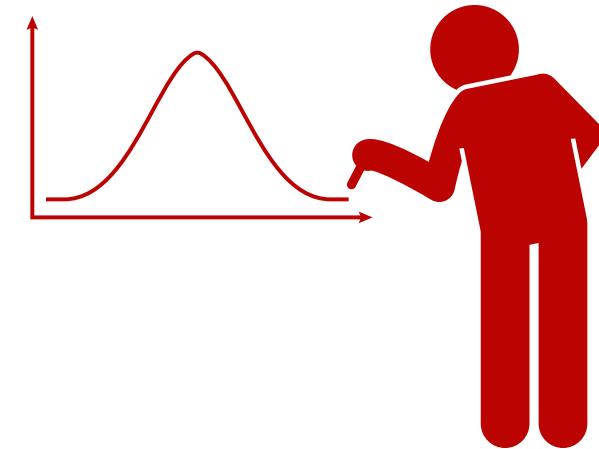
## Missing Values Checking

- 11 invisible **TotalCharges** missing.
- Dihapus karena **tenure**-nya 0.



## Duplicated Values Checking

- Tidak ada duplikasi data

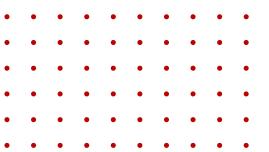
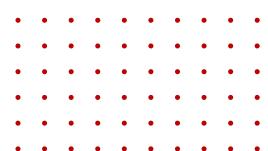


## Normality Checking

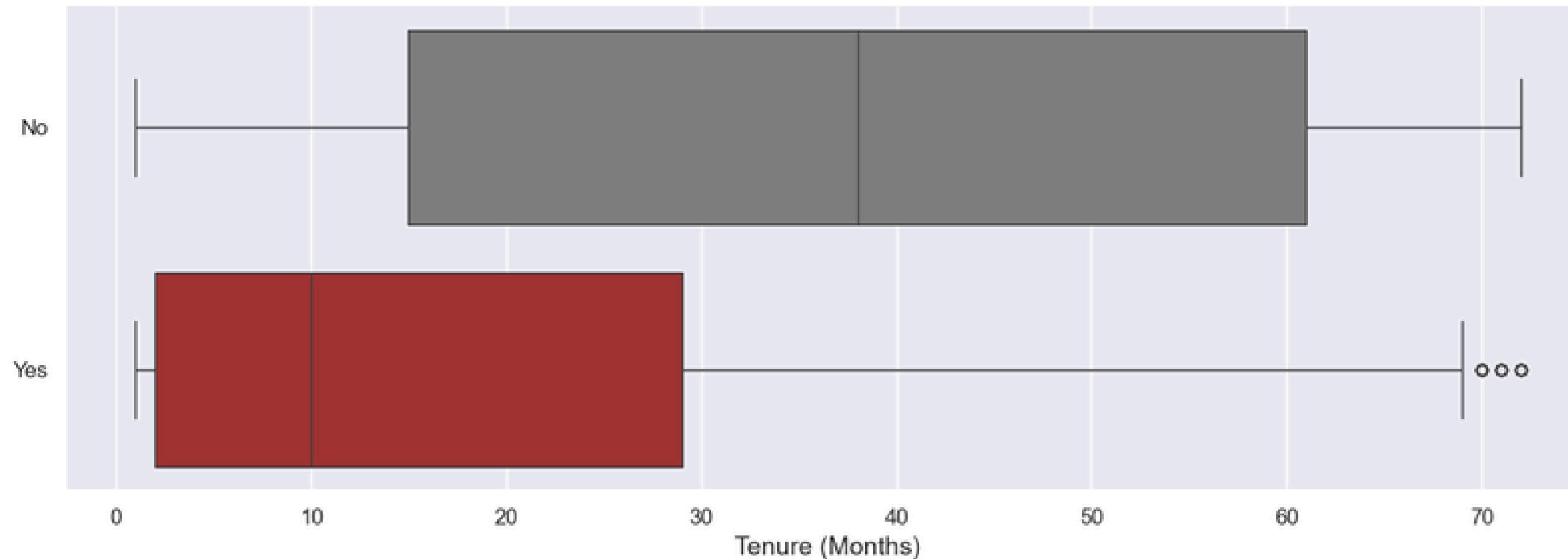
- Tiga kolom numerik: **tenure**, **TotalCharges**, dan **Monthly Charges**.
- Distribusi **tidak normal**



