

Reducing Customer Acquisition Costs and Improving Retention Strategies in Telecommunications

“Leveraging Descriptive and Predictive Machine Learning Insights”

Analyzed by Rofi Rosdiani and Dafa Muhammad Althaf



Table of Contents



01 Business Problem Understanding

...



02 Data Understanding & Cleaning

...



03 Data Analysis

...



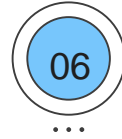
04 Data Preprocessing

...



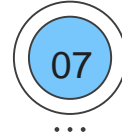
05 Modelling and Evaluation

...



06 Conclusion & Recommendation

...



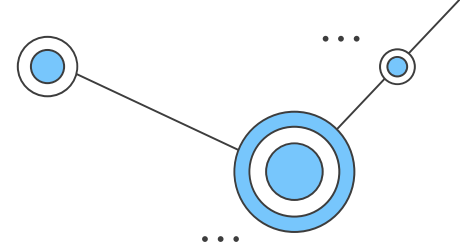
07 Business Implementation

...



08 References

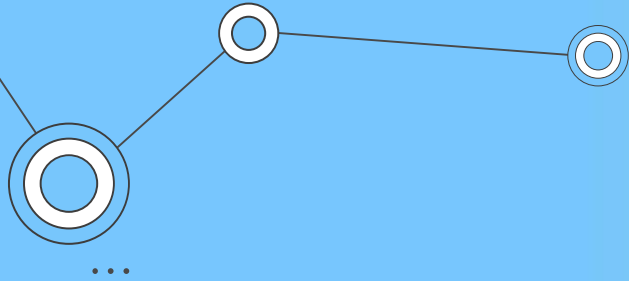
...



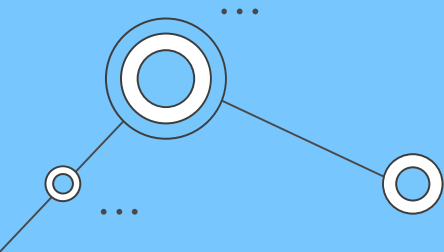


01

**Business Problem
Understanding**



Dalam dunia telekomunikasi, **churn** adalah **fenomena berakhirnya hubungan kontraktual atau non-kontraktual antara pelanggan dan penyedia layanan dalam periode tertentu**, diukur sebagai persentase dari total pelanggan.



Understanding the Problem



Problem Statement

Biaya akuisisi pelanggan baru **5 – 25x lebih besar** daripada mempertahankan pelanggan lama.

...



Goals

- Mengidentifikasi faktor-faktor pengaruh churn
- Memprediksi pelanggan yang akan churn/loyal untuk menghindari biaya akuisisi

...

Analytic Approach

- Menemukan pola – pola yang membedakan antara pelanggan churn dan tidak churn (loyal)
- Membangun model klasifikasi untuk memprediksi probabilitas pelanggan yang churn dan tidak
- Menggunakan beberapa algoritma klasifikasi, seperti:
 - Logistic Regression
 - Decision Tree
 - K-Nearest Neighbors (KNN)
 - Random Forest
 - XGBoost
 - LightGBM

Stakeholders yang terlibat

Manajemen Eksekutif

Mengambil keputusan strategis berbasis data churn, seperti alokasi anggaran retensi atau pengembangan produk

Tim Pemasaran

Menggunakan hasil prediksi untuk membuat program retensi yang lebih terarah kepada pelanggan berisiko.



Metric Evaluation

		PREDICTED	
		Churn (1)	Non Churn (0)
ACTUAL	Churn (1)	TRUE POSITIVE (TP)	FALSE NEGATIVE (FN)
		Model memprediksi churn, dan memang benar churn	Model prediksi tidak churn, tapi ternyata churn
	Non Churn (0)	FALSE POSITIVE (FP)	TRUE NEGATIVE (TN)
		Model prediksi churn, padahal tidak churn	Model prediksi tidak churn, dan memang tidak churn

- **True Positive (TP)** : Model memprediksi customer akan churn dan memang benar churn
- **True Negative (TN)** : Model memprediksi customer tidak akan churn dan memang tidak churn
- **Type 1 error (False Positive/FP)** : Model memprediksi customer akan churn, padahal sebenarnya tidak churn.
 - Konsekuensi: Perusahaan mengalokasikan biaya retensi (diskon, promo, insentif) kepada customer yang sebenarnya loyal. Ini menyebabkan pemborosan biaya karena strategi retensi diterapkan pada customer yang tetap akan bertahan tanpa insentif.
- **Type 2 error (False Negative/FN)** : Model memprediksi customer tidak akan churn, padahal sebenarnya churn.
 - Konsekuensi : Customer yang benar-benar akan churn tidak mendapatkan perhatian atau intervensi, sehingga:
 - Customer benar-benar pergi tanpa upaya pencegahan kemudian berdampak pada hilangnya pendapatan
 - Perusahaan harus mengeluarkan biaya akuisisi yang lebih besar untuk mencari customer baru.



02

Data Understanding & Data Cleaning

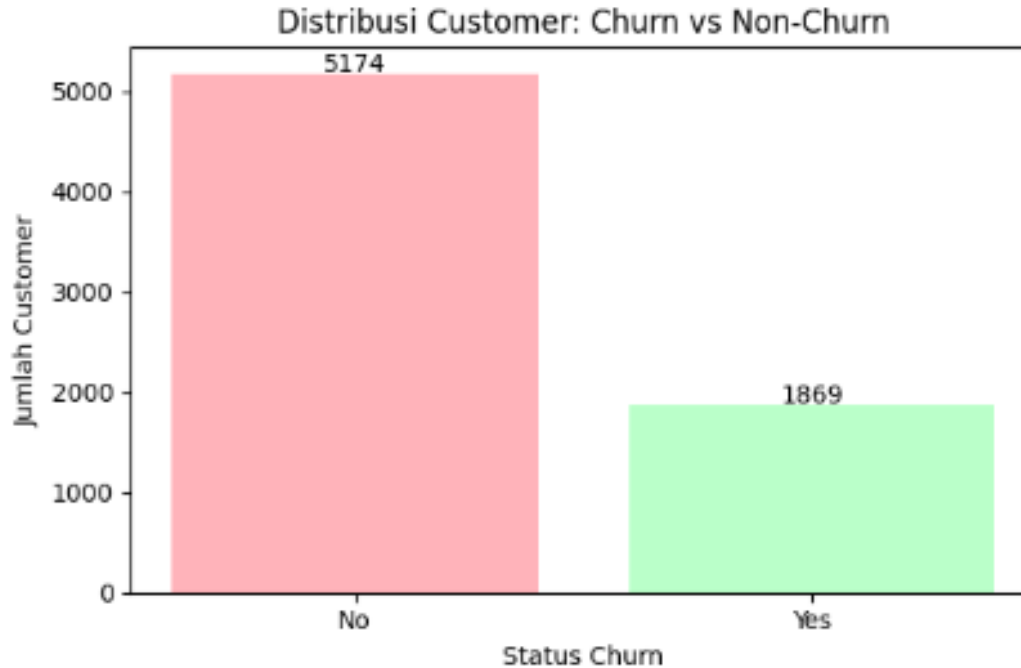
Data Understanding

Kolom	Deskripsi	Tipe Data Awal	Tipe Data Seharusnya
customerID	ID unik tiap pelanggan	object	object (string)
gender	Jenis kelamin pelanggan (Male / Female)	object	category
SeniorCitizen	Apakah pelanggan adalah warga senior (1 = ya, 0 = tidak)	int64	category
Partner	Apakah pelanggan memiliki pasangan	object	category
Dependents	Apakah pelanggan memiliki tanggungan	object	category
tenure	Lama berlangganan (dalam bulan)	int64	int64
PhoneService	Apakah pelanggan berlangganan layanan telepon	object	category
MultipleLines	Apakah pelanggan memiliki lebih dari satu jalur telepon	object	category
InternetService	Jenis layanan internet	object	category
OnlineSecurity	Layanan keamanan online	object	category
OnlineBackup	Layanan pencadangan online	object	category
DeviceProtection	Proteksi perangkat	object	category
TechSupport	Dukungan teknis	object	category
StreamingTV	Layanan streaming TV	object	category
StreamingMovies	Layanan streaming film	object	category
Contract	Jenis kontrak	object	category
PaperlessBilling	Apakah pelanggan menggunakan tagihan tanpa kertas	object	category
PaymentMethod	Metode pembayaran	object	category
MonthlyCharges	Biaya bulanan (\$)	float64	float64
TotalCharges	Total biaya yang telah dibayarkan (\$)	object	float64
Churn	Status churn	object	category

Summary:

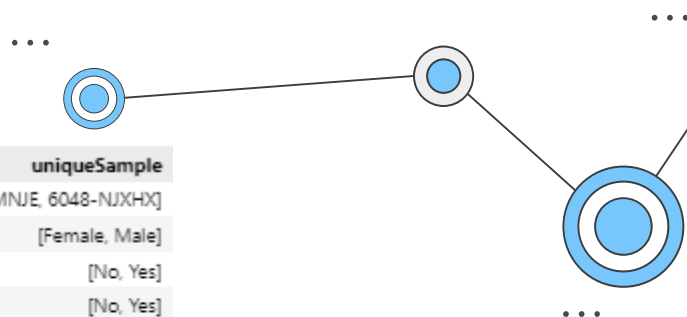
1. Dataset terdiri dari **7.043 baris** dan **21 kolom fitur**
2. Terdapat **3 fitur numerik** dan **18 fitur lainnya bertipe kategorik**
3. Target variabel: **`Churn`** (apakah pelanggan berhenti atau tidak)

Pengecekan Data Balance/Imbalance



Dataset bersifat imbalanced, di mana proporsi kelas positif (customer yang churn atau berhenti) hanya sebesar 26.54%, sedangkan proporsi kelas negatif (customer yang non-churn atau loyal) mencapai 73.46%.

Data Cleaning



	dataFeatures	dataType	null	nullPct	unique	uniqueSample
0	customerID	object	0	0.0	7043	[2907-ILJBN, 5014-WUQMG, 1942-OQFRW, 5989-OMNJE, 6048-NJXHx]
1	gender	category	0	0.0	2	[Female, Male]
2	SeniorCitizen	category	0	0.0	2	[No, Yes]
3	Partner	category	0	0.0	2	[No, Yes]
4	Dependents	category	0	0.0	2	[Yes, No]
5	tenure	int64	0	0.0	73	[56, 16, 48, 13, 70]
6	PhoneService	category	0	0.0	2	[No, Yes]
7	MultipleLines	category	0	0.0	3	[Yes, No, No phone service]
8	InternetService	category	0	0.0	3	[Fiber optic, DSL, No]
9	OnlineSecurity	category	0	0.0	3	[No internet service, Yes, No]
10	OnlineBackup	category	0	0.0	3	[No internet service, No, Yes]
11	DeviceProtection	category	0	0.0	3	[No internet service, No, Yes]
12	TechSupport	category	0	0.0	3	[No internet service, No, Yes]
13	StreamingTV	category	0	0.0	3	[No, Yes, No internet service]
14	StreamingMovies	category	0	0.0	3	[No, No internet service, Yes]
15	Contract	category	0	0.0	3	[Two year, Month-to-month, One year]
16	PaperlessBilling	category	0	0.0	2	[Yes, No]
17	PaymentMethod	category	0	0.0	4	[Credit card (automatic), Electronic check, Bank transfer (automatic), Mailed check]
18	MonthlyCharges	float64	0	0.0	1585	[73.75, 35.65, 20.0, 59.55, 20.1]
19	TotalCharges	object	0	0.0	6531	[354.05, 3382.3, 585.95, 147.8, 555.4]
20	Churn	category	0	0.0	2	[Yes, No]

Summary

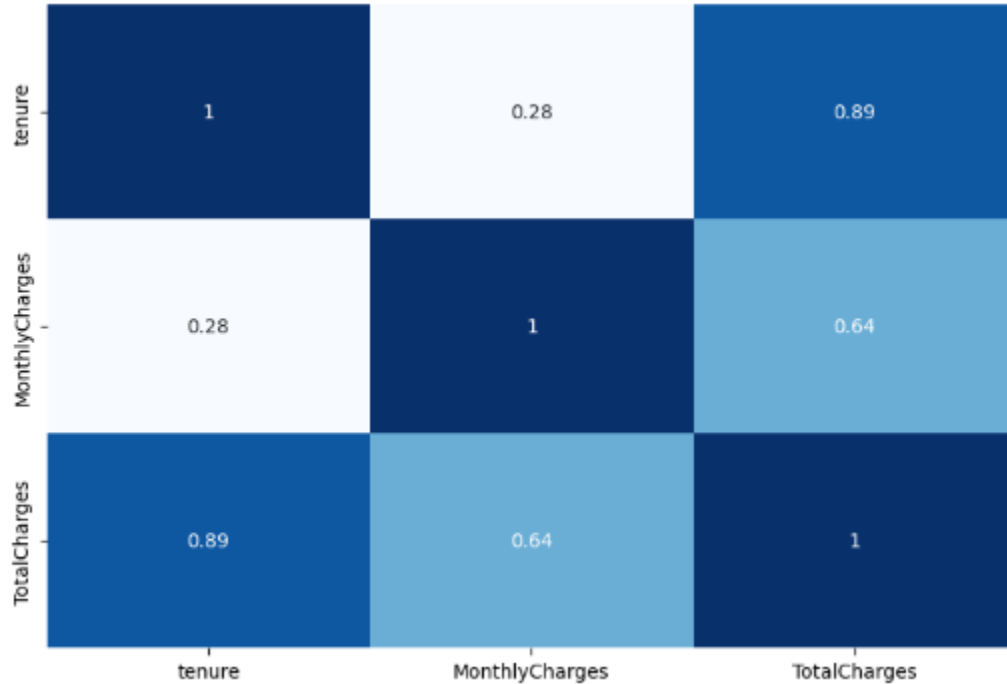
- Tidak terdapat duplikasi data dan missing values
- Ditemukan 11 baris data dengan `tenure = 0` dan Total Charges dengan nilai spasi → Dihapus
- Tidak ditemukan nilai yang berada di luar batas bawah dan batas atas (outlier)
- Setelah dilakukan cleaning, maka saat ini dataset berjumlah 7032 baris



03

Data Analysis

Pengecekan Korelasi Numerikal Fitur



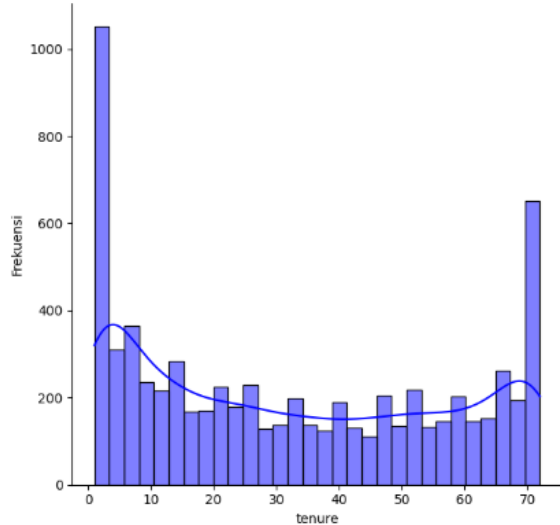
SUMMARY

- Variabel numerik tenure, MonthlyCharges, dan TotalCharges menunjukkan hubungan **korelasi positif** satu sama lain.
- Korelasi antara tenure dan TotalCharges **sangat kuat**, dengan nilai sebesar 0.89.
- MonthlyCharges memiliki **korelasi sedang** terhadap TotalCharges (0.64) dan **korelasi rendah** terhadap tenure (0.28).

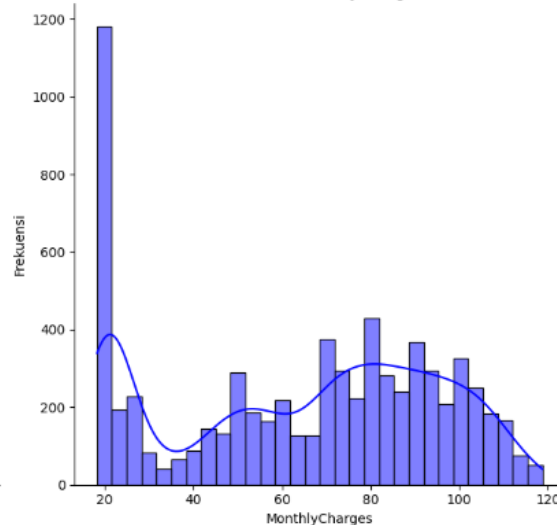


Pengecekan Distribusi Data

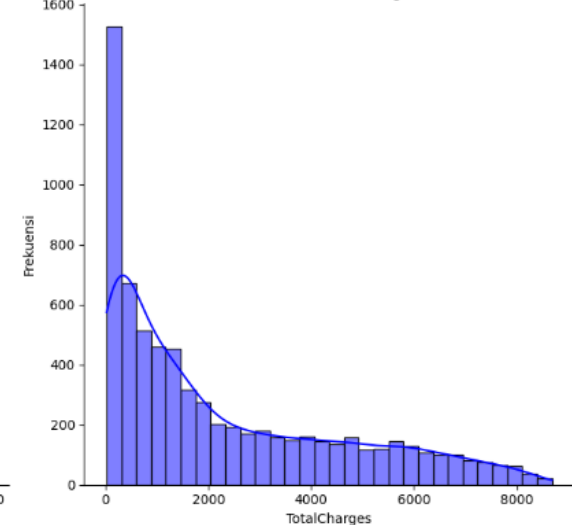
Distribusi tenure



Distribusi MonthlyCharges



Distribusi TotalCharges



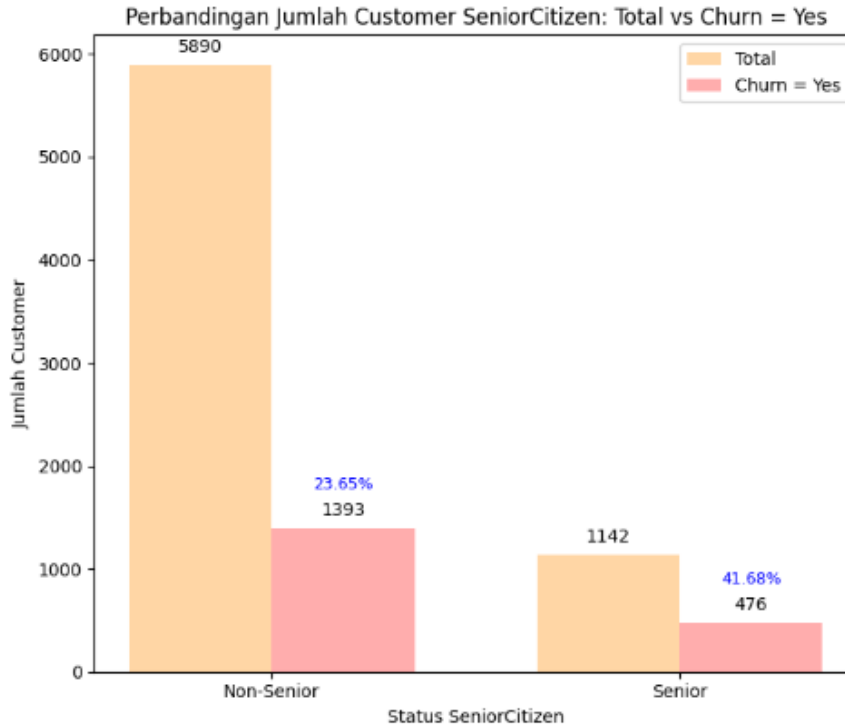
- **Tenure:** Distribusi tenure menunjukkan pola U-shape, dengan dua kelompok utama pelanggan: yang baru bergabung dan yang sudah sangat lama berlangganan. Pelanggan dengan masa langganan menengah relatif lebih sedikit.
- **MonthlyCharge:** Distribusi tidak simetris dan tampak bimodal (memiliki dua puncak).
 - Puncak pertama di sekitar \$20, menunjukkan banyak pelanggan yang membayar biaya bulanan rendah—kemungkinan karena hanya menggunakan layanan dasar seperti telepon/internet tanpa tambahan.
 - Puncak kedua menyebar antara \$70–\$100, menunjukkan kelompok pelanggan yang menggunakan lebih banyak layanan atau layanan premium.
- **TotalCharge:** Distribusi ini sangat miring ke kanan (right-skewed). Sebagian besar pelanggan memiliki total tagihan rendah, terkonsentrasi di bawah \$2000

Bivariate Analysis: Categorical Features vs Target (Churn)

Senior Citizen vs Churn

- Mayoritas pelanggan bukan warga senior (5890 dari 7032)
- Warga senior lebih mungkin churn dibanding non-senior:
 - Non-senior: hanya 23.65% yang churn.
 - Senior: 41.68% yang churn — hampir dua kali lipat lebih tinggi dari non-senior.