# Exploratory Data Analysis



# Boxplot of Tenure by Churn Status

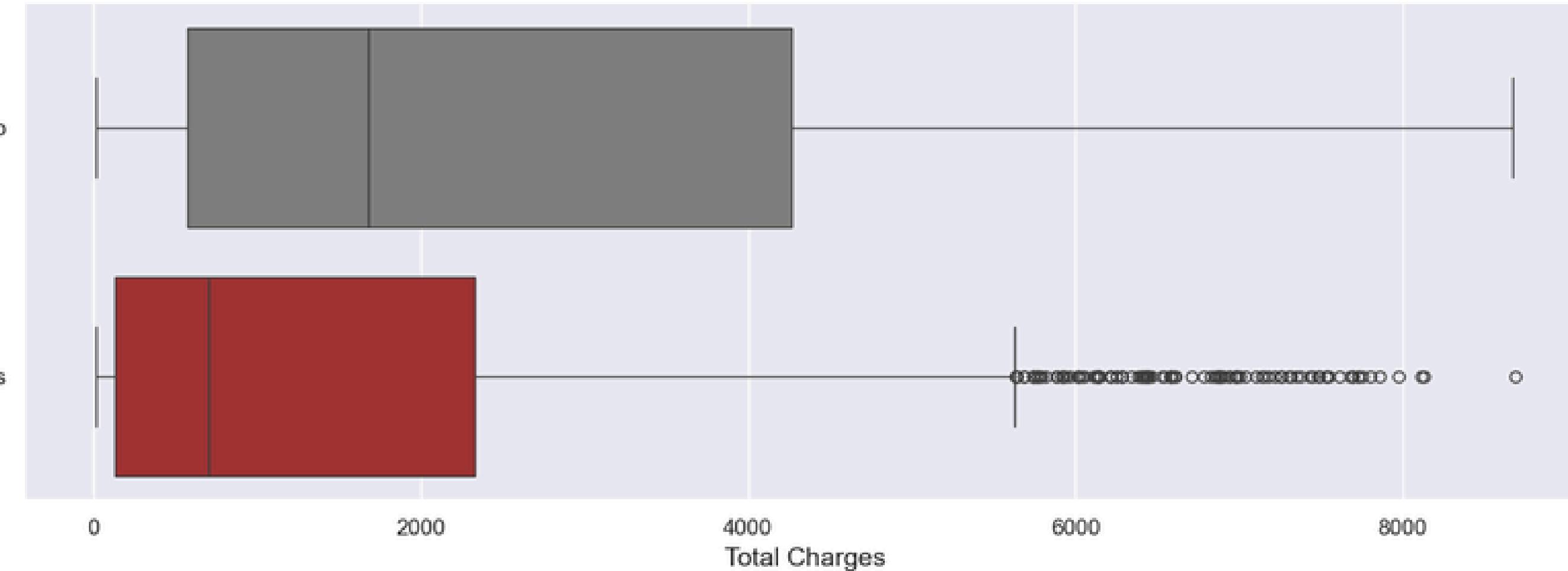


## Business Insights:

- Banyak churn terjadi **di awal masa langganan**.
- Pelanggan churn berhenti lebih cepat (**10 bulan**), dibanding non-churn yang bertahan lebih lama (38 bulan).

Churn	count	mean	std	min	25%	50%	75%	max	IQR
No	5163	37.65001	24.07694	1	15	38	61	72	46
Yes	1869	17.97913	19.53112	1	2	10	29	72	27

# Boxplot of Total Charges by Churn Status

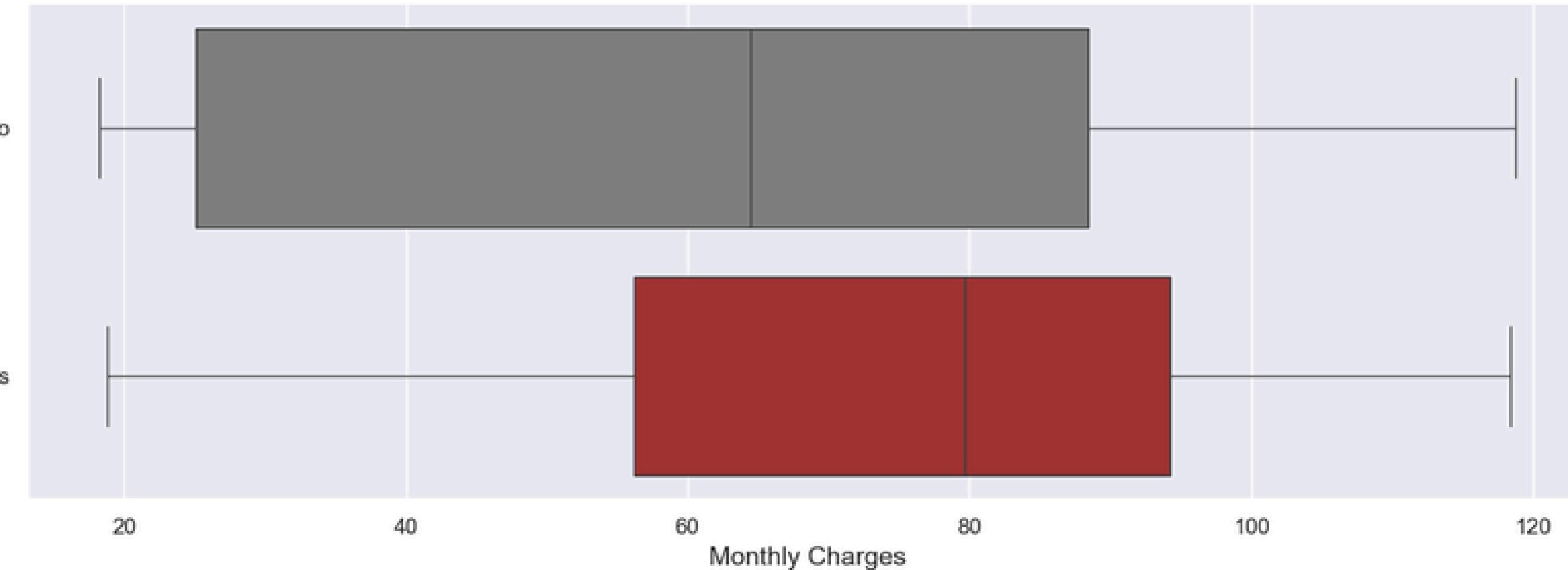


## Business Insights:

- Banyak churn terjadi saat **nilai pelanggan masih rendah**.
- Terlihat dari **total pengeluaran mereka yang lebih kecil (\$704)** dibandingkan dengan pelanggan non-churn (\$1.683).

Churn	count	mean	std	min	25%	50%	75%	max	IQR
No	5163	2555.3441	2329.4569	18.8	577.83	1683.6	4264.13	8672.5	3686.3
Yes	1869	1531.7960	1890.8229	18.85	134.5	703.55	2331.3	8684.8	2196.8