Kesimpulan: Warga senior memiliki kecenderungan lebih tinggi untuk berhenti berlangganan. Ini bisa jadi sinyal bahwa mereka butuh perhatian khusus, misalnya dalam hal layanan pelanggan atau paket yang lebih sesuai kebutuhan.

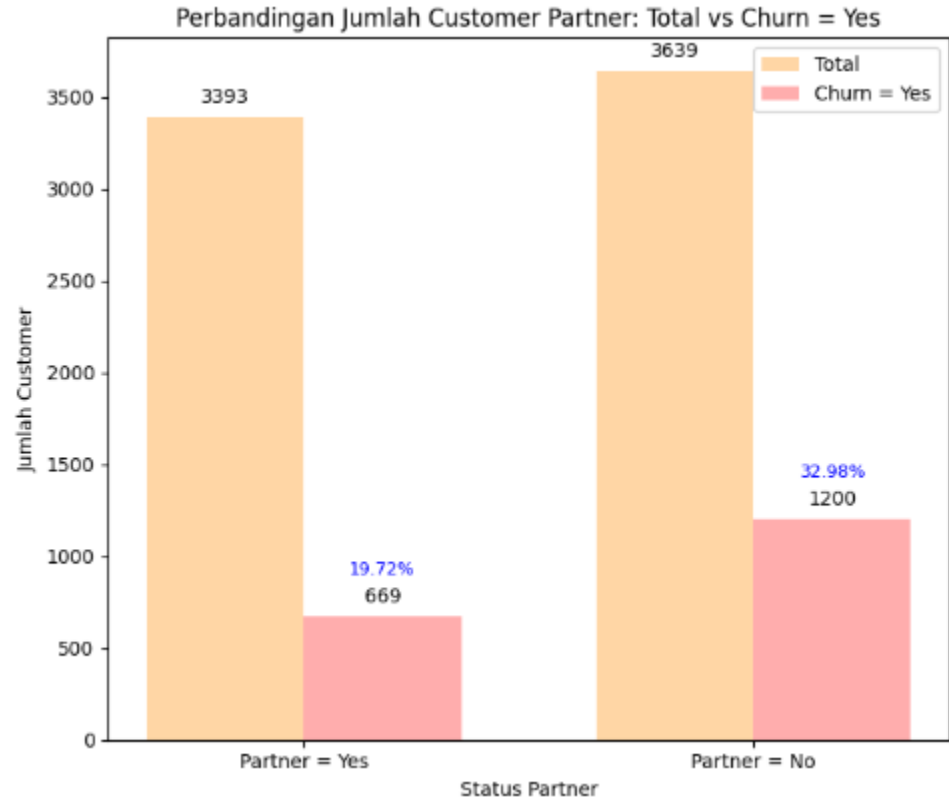


Bivariate Analysis: Categorical Features vs Target (Churn)

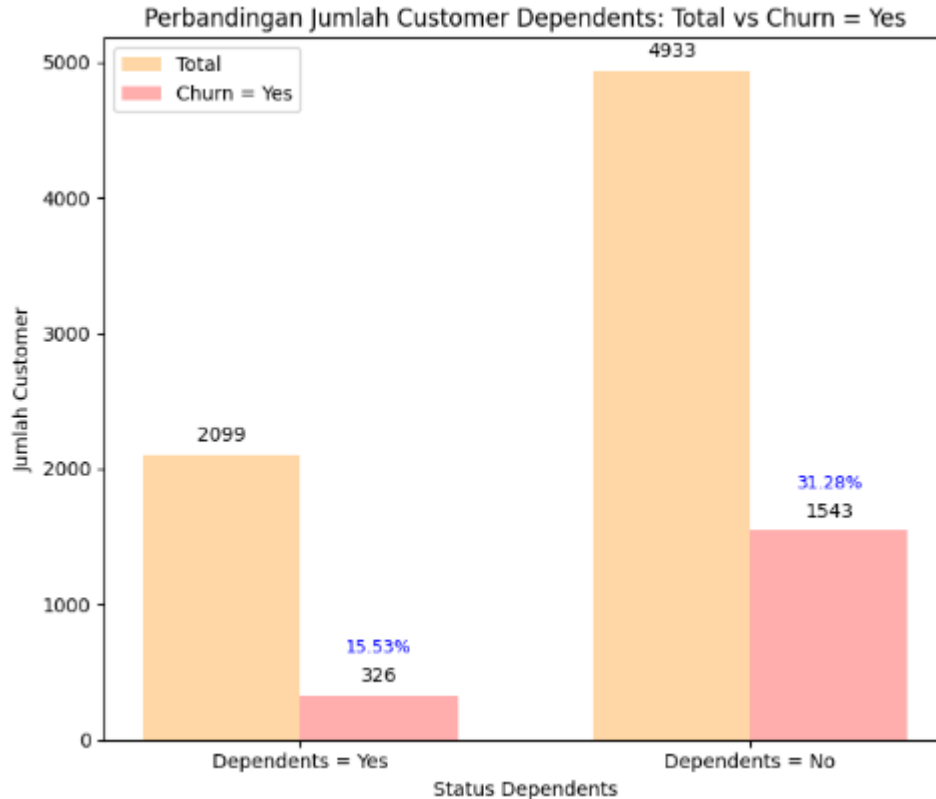
Partner vs Churn

- Jumlah pelanggan terbagi cukup merata antara yang punya pasangan (Partner = Yes) dan tidak.
- Pelanggan tanpa pasangan lebih mungkin churn:
 - Mempunyai pasangan: hanya 19.72% yang churn.
 - Tanpa pasangan: 32.98% yang churn.

Kesimpulan: Pelanggan yang tidak memiliki pasangan memiliki risiko churn yang lebih tinggi. Mungkin karena mereka lebih sensitif terhadap biaya atau kurang terikat secara emosional dengan layanan. Ini bisa jadi peluang untuk pendekatan personal atau penawaran khusus.



... Bivariate Analysis: Categorical Features vs Target (Churn)



Dependents vs Churn

- Sebagian besar pelanggan tidak memiliki tanggungan (lebih 2x banyak dibanding yang punya tanggungan).
- Pelanggan tanpa tanggungan jauh lebih mungkin churn:
 - Dengan tanggungan: hanya 15.53% yang churn.
 - Tanpa tanggungan: 31.28% yang churn — lebih dari dua kali lipat.

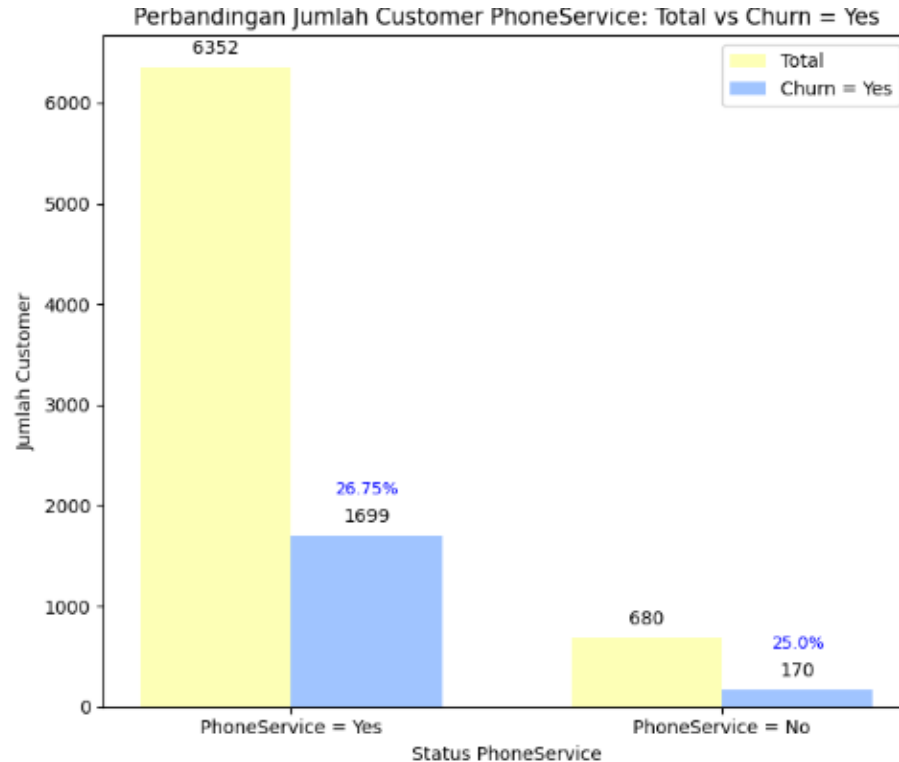
Kesimpulan: Pelanggan dengan tanggungan cenderung lebih setia, mungkin karena mereka lebih bergantung pada layanan untuk keluarga. Ini menunjukkan bahwa pelanggan tanpa tanggungan perlu perhatian lebih agar tidak churn.

Bivariate Analysis: Categorical Features vs Target (Churn)

Phone Service vs Churn

- Sebagian besar pelanggan menggunakan layanan telepon (lebih dari 9 dari 10 pelanggan memiliki PhoneService).
- Pelanggan dengan atau tanpa PhoneService memiliki tingkat churn yang hampir sama:
 - Dengan PhoneService: 26.75% churn.
 - Tanpa PhoneService: 25.0% churn.

Kesimpulan: PhoneService tidak memiliki pengaruh besar terhadap churn, karena selisih tingkat churn di kedua kelompok sangat kecil. Ini menunjukkan bahwa keberadaan layanan telepon bukanlah faktor utama dalam keputusan pelanggan untuk berhenti.



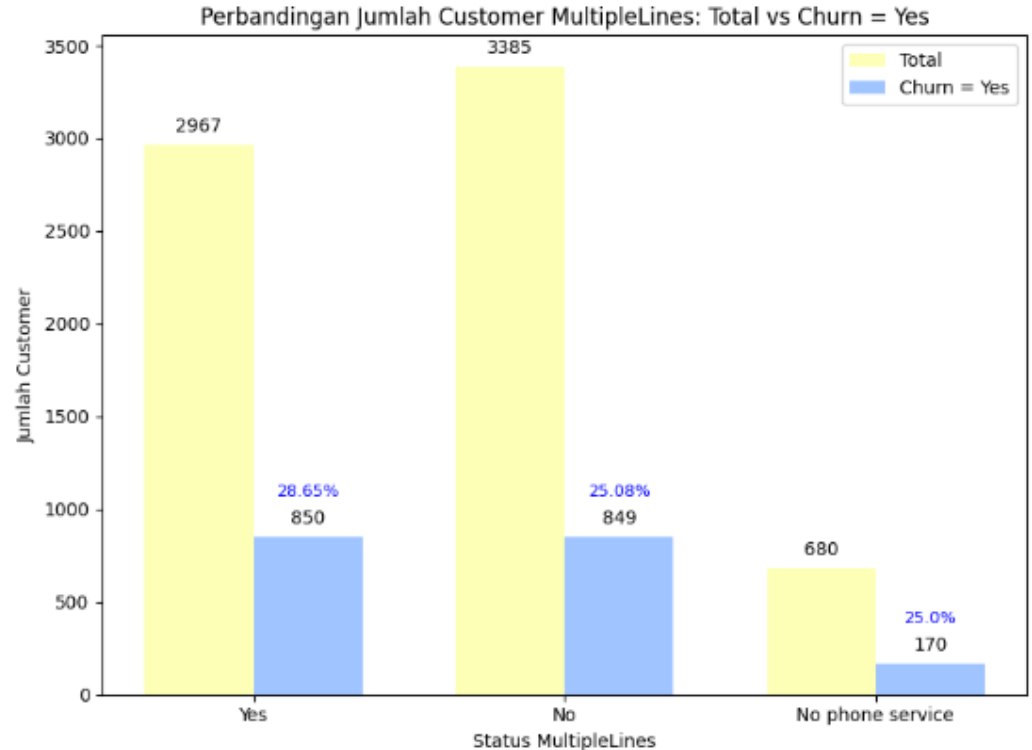
Multiple Lines vs Churn

➤ Pelanggan dengan layanan multiple lines memiliki tingkat churn paling tinggi:

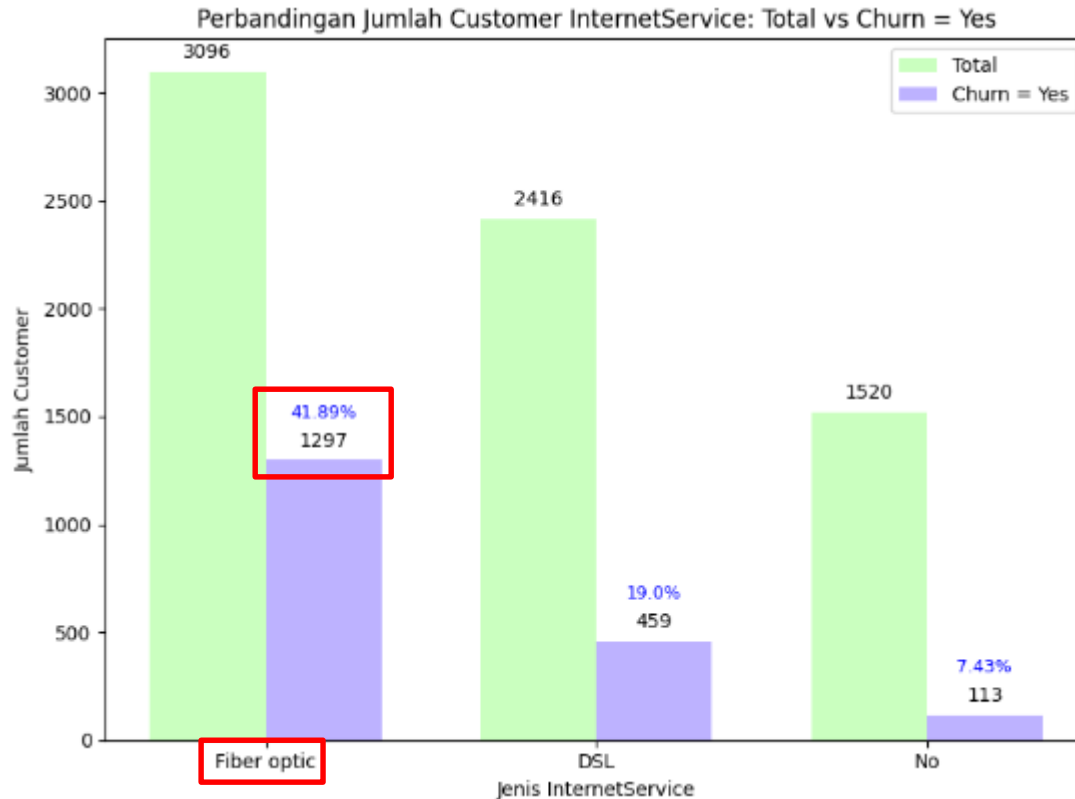
- **MultipleLines = Yes: 28.65%**
- MultipleLines = No: 25.08%
- Tidak punya layanan telepon (No phone service): 25.0%

Kesimpulan:

Memiliki multiple lines sedikit meningkatkan risiko churn dibanding pelanggan lain. Namun, perbedaan tidak terlalu besar, jadi ini mungkin bukan faktor dominan. Tetap bisa jadi sinyal bahwa pelanggan dengan lebih dari satu jalur telepon perlu lebih diperhatikan atau diberi insentif agar tetap bertahan.



Bivariate Analysis: Categorical Features vs Target (Churn)



Internet Service vs Churn

Pelanggan **Fiber optic** jauh lebih **rentan churn** dibanding yang lain - bahkan **2x lebih tinggi** dibanding **DSL** dan **hampir 6x lipat** dibandingkan yang tidak pakai internet.

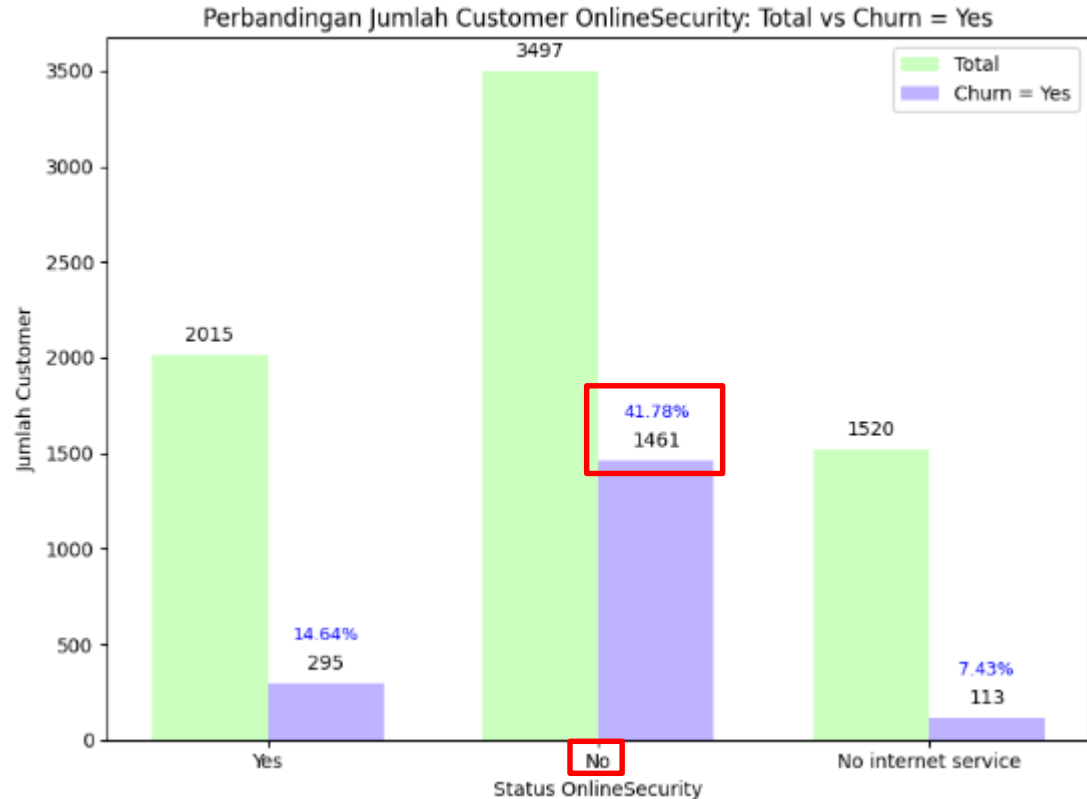
Kemungkinan karena:

- Harga Fiber optic lebih tinggi.
- Harapan pelanggan atas kualitas layanan juga lebih tinggi.

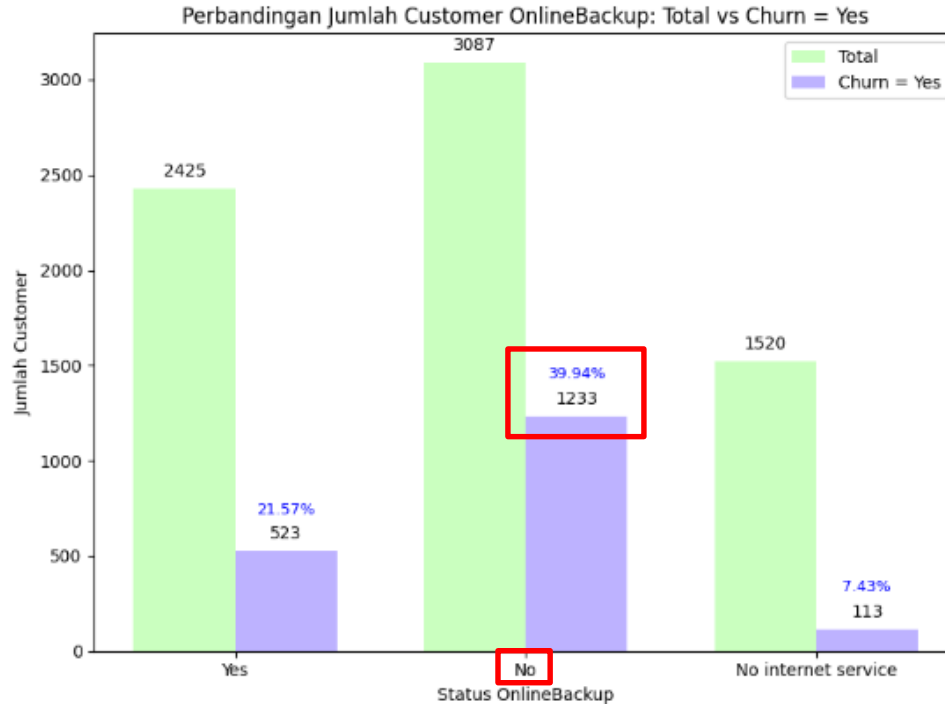
Bivariate Analysis: Categorical Features vs Target (Churn)

Online Security vs Churn

- **Sebagian besar pelanggan tidak menggunakan layanan OnlineSecurity** - lebih dari **1,5x lipat** dibanding yang menggunakannya.
- **Pelanggan yang tidak berlangganan OnlineSecurity sangat rentan churn (41.78%)**, kemungkinan karena merasa kurang aman atau tidak puas. Memberikan atau mempromosikan layanan keamanan online bisa menjadi strategi penting untuk mempertahankan pelanggan.