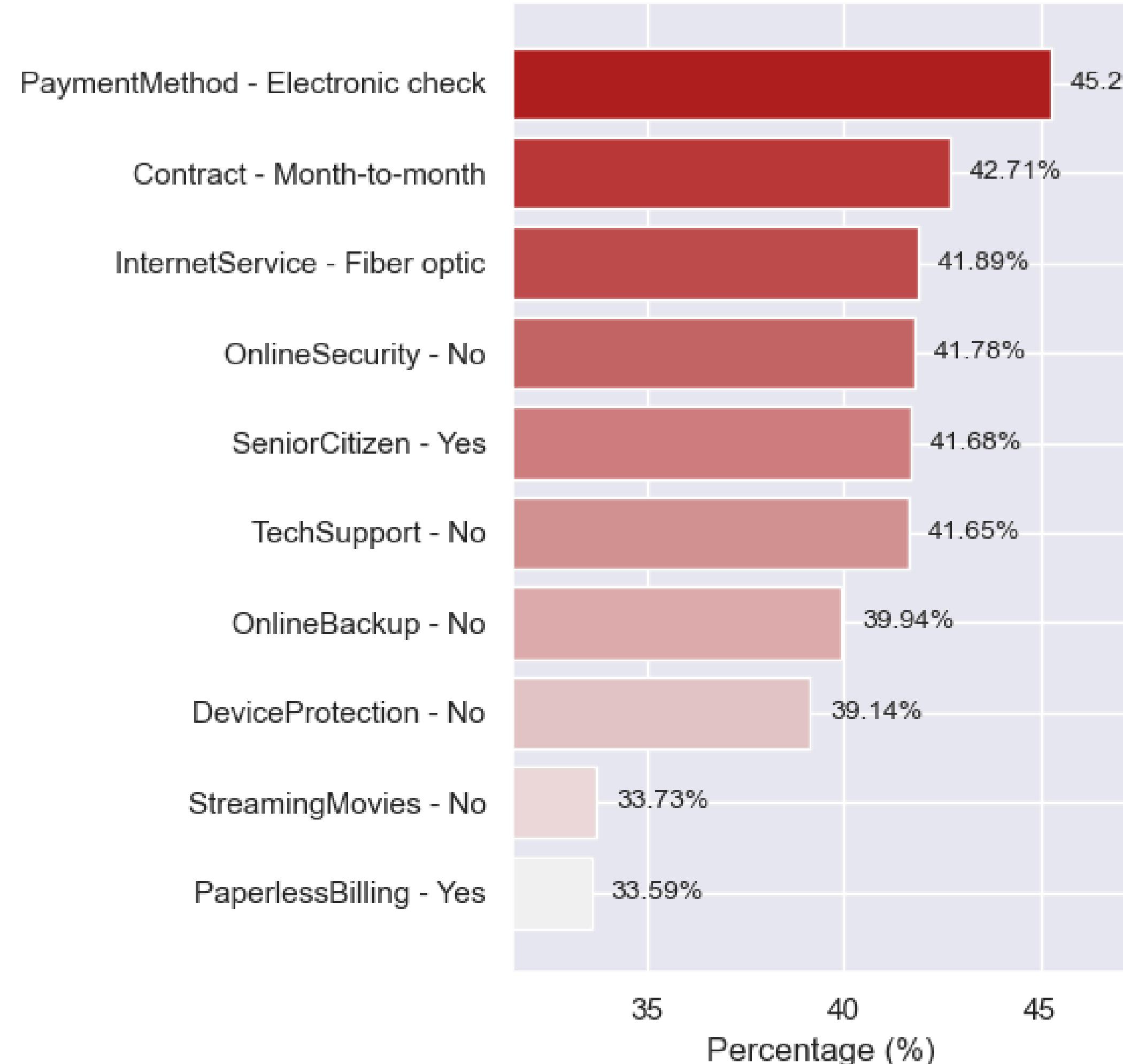
# Boxplot of Monthly Charges by Churn Status



## Business Insights

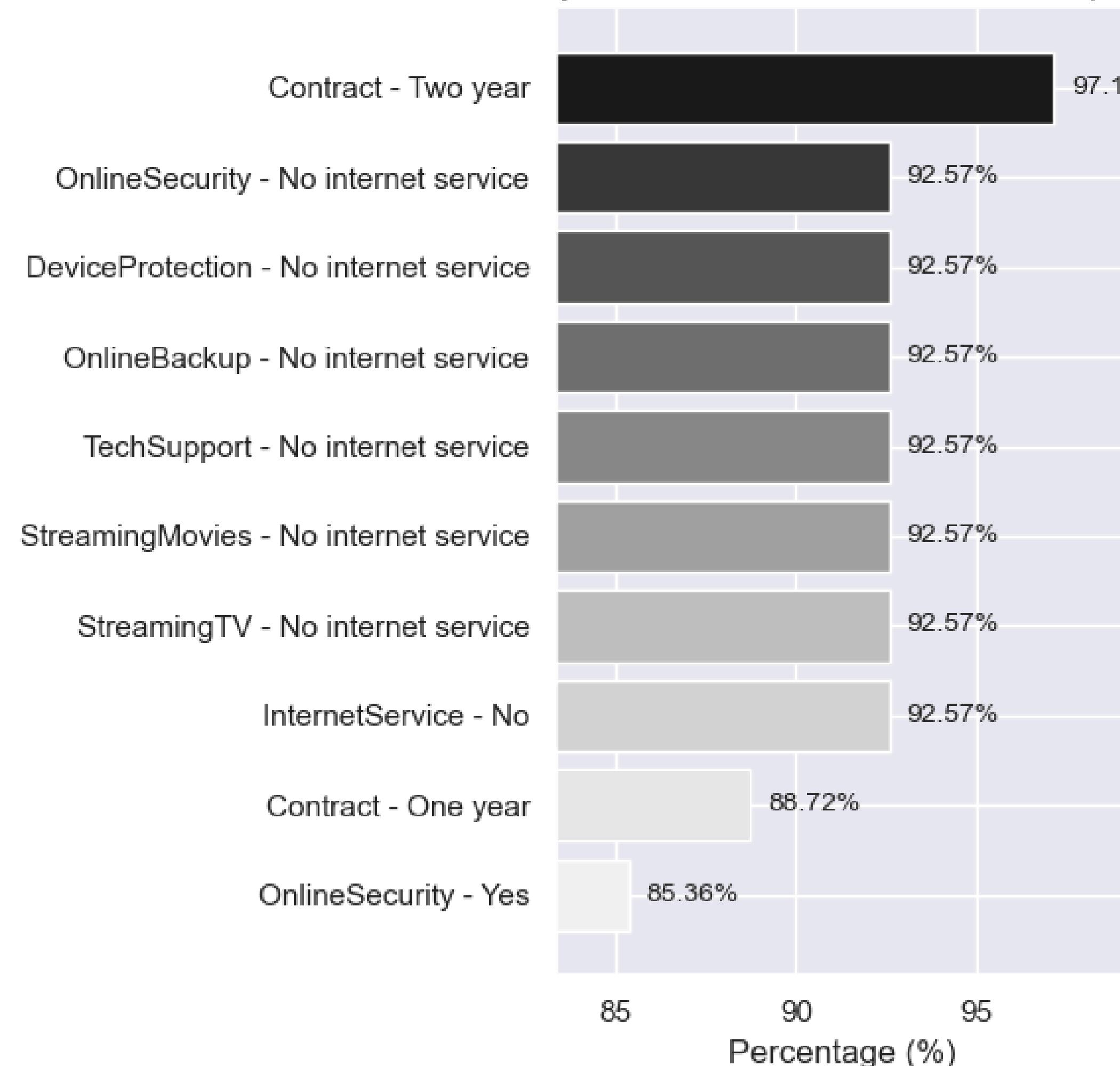
- Semakin mahal biaya bulanan, makin tinggi risiko churn.
- Pelanggan churn punya tagihan bulanan lebih tinggi (median \$79.65) dibandingkan non-churn (median \$64.45).