Bivariate Analysis: Categorical Features vs Target (Churn)



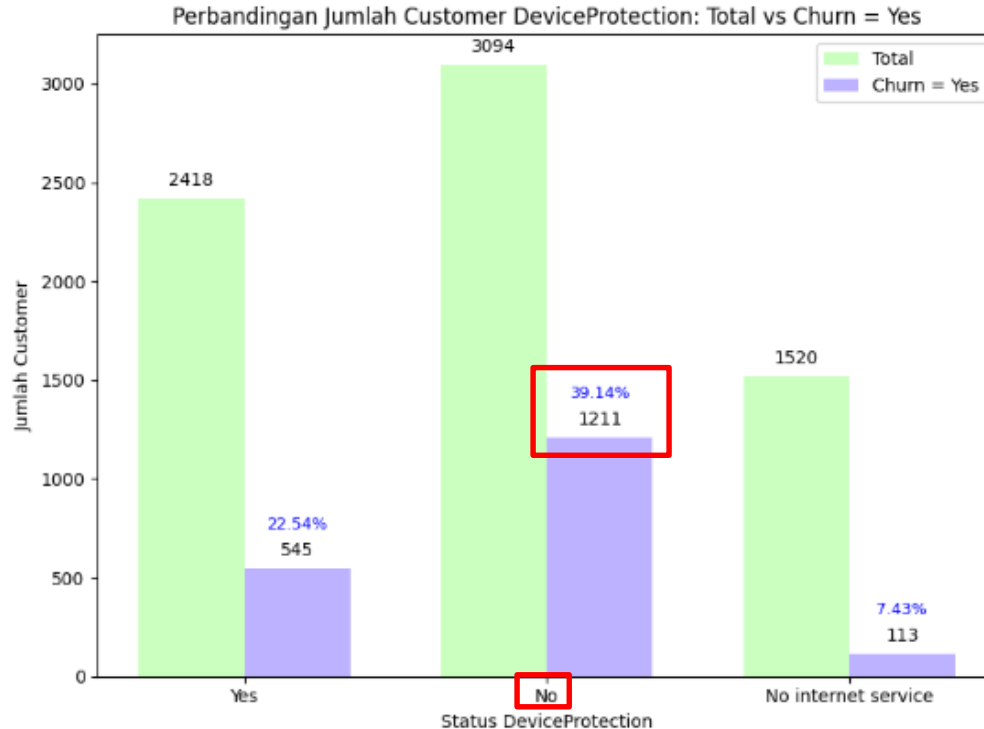
Online Backup vs Churn

- Lebih banyak pelanggan tidak menggunakan layanan OnlineBackup dibanding yang menggunakannya.
- Di antara pelanggan yang memiliki layanan internet, pelanggan yang menggunakan OnlineBackup cenderung churn lebih rendah (21.57%) dibanding yang tidak menggunakannya (39.94%). Hal ini menunjukkan bahwa layanan tambahan seperti OnlineBackup berpotensi meningkatkan retensi pelanggan internet.

... Bivariate Analysis: Categorical Features vs Target (Churn)

Device Protection vs Churn

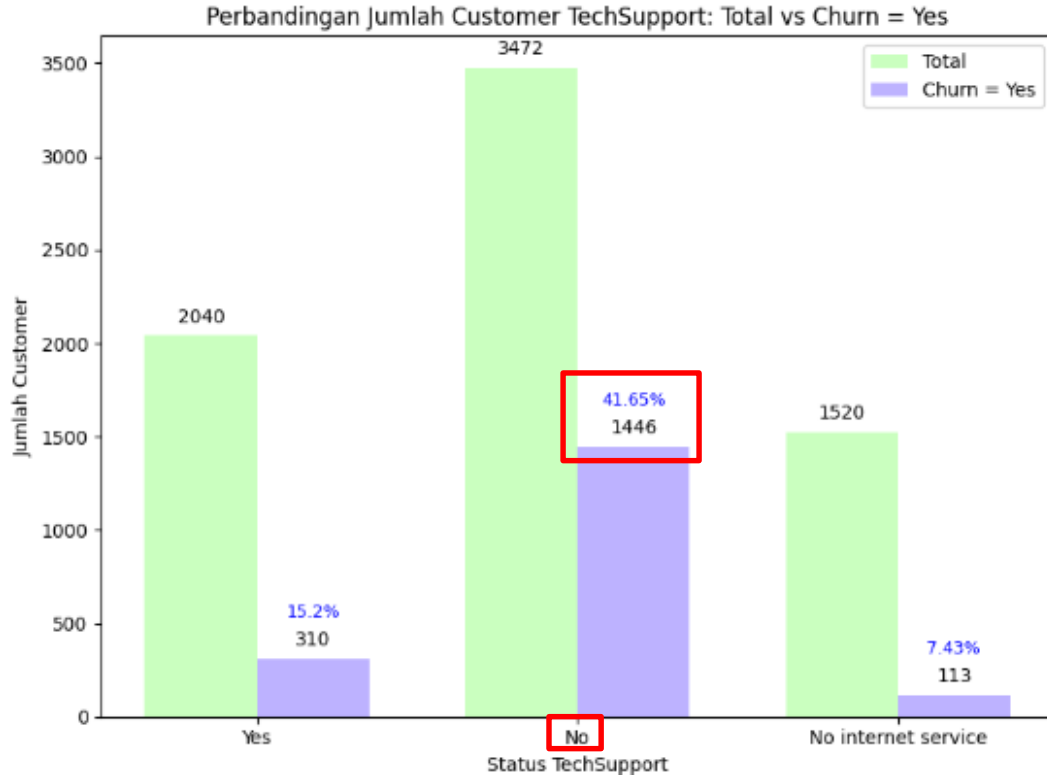
- Lebih banyak pelanggan tidak menggunakan DeviceProtection dibanding yang menggunakannya.
- Pelanggan yang tidak menggunakan layanan Device Protection lebih berisiko churn (39.14%), menunjukkan bahwa fitur keamanan seperti ini bisa menjadi faktor yang meningkatkan kepuasan dan retensi.



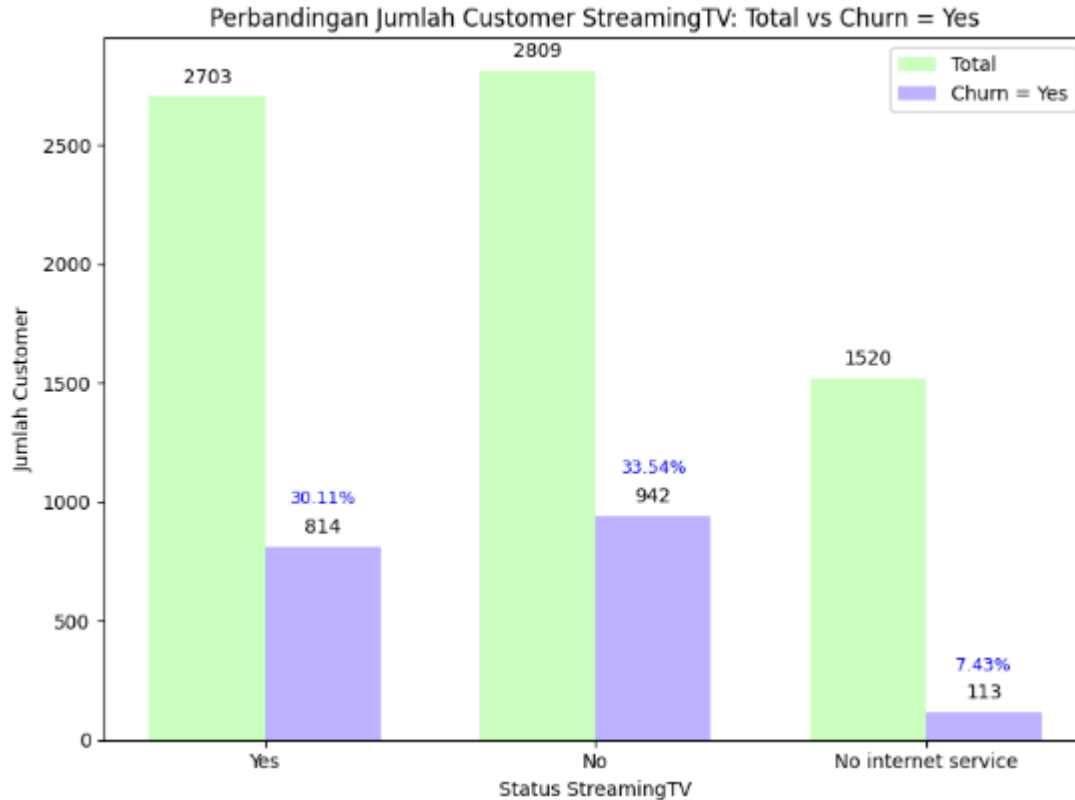
Bivariate Analysis: Categorical Features vs Target (Churn)

Tech Support vs Churn

- Sebagian besar pelanggan tidak menggunakan layanan TechSupport — hampir 2x lebih banyak dibanding yang menggunakannya.
- **Tidak adanya TechSupport berkorelasi kuat dengan churn.** Memberikan akses yang mudah ke layanan bantuan teknis bisa meningkatkan loyalitas pelanggan secara signifikan.



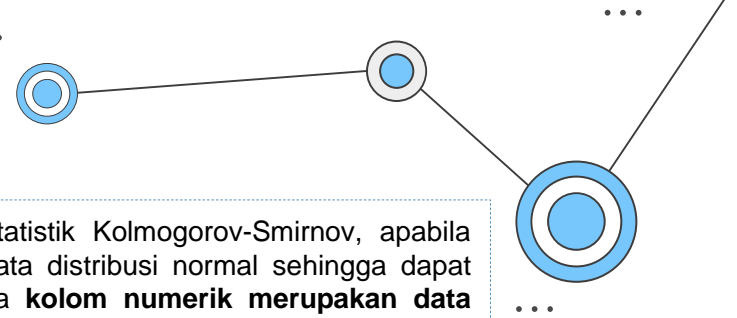
... Bivariate Analysis: Categorical Features vs Target (Churn)



Streaming TV vs Churn

- Jumlah pelanggan yang menggunakan dan tidak menggunakan layanan StreamingTV hampir seimbang
- **StreamingTV tampaknya tidak memiliki pengaruh besar terhadap churn** — perbedaan churn antara pengguna dan non-pengguna relatif kecil. Namun, tetap ada indikasi bahwa pelanggan yang memanfaatkan layanan hiburan lebih cenderung bertahan.