Churn	count	mean	std	min	25%	50%	75%	max	IQR
No	5163	61.30741	31.09456	18.25	25.1	64.45	88.475	118.75	63.375
Yes	1869	74.44133	24.66605	18.85	56.15	79.65	94.2	118.35	38.05



## Top 10 Churn Percentages

- Pelanggan lansia memiliki tingkat churn tinggi (41,68%), menandakan bahwa **senior citizen merupakan segmen rentan** yang memerlukan perhatian khusus dalam strategi retensi.
- Pelanggan tanpa layanan tambahan seperti OnlineSecurity, TechSupport, dan OnlineBackup memiliki **risiko churn lebih tinggi (~40%)**, menunjukkan bahwa **layanan bernilai tambah berperan penting dalam meningkatkan loyalitas**.
- Metode pembayaran manual seperti **Electronic check (45,29%)** dan **kontrak bulanan (42,71%)** menjadi **pemicu churn utama**, mencerminkan bahwa **pelanggan dengan fleksibilitas tinggi dan pembayaran tidak otomatis lebih rawan churn**.
- Pengguna Fiber Optic menunjukkan churn tinggi (41,89%), yang mengindikasikan potensi masalah dalam persepsi atau pengalaman pelanggan terhadap layanan.

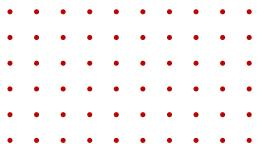


## Top 10 Not Churn Percentages

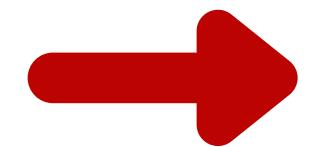
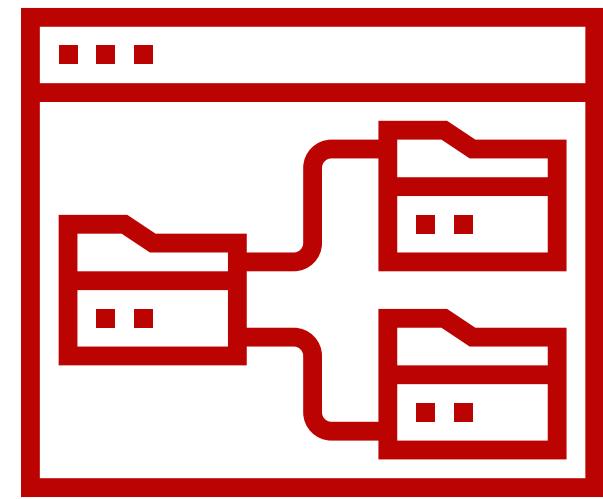
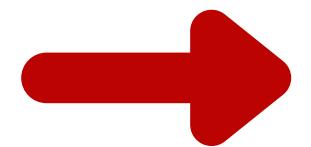
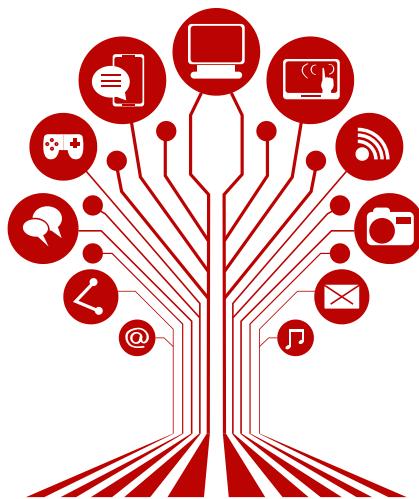
- Pelanggan dengan kontrak dua tahun memiliki tingkat retensi tertinggi (97,15%), menunjukkan bahwa komitmen jangka panjang efektif dalam menekan churn.
- Pelanggan tanpa layanan internet menunjukkan retensi sangat tinggi (92,57%) pada fitur seperti OnlineSecurity, DeviceProtection, hingga Streaming, karena layanan tersebut tidak relevan dan tidak digunakan.
- Pelanggan dengan layanan OnlineSecurity aktif memiliki tingkat retensi 85,36%, menandakan bahwa fitur tambahan bernilai dapat meningkatkan loyalitas pelanggan secara signifikan.



# Model Benchmarking & Hyperparameter Tuning



# Data Preparation



## Feature Engineering

1. Konversi "No Phone Service" / "No Internet Service" menjadi "No"
2. Mapping SeniorCitizen
  - 0 → No
  - 1 → Yes
3. Mapping target
  - Churn → 1
  - No → 0

## Feature, Target, dan Data Splitting

1. Memisahkan kolom **churn** sebagai target (y) dan **kolom lainnya** selain customer ID sebagai **features (X)**
2. Memisahkan data untuk train dan test dengan proporsi **70% data train** dan **30% data test**

## Pre-processing

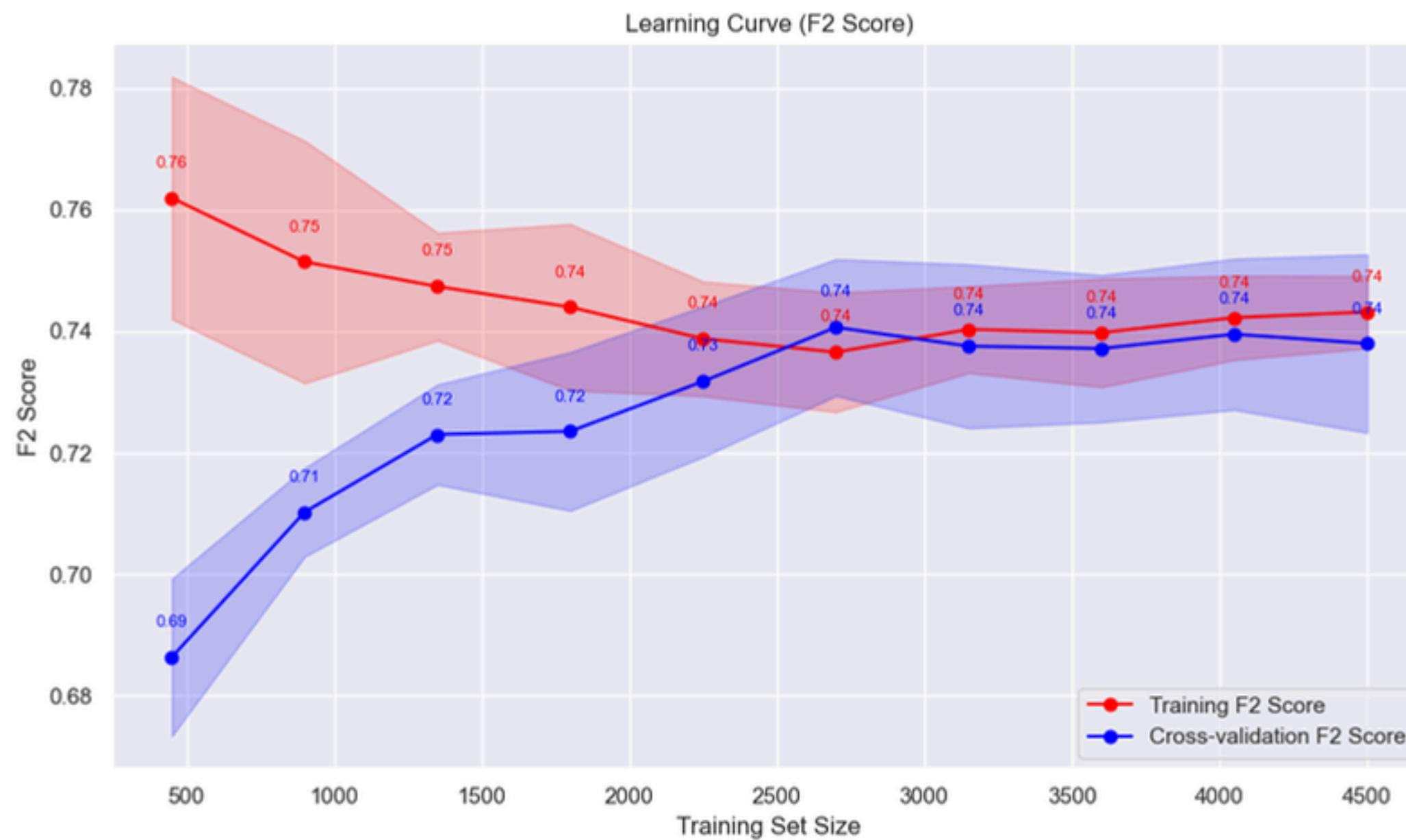
1. **Encoding** kolom kategorik menggunakan **One Hot Encoding**.
2. **Scaling** kolom numerik dengan **Robust Scaler**

# Model Benchmarking

	Mean F2 Score	Standar Deviasi	All Score
<b>Borderline SMOTE + Logreg</b>	0.738200	0.013	0.73, 0.76, 0.72, 0.73, 0.75
<b>SMOTENN + Adaboost</b>	0.737689	0.012	0.74, 0.75, 0.72, 0.74, 0.74
<b>Borderline SMOTE + Bagging</b>	0.737615	0.013	0.73, 0.76, 0.72, 0.73, 0.74
<b>Logreg</b>	0.566991	0.023	0.56, 0.59, 0.55, 0.59, 0.53
<b>Bagging (estimator: logreg)</b>	0.565802	0.020	0.56, 0.59, 0.55, 0.59, 0.54
<b>Stacking</b>	0.556912	0.019	0.56, 0.58, 0.53, 0.58, 0.54

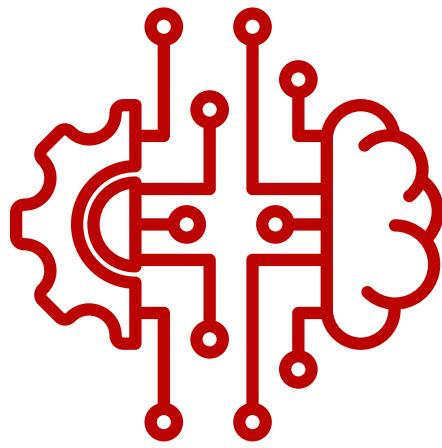
- **Resampling** berhasil **meningkatkan** performa model, terbukti dari peningkatan **F2 score** semua model.
- Model Logistic Regression + Borderline SMOTE menjadi model terbaik dengan F2 Score tertinggi **0.73820**, menunjukkan bahwa model bekerja lebih optimal setelah penyeimbangan data.

# Learning Curve



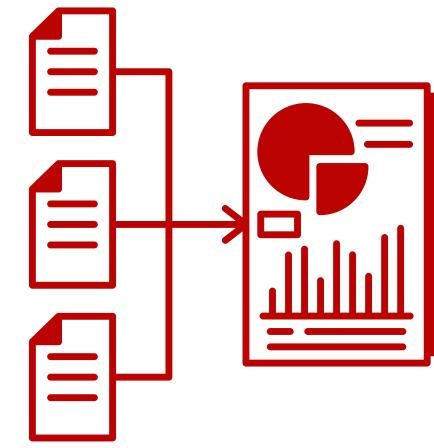
- **Model stabil & generalis**, gap train-test sudah kecil.
- **Tidak** ada sinyal **underfitting**
- **Overfitting tidak signifikan**. Overfitting hanya di dataset kecil. Pada dataset besar, training & CV F2 Score mendekati.