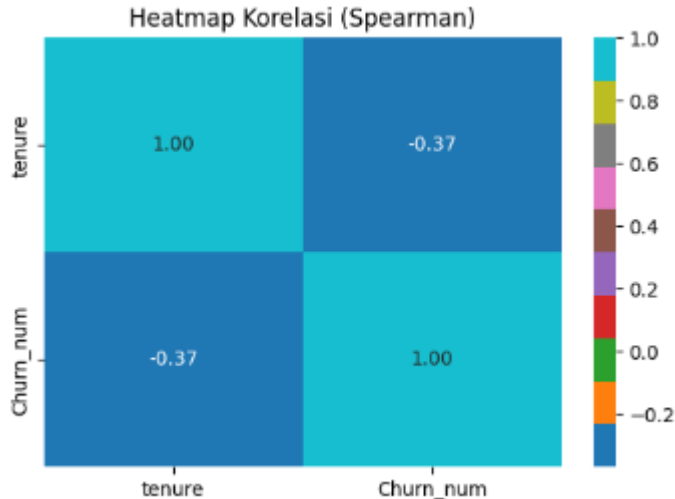
Melakukan Uji-Statistik: Fitur Tenure, Contract, and ... Monthly Charges terhadap Churn



	Kolom	KS_Statistic	P_Value	Distribusi_Normal
0	tenure	0.890077	0.0	Tidak
1	MonthlyCharges	1.000000	0.0	Tidak
2	TotalCharges	1.000000	0.0	Tidak

Berdasarkan hasil uji statistik Kolmogorov-Smirnov, apabila $p_value > 0,05$ maka data distribusi normal sehingga dapat dibuktikan bahwa semua **kolom numerik merupakan data distribusi tidak normal**

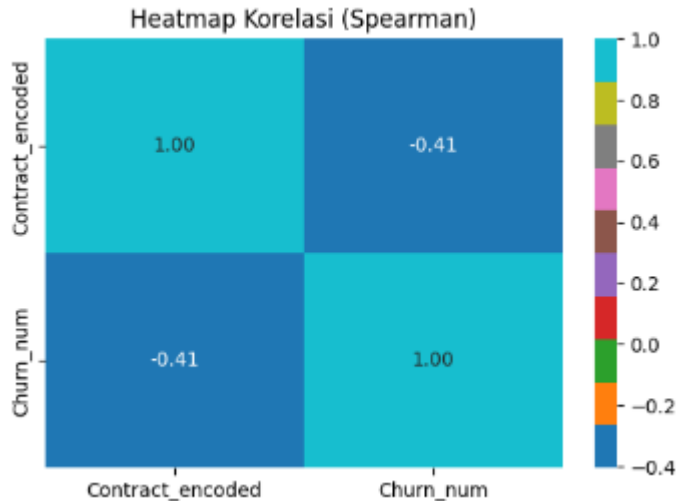
1. Analisis hubungan lama berlangganan (*tenure*) terhadap kemungkinan pelanggan melakukan churn



- Terdapat **hubungan negatif yang signifikan secara statistik** antara lama berlangganan dan status churn.
- Pelanggan yang sudah **berlangganan lebih lama cenderung lebih loyal dan tidak churn.**

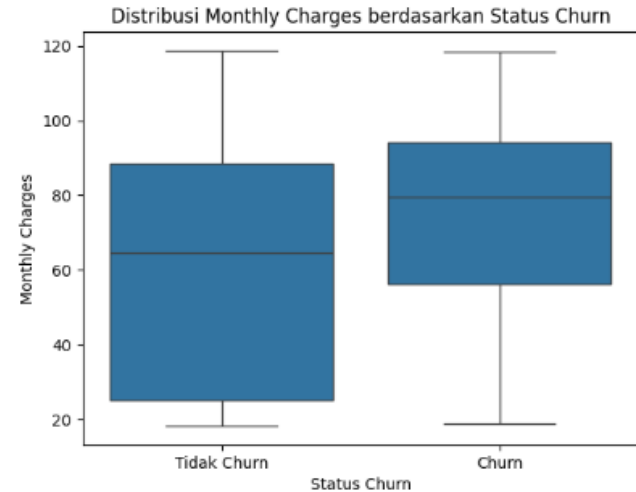
Melakukan Uji-Statistik: Fitur Tenure, Contract, and Monthly Charges terhadap Churn

2. Analisis hubungan probabilitas jenis kontrak terhadap pelanggan berhenti berlangganan (*churn*)



Nilai koefisien korelasi **-0.41** menunjukkan adanya **hubungan negatif sedang** antara jenis kontrak dan churn - artinya, **semakin panjang atau mengikat jenis kontraknya (misalnya kontrak tahunan), maka kemungkinan pelanggan untuk churn (berhenti) cenderung lebih rendah.**

3. Analisis perbedaan signifikan rata-rata tagihan bulanan (*MonthlyCharges*) antara pelanggan yang churn dan yang tidak



Pelanggan yang membayar **lebih mahal setiap bulan cenderung lebih sering berhenti berlangganan.**



04

Data Preprocessing

Data Pre Processing: Feature Engineering

Drop kolom yg
sebelumnya
dilakukan
encoding manual

Lakukan
Encoding secara
otomatis

Splitting Data

Check
Multikolinearisme

Lakukan Scalling

• Kolom seperti
'Churn_num',
'tenure_group',
'Contract_encoded'

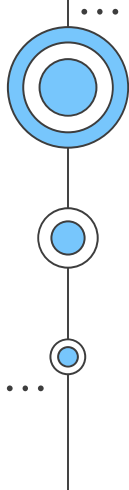
• **One-Hot Encoding:**
gender, Partner,
Dependents,
PhoneService,
PaperlessBilling,
MultipleLines,
InternetService,
OnlineSecurity,
OnlineBackup,
DeviceProtection,
TechSupport,
StreamingTV,
StreamingMovies, dan
PaymentMethod.
• **Ordinal Encoding:**
Contract

• Define target (x) dan
features (y)
• Split data 80:20
• Fit to x_train

• Pengecekan dengan VIF
• Penanganan VIF

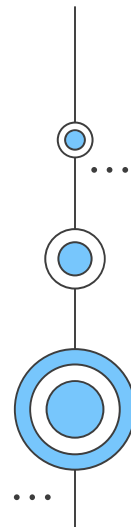
• RobustScaler → lebih
tahan terhadap distribusi
yang skewed dan nilai
ekstrem





05

Modelling & Evaluation



MODELLING & EVALUATION: Model Selection

Pada tahap pemodelan, akan diterapkan sejumlah algoritma machine learning, yaitu: Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree, Random Forest, XGBoost (Extreme Gradient Boosting), dan LightGBM (Light Gradient Boosting Machine).

Evaluasi performa model machine learning terhadap metrik recall sebelum dan sesudah penanganan multikolinearitas (melalui VIF)

	recall_before	recall_after	delta
model			
Logistic Regression	0.803127	0.784648	-0.018479
LightGBM	0.793888	0.783937	-0.009950
XGBoost	0.786780	0.776119	-0.010661
Random Forest	0.788913	0.774698	-0.014215
KNN	0.757640	0.746979	-0.010661
Decision Tree	0.735608	0.717839	-0.017768

Pemilihan Logistic Regression (sebelum VIF treatment):

1. Recall Tertinggi: 0.8031
2. Model yang Interpretatif: Koefisien regresi membantu dalam memahami pengaruh fitur.
3. Performa Stabil: Meski ada penurunan, tetap kompetitif dibanding model lain.
4. Sesuai untuk Masalah dengan Fokus pada Deteksi Positif (FN minim).

Tabel Perbandingan Performa Recall: Train vs Test

Model	Recall (Train)	Recall (Test)	Selisih (Test - Train)	Penurunan (%)
Logistic Regression	0.8031	0.5614	-0.2417	-30.9%
LightGBM	0.7938	0.5374	-0.2564	-32.3%
XGBoost	0.7868	0.5133	-0.2735	-34.7%
Random Forest	0.7889	0.5133	-0.3026	-38.3%
KNN	0.7576	0.4438	-0.3138	-41.4%
Decision Tree	0.7356	0.5080	-0.2276	-30.9%

Logistic Regression tetap menjadi **pilihan terbaik secara keseluruhan**, dengan keseimbangan antara performa di data train dan test serta **penurunan paling terkendali** dibanding model lain.

MODELLING & EVALUATION:

Penanganan Data Imbalance – Logistic Regression

1. Penerapan regularisasi L1 pada model Logistic Regression

- L1 dipilih karena tidak hanya membantu mencegah overfitting, tetapi juga secara otomatis menghapus fitur yang kurang penting. Dengan cara ini, model menjadi lebih sederhana, efisien, dan mudah dipahami—tanpa mengorbankan performa prediksi secara signifikan.

2. Variabel Penanganan:

- Oversampling with K-Fold, using SMOTE and RobustScaler
- Evaluation Metrics Without Oversampling
- Evaluation Metrics With Oversampling
- Undersampling dengan RandomUnderSampler (10-Fold CV)
- Evaluation Metrics Without Undersampling
- Evaluation Metrics With Undersampling
- Hybrid Sampling Using SMOTE + Tomek Links
- Evaluation Metrics Without SMOTE + TOMEK
- Evaluation Metrics With SMOTE + TOMEK
- Penalized Model

Perbandingan antara Oversampling, Undersampling, Hybrid, dan Penalized

Teknik	Train Recall	Test Recall	Perubahan vs Baseline
-Baseline-	0.5530	0.5498	-
Oversampling	0.7988	0.7973	▲ +0.25 (Test)
Undersampling	0.8005	0.8040	▲ +0.25 (Test)
Hybrid	0.8005	0.7980	▲ +0.25 (Test)
Penalized Model	0.7980	0.7887	▲ +0.24 (Test)

Kriteria	Oversampling	Undersampling	Hybrid	Penalized Model
Kemudahan	✅ Tinggi	⚠️ Sedang	⚠️ Kompleks	✅ Tinggi
Retain Info	⚠️ Sebagian	❌ Terbatas	✅ Optimal	✅ Optimal
Performa	0.7973	0.8040	0.7980	0.7887

Hybrid Sampling adalah pilihan optimal untuk model ini karena berhasil memadukan kekuatan dari **SMOTE** dan **undersampling**, menjaga **keseimbangan data tanpa kehilangan informasi penting**, serta menunjukkan **performa tinggi dan stabil** pada metrik recall — yang sangat krusial dalam deteksi kasus minoritas.