# Model Terbaik



## Logistic Regression

Merupakan model prediktif yang digunakan untuk klasifikasi biner. **Logreg memodelkan probabilitas sebuah output sebagai fungsi logistik dari input fitur**. Model ini tidak menghasilkan kelas langsung, tetapi probabilitas, yang kemudian dikonversi menjadi kelas.



## Cara Kerja?

### 1. Linear Combination (Logit Function)

Fungsi logit menghitung kombinasi linear dari fitur:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

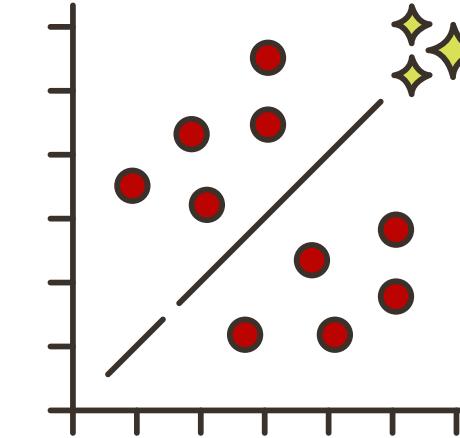
### 2. Fungsi sigmoid

Mengubah nilai z menjadi probabilitas antara 0-1:

$$\hat{y} = \sigma(z) = \frac{1}{1 + e^{-z}}$$

### 3. Prediksi

- Jika  $\hat{y} \geq 0.5$ , prediksi = 1
- Jika  $\hat{y} < 0.5$ , prediksi = 0



## Borderline SMOTE?

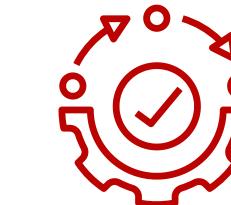
BorderlineSMOTE **fokus** pada **data minoritas** yang berada **di dekat batas klasifikasi (borderline)**. Titik-titik ini dianggap paling rentan terhadap salah klasifikasi, sehingga algoritma hanya membuat data sintetis di sekitar mereka.

# Hyperparameter Tuning Model



Tujuan

Meningkatkan **performa model** dengan mencari **kombinasi parameter terbaik**

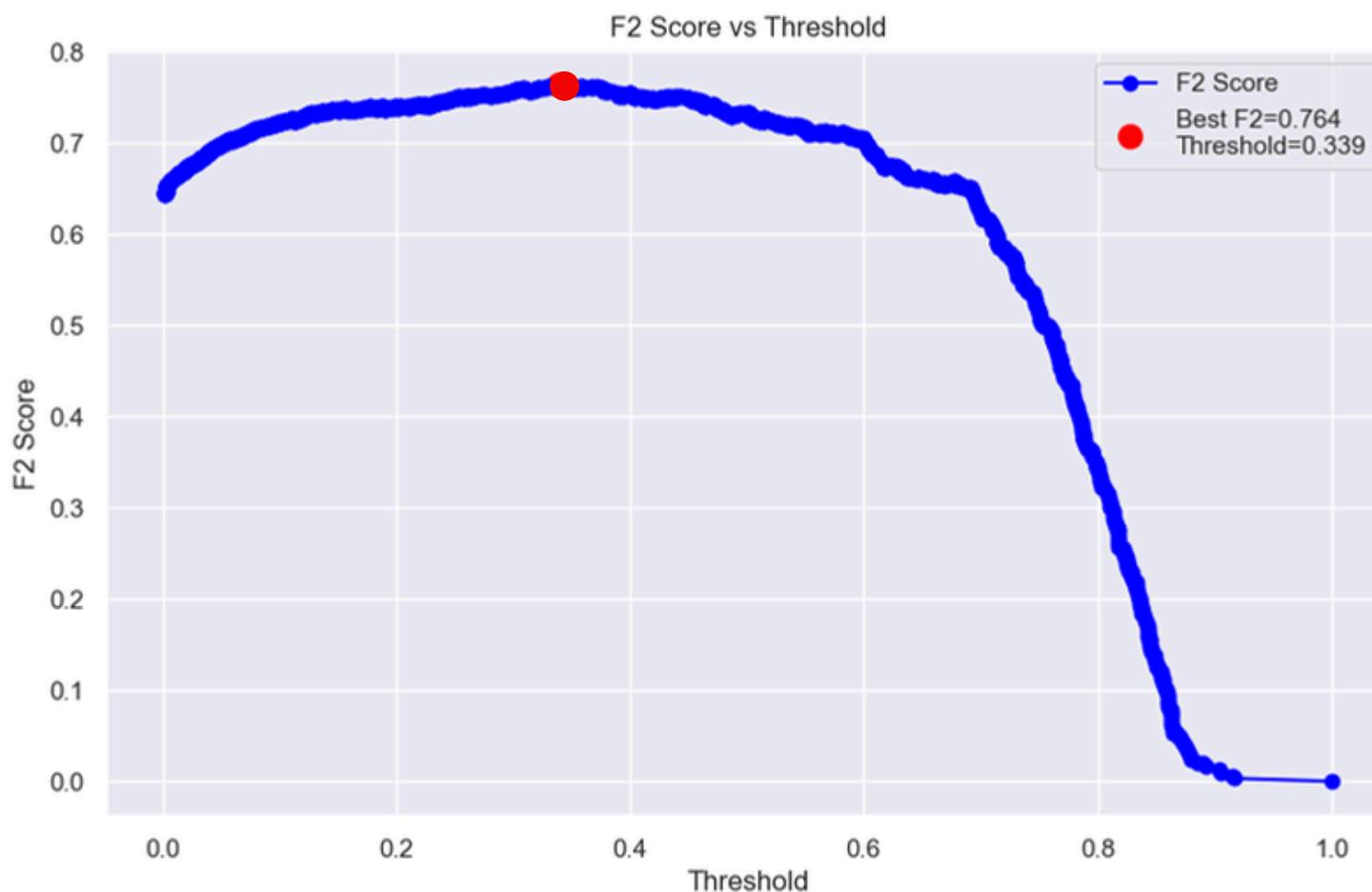


Metode

**Randomized Search** adalah metode pencarian hyperparameter secara acak untuk **menemukan kombinasi parameter terbaik**. Randomized Search hanya mencoba sejumlah kombinasi **secara acak dari ruang parameter yang telah ditentukan**

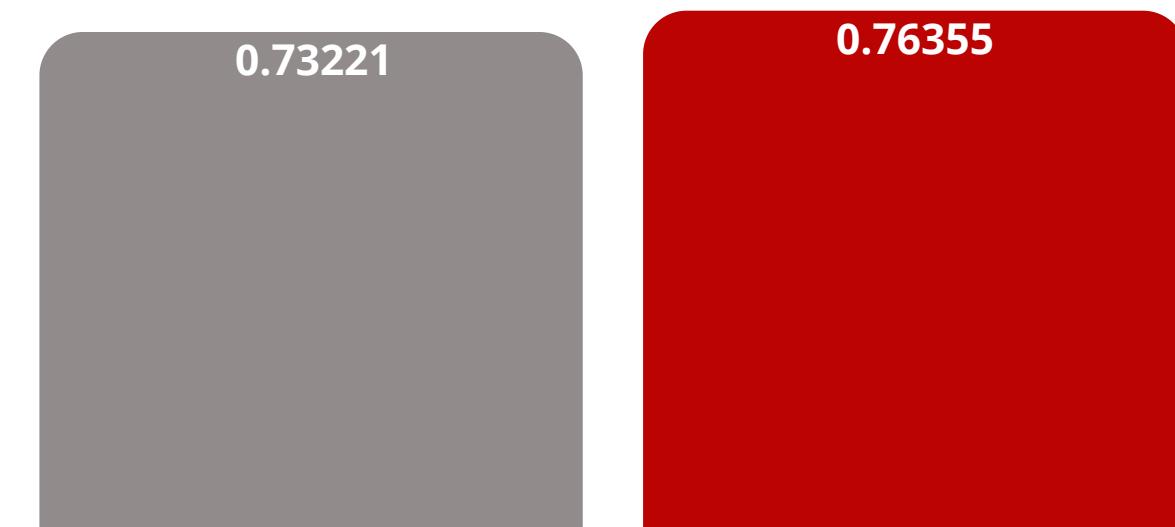
Best Params	Mean F2 Score	Std F2 Score	All Score
<ul style="list-style-type: none"><li>• <b>solver = saga</b></li><li>• <b>penalty = l1</b></li><li>• <b>C = 1</b></li></ul>	0.7391	0.013	0.72, 0.76, 0.72, 0.73, 0.74

# Threshold Probability Adjustment



F2 Score Pada Threshold Berbeda

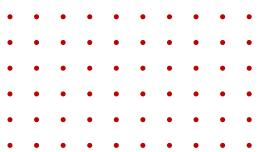
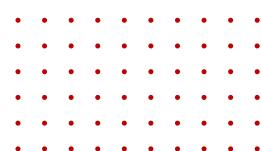
● Threshold 0.5      ● Threshold 0.3220



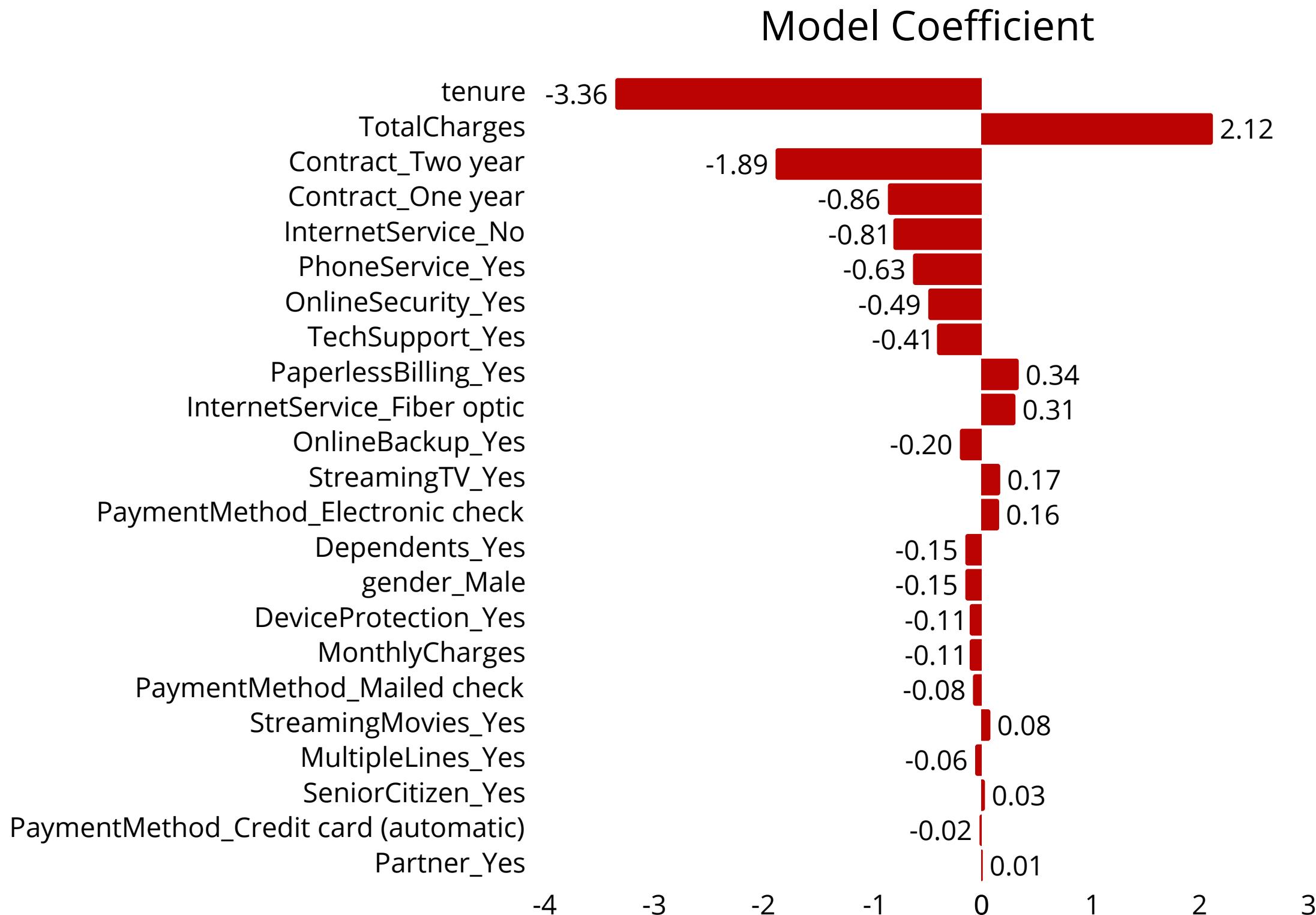
- Threshold optimal ditemukan di **0.3385** dengan **F2 Score 0.7636**, lebih baik dibanding threshold default 0.5 (F2 Score: 0.73221).
- Penyesuaian threshold ini meningkatkan kemampuan model dalam menangkap pelanggan yang berpotensi churn.