MODELLING & EVALUATION:

Hyperparameter Tuning with Gridsearch

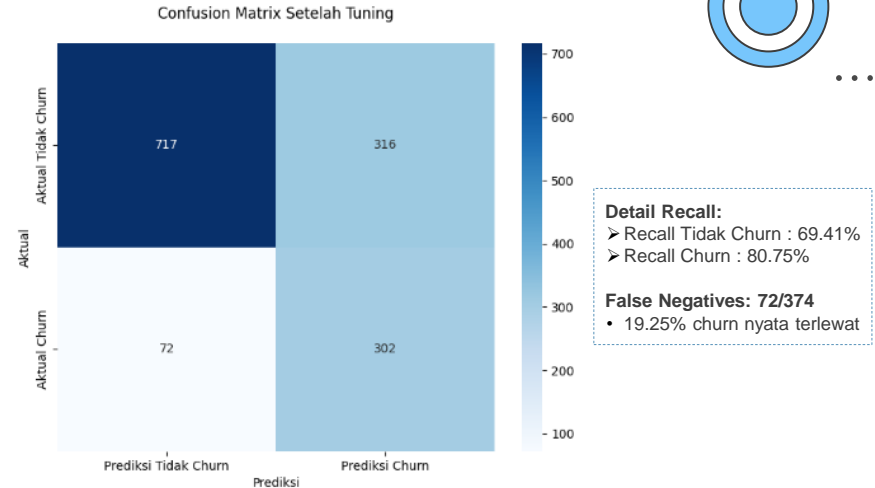
1. Hyperparameter tuning yang akan dilakukan menggunakan metode GridSearch (dengan mencoba seluruh **kombinasi hyperparameter**), sebagai berikut:

- solver = lbfgs, newton-cg, liblinear, sag, saga
- C = 1000, 500, 100, 50, 10, 5, 1, 0.5, 0.1, 0.05, 0.01, 0.005, 0.001, 0.0005, 0.0001, 0.00005
- max_iter = 100, 200, 300, 400, 500, 1000
- penalty : l1, l2, elasticnet, none

2. Best Parameters & Recall:

- solver: liblinear - Algoritma optimasi yang digunakan adalah LIBLINEAR, yang cocok untuk regularisasi L1
- C: 0.01 - Nilai regularisasi yang sangat kecil (regularisasi kuat), artinya model lebih sederhana untuk menghindari overfitting
- max_iter: 100 - Model dilatih dengan maksimum 100 iterasi
- penalty: l1 - Menggunakan regularisasi L1 (Lasso) yang bisa menghasilkan sparse model (beberapa koefisien menjadi nol)
- **Best Recall: 0.83 - Model mampu mengidentifikasi 83% dari semua kasus positif yang ada dalam data.**

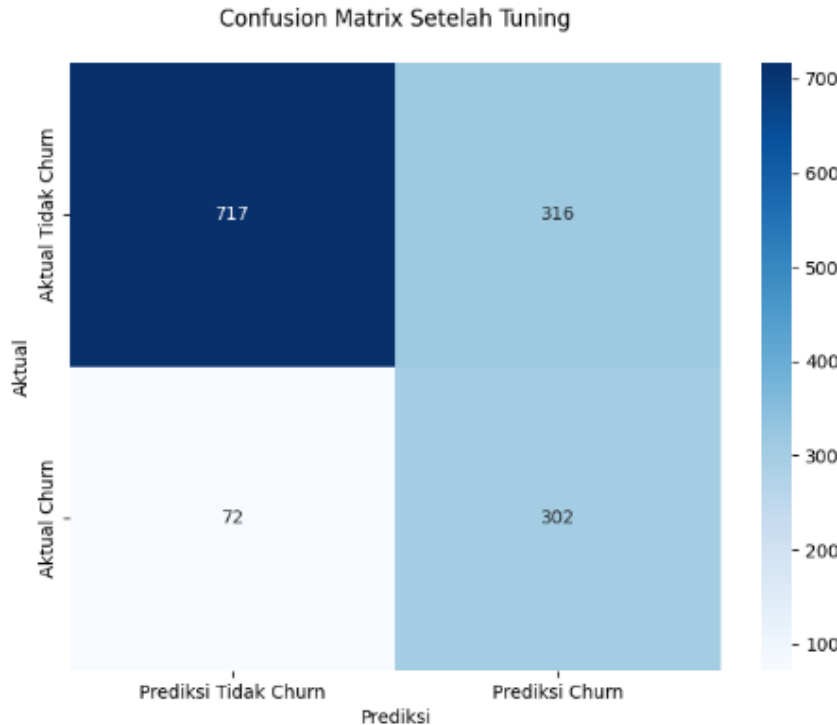
Perbandingan Performa Sebelum vs Sesudah Tuning



Berdasarkan hasil evaluasi, tuning yang dilakukan terhadap model Logistic Regression menghasilkan **sedikit peningkatan performa**, terutama pada metrik recall yang naik dari **0.77 menjadi 0.80** (kenaikan sekitar 3%).

Dengan demikian, model yang akan digunakan adalah yang **sesudah dituning**, yaitu **logistic regression yang telah dilakukan scaling dan Hybrid (SMOTETomek)**.

MODELLING & EVALUATION: Confusion Matrix in detail (After tuning)



Informasi dari confusion matrix pada test set :

- Jumlah customer yang aktualnya churn dan diprediksi akan melakukan churn (**True Positive**) : 302 orang
- Jumlah customer yang aktualnya churn tetapi diprediksi tidak akan melakukan churn (**False Negative**): 72 orang
- Jumlah customer yang aktualnya tidak churn dan diprediksi tidak akan churn (**True Negative**) : 717 orang
- Jumlah customer yang aktualnya tidak churn tetapi diprediksi akan churn (**False Positive**) : 316 orang
- Jumlah customer yang diprediksi akan churn sebanyak $302 + 316 = 618$ orang
- Jumlah customer yang diprediksi tidak akan churn sebanyak $72 + 717 = 789$ orang

Berdasarkan hasil evaluasi :

- Recall : Dari total customer yang aslinya churn, **80% yang berhasil diprediksi dengan benar.**

MODELLING & EVALUATION: Feature Coefficients

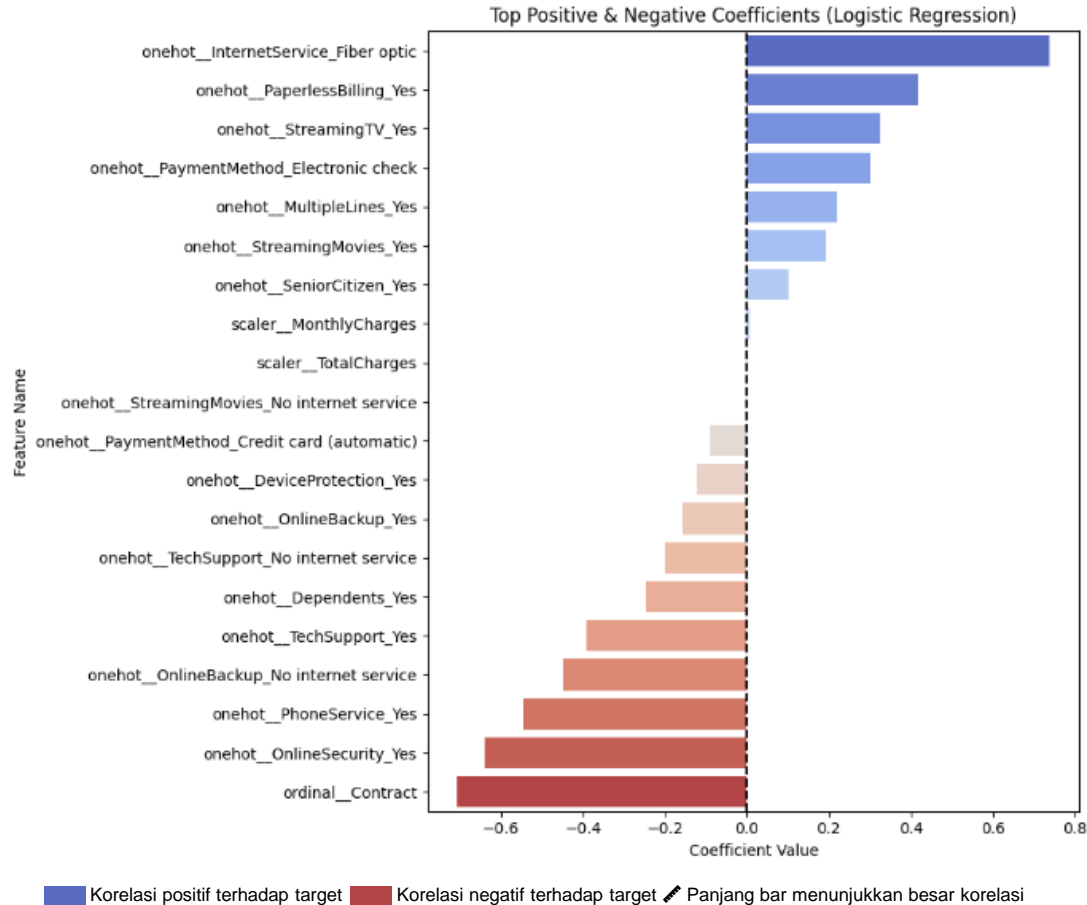
Feature	Coefficient	Odds Ratio	Interpretasi
InternetService_Fiber optic	0.738328	2.092434	Pelanggan dengan layanan Fiber Optic memiliki risiko 2 kali lebih tinggi untuk churn dibandingkan pelanggan dengan layanan DSL
PaperlessBilling_Yes	0.416661	1.516888	Pelanggan dengan tagihan digital memiliki risiko churn 1.5 kali lebih tinggi dibandingkan yang menggunakan tagihan kertas
StreamingTV_Yes	0.324487	1.383321	Pelanggan yang menggunakan layanan StreamingTV memiliki risiko churn 1.38 kali lebih besar dibandingkan pelanggan tanpa layanan StreamingTV
PaymentMethod_Electronic check	0.301732	1.352198	Pelanggan yang menggunakan pembayaran dengan elektronik cek memiliki risiko churn 1.35 kali lebih besar dibandingkan metode pembayaran Bank Transfer/Mailed Check
MultipleLines_Yes	0.220666	1.246907	Pelanggan yang memiliki layanan MultipleLines memiliki risiko churn 1.24 kali lebih besar dibandingkan pelanggan tanpa layanan MultipleLines
StreamingMovies_Yes	0.192369	1.212118	Pelanggan yang menggunakan layanan StreamingMovies memiliki risiko churn 1.21 kali lebih besar dibandingkan pelanggan tanpa layanan StreamingMovies
SeniorCitizen_Yes	0.100480	1.105702	Pelanggan lansia memiliki risiko churn 1.10 kali lebih besar dibandingkan pelanggan dengan usia muda
TotalCharges**	0.000147	1.000147	Pelanggan dengan total charge sebesar \$1000 memiliki risiko churn 1.14 kali lebih besar daripada pelanggan dengan pembayaran total senilai \$100 <disisi lain> Pelanggan dengan total charge sebesar \$100 memiliki risiko churn 0.87 kali lebih rendah dibandingkan dengan pelanggan yg membayar total tagihan sebesar \$1000
MonthlyCharges**	0.007206	1.007232	Pelanggan dengan tagihan per bulan sebesar \$118 memiliki risiko churn 2.05 kali lebih besar daripada pelanggan yang membayar biaya bulanan sebesar \$18 <disisi lain> Pelanggan yang membayar biaya bulanan sebesar \$18 memiliki risiko churn 0.48 kali lebih rendah dibandingkan dengan pelanggan yg tagihan bulanan sebesar \$118
StreamingMovies_No internet service	-0.000139	0.999861	Tidak memiliki internet tidak membuat churn lebih rendah atau lebih tinggi daripada pelanggan yang punya internet tapi tidak pakai StreamingMovies
Tenure**	-0.048834	0.952339	Pelanggan dengan lama berlangganan 72 bulan, risiko churn 0.03 kalinya pelanggan yang lama berlangganan nya baru 6 bulan <disisi lain> Pelanggan dengan lama berlangganan 6 bulan, risiko churnnya 25 kali lebih besar dari pelanggan yang lama berlangganan nya sudah 72 bulan