# Model Interpretation



# Model Interpretation



**Makna Koefisien:**

**Koefisien menunjukkan pengaruh suatu fitur terhadap logit (log-odds) dari target.**

- Positif → kenaikan fitur meningkatkan peluang ke kelas 1 (positif).
- Negatif → kenaikan fitur menurunkan peluang ke kelas 1.
- Nol → fitur tidak berpengaruh terhadap logit (log-odds).

**Interpretasi secara probabilitas:**

$$\text{odds ratio} = e^{\beta}$$

- $> 1 \rightarrow$  fitur meningkatkan kemungkinan kelas positif (misalnya, churn).
- $< 1 \rightarrow$  fitur menurunkan kemungkinan kelas positif.

**Contoh:**

Koefisien Tenure = -3.36

Odds ratio Tenure = 0.035

**Setiap kenaikan 1 unit tenure, peluang churn hanya 3.5% dari sebelumnya (menurun tajam).**

Koefisien TotalCharges = 2,12

Odds ratio Tenure = 8.33

**Setiap kenaikan 1 unit TotalCharges, peluang churn meningkat 833% (8 kali lipat) dari sebelumnya.**

# Perhitungan Kerugian Bisnis



\$ 694



\$ 139

## Tanpa Machine Learning



1033  
Not Churn



374  
Churn

Kerugian = Jumlah Customer Churn \* CAC

Kerugian tanpa ML  
**\$ 259,556**

## Hyperparameter Tuning

		Predicted	
		0	1
Actual	0	693 True Negative	340 False Positive
	1	59 False Negative	315 True Positive

Kerugian = (FP \* CRC) + (FN \* CAC)

Kerugian dengan Model HT  
**\$ 88,206**

## Hyperparameter Tuning & Threshold Adjustment

		Predicted	
		0	1
Actual	0	576 True Negative	457 False Positive
	1	22 False Negative	352 True Positive

Kerugian = (FP \* CRC) + (FN \* CAC)

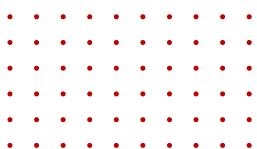
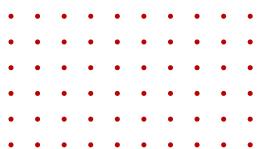
Kerugian dengan Model HT dan TA  
**\$ 78,791**

Aplikasi model ML yang telah dilakukan Hyperparameter Tuning dan Adjustment Threshold dapat menekan kerugian mencapai

↓ 69%



# Kesimpulan & Rekomendasi



# Business Conclusions

01

Pelanggan yang **cenderung Churn** memiliki karakteristik sebagai berikut :

- Masa langganan pendek (10 bulan)
- TotalCharges rendah (\$703.55)
- MonthlyCharges tinggi (\$79.65)
- Kontrak bulanan / month-to-month (43%)
- Pembayaran dengan Electronic Check (45%)
- Tidak menggunakan layanan tambahan
- Senior citizen (42%)
- Tanpa pasangan (33%), tanpa dependents (31%)
- Menggunakan paperless billing (34%)

02

Pelanggan yang **cenderung bertahan** memiliki karakteristik sebagai berikut:

- Masa langganan panjang (38 bulan)
- TotalCharges tinggi (\$1683.60)
- MonthlyCharges lebih rendah (\$64.45)
- Kontrak 2 tahun (97%) atau 1 tahun (89%)
- Pembayaran otomatis: Credit Card (85%), Bank Transfer (83%)
- Menggunakan layanan tambahan seperti OnlineSecurity (85%), TechSupport (85%), OnlineBackup (78%), DeviceProtection (77%)
- Menghindari paperless billing (84%)
- Memiliki pasangan (80%), memiliki dependents (84%)
- Bukan senior citizen (76%)

01

Fokus pada pelanggan baru di 12 bulan pertama

02

Sesuaikan tarif bulanan bagi pelanggan rentan di awal masa langganan

03

Dorong migrasi ke kontrak jangka panjang

04

Alihkan metode pembayaran dari Electronic Check ke sistem otomatis

05

Bundling layanan keamanan sebagai nilai tambah

06

Evaluasi ulang value proposition dari layanan Fiber Optic

## **Business Recommendations**



# Machine Learning Conclusions

- 1 Penggunaan model terbaik (Bagging estimator logistic regression) mampu menghasilkan **F2-Score sebesar 0.76.**
- 2 **Model** mampu **mengurangi kerugian finansial sebesar 69%** (\$180,765) dandingkan tidak menggunakan model.
- 3 Faktor utama churn meliputi: **tenure, total charges, contract type, dan layanan tambahan.**



## Machine Learning Recommendations

- 1 Implementasikan **model prediksi churn** sebagai **sistem peringatan dini**.
- 2 Gunakan **threshold bisnis** (tidak default 0.5) untuk memaksimalkan **F2-Score**.
- 3 Lakukan **monitoring berkala** terhadap **performa model** (drift check).

# LIMITASI MODEL



Tidak semua fitur penting **tersedia** (misalnya: interaksi customer service, feedback pelanggan)



Model **belum diuji** dalam skenario produksi **aktual**



AHA Consulting Group

# TERIMA KASIH