** Fitur Numerik perlu dilakukan analisis berbeda daripada fitur kategorik lainnya

MODELLING & EVALUATION: Feature Coefficients

Feature	Coefficient	Odds Ratio	Interpretasi
PaymentMethod_Credit card (automatic)	-0.089205	0.914658	Pelanggan yang menggunakan pembayaran dengan Credit Card memiliki kemungkinan churn 0.91 kali lebih rendah dibandingkan dengan pembayaran BankTransfer/MailedCheck
DeviceProtection_Yes	-0.121470	0.885618	Pelanggan yang berlangganan DeviceProtection memiliki kemungkinan churn 0.88 kali lebih rendah dibandingkan pelanggan tanpa layanan DeviceProtection
OnlineBackup_Yes	-0.157737	0.854074	Pelanggan yang berlangganan OnlineBackup memiliki kemungkinan churn 0.86 kali lebih rendah dibandingkan pelanggan tanpa layanan OnlineBackup
TechSupport_No internet service	-0.199377	0.819241	Pelanggan tanpa internet justru 0.81 lebih kecil risiko churn dibanding pelanggan yang punya internet tapi tidak pakai layanan dukungan teknis
Dependents_Yes	-0.245881	0.782016	Pelanggan yang memiliki tanggungan (Dependent) mempunyai kemungkinan churn 0.78 kali lebih rendah dibandingkan pelanggan yang tidak memiliki tanggungan
TechSupport_Yes	-0.391371	0.676129	Pelanggan yang berlangganan TechSupport memiliki kemungkinan churn 0.67 kali lebih rendah dibandingkan pelanggan tanpa layanan TechSupport
OnlineBackup_No internet service	-0.447965	0.638927	Pelanggan tanpa layanan internet memiliki kemungkinan churn 0.64 kali lebih rendah dibandingkan pelanggan yang punya internet tapi tidak pakai OnlineBackup. Artinya, pelanggan yang tidak punya internet justru lebih loyal daripada yang punya internet tapi tidak pakai layanan tambahan.
PhoneService_Yes	-0.545034	0.579822	Pelanggan yang berlangganan PhoneService memiliki kemungkinan churn 0.57 kali lebih rendah dibandingkan pelanggan tanpa layanan PhoneService
OnlineSecurity_Yes	-0.638105	0.528293	Pelanggan yang berlangganan OnlineSecurity memiliki kemungkinan churn 0.52 kali lebih rendah dibandingkan pelanggan tanpa layanan OnlineSecurity
Contract	-0.705923	0.493653	Semakin panjang jangka waktu kontrak (misal: kontrak 1 atau 2 tahun dibandingkan bulanan), maka kemungkinan churn 0.49 kali lebih rendah
InternetService_No	0.000000	1.000000	Tidak Berpengaruh
StreamingTV_No internet service	0.000000	1.000000	Tidak Berpengaruh

MODELLING & EVALUATION: Visualisasi Feature Coefficients





06

**Conclusion &
Recommendation**

CONCLUSION

1. Faktor Pendorong Churn (Meningkatkan Risiko)

- **Layanan Fiber Optic** → Risiko churn **2x lebih tinggi** dibanding DSL
- **Tagihan Tanpa Kertas (Paperless Billing)** → Kemungkinan churn **1.5x lebih tinggi**
- **Pembayaran via Electronic Check** → Meningkatkan risiko churn hingga **1.35x**
- **Layanan Streaming (TV & Movies)** → Meningkatkan risiko churn **1.2 – 1.38x**
- **Biaya Bulanan Tinggi (Monthly Charges)** → Pelanggan dengan biaya tinggi lebih berisiko churn
- **Total Tagihan Tinggi (Total Charges)** → Semakin besar total pembayaran, makin tinggi potensi churn (meski dampaknya kecil)
- **Pelanggan Lansia (Senior Citizen)** → Risiko churn sedikit lebih tinggi dibanding pelanggan muda

2. Faktor Pelindung Churn (Menurunkan Risiko)

- **Kontrak Jangka Panjang** → Risiko churn **50% lebih rendah** dari kontrak bulanan
- **Lama Berlangganan (Tenure)** → Pelanggan lama **jauh lebih setia** (misal: 72 bulan vs 6 bulan → risiko jauh turun)
- **Layanan Tambahan** → **OnlineSecurity, TechSupport, OnlineBackup** menurunkan risiko churn (**0.52–0.67x**)
- **Layanan Telepon (Phone Service)** → Menurunkan risiko churn
- **Tidak Memiliki Layanan Internet** → Pelanggan **tanpa internet** justru **lebih loyal** (dibanding yang punya internet tapi tak pakai fitur tambahannya)

3. Faktor Netral / Tidak Signifikan (Fitur-fitur ini tidak berpengaruh secara signifikan terhadap churn):

MultipleLines, DeviceProtection, Dependents, StreamingMovies_No internet service, TechSupport_No internet service, OnlineBackup_No internet service, PaymentMethod_Credit card (automatic), InternetService_No, StreamingTV_No internet service



RECOMMENDATION: Berdasarkan peran Stakeholders



Manajemen Eksekutif

Fokus: *Pengambilan keputusan strategis dan alokasi sumber daya*

Isu & Data	Rekomendasi Strategi
Fiber Optic (OR = 2.09): risiko churn tinggi	- Audit kualitas jaringan dan kepuasan pelanggan fiber optic per kuartal (target NPS ≥ 40) - Survei churner untuk identifikasi 3 alasan utama berpindah
Metode pembayaran Electronic Check (OR = 1.35)	- Insentif migrasi 5% ke metode autodebit/kartu kredit (target: konversi 30% dalam 3 bulan) - Reminder jatuh tempo otomatis via WhatsApp/SMS
Kontrak jangka panjang (OR = 0.49): proteksi churn kuat	- Diskon 10–15% untuk upgrade dari kontrak bulanan ke tahunan (target uptake 25%) - Bonus upgrade kecepatan 20 Mbps bagi yang perpanjang lebih awal
Tenure pelanggan: <6 bulan = 25x lebih rentan churn dibanding 72 bulan	- Program onboarding 90 hari dengan edukasi layanan (target churn <15% dalam 3 bulan) - Loyalty cashback Rp50.000 setelah bertahan 12 bulan
Tagihan bulanan tinggi lebih rentan churn	- Review ulang struktur harga untuk pelanggan high-bill, uji coba diskon adaptif berbasis segmentasi ARPU (target churn turun 10%)



Tim Pemasaran

Fokus: *Eksekusi program retensi & kampanye pelanggan*

Isu & Data	Rekomendasi Strategi
Paperless Billing (OR = 1.52): risiko churn tinggi pelanggan digital-savvy	- Kampanye edukasi manfaat + loyalty tier (target kepuasan >80%) - Jalur cepat penanganan keluhan pelanggan paperless melalui WhatsApp Business
OnlineSecurity (OR = 0.53) & TechSupport (OR = 0.67): pelindung churn kuat	- Bundling add-on pelindung churn: OnlineSecurity + TechSupport diskon 30% (target aktivasi add-on +20%)
Pelanggan tidak punya layanan tambahan (add-on)	- Segmentasi pelanggan "low engagement" dan kirim penawaran add-on protektif (target konversi bundling 25% dalam 6 bulan)
SeniorCitizen & Dependents: churn relatif rendah	- Buat loyalty program khusus segmen keluarga & lansia, seperti "Family Plan" dengan harga spesial atau bonus kuota
StreamingTV & StreamingMovies (OR >1.2): pendorong churn ringan	- Tawarkan diskon add-on lainnya jika pelanggan sudah menggunakan streaming (contoh: TechSupport), untuk menyeimbangkan risiko churn

RECOMMENDATION: Pengembangan Model dan Limitasi Model



Rekomendasi Pengembangan Model

- Mencoba algoritma dan juga mencoba hyperparameter tuning kembali serta oversampling lain (SMOTENC, ADASYN, dll).
- Tambahkan fitur perilaku baru: pola bayar, pemakaian, interaksi CS.
- Gunakan metrik lain seperti F1/F2 Score untuk klasifikasi tidak seimbang.



Limitasi Model

- TotalCharges kurang akurat (kemungkinan ada biaya tambahan yg tidak tersedia).
- Data perilaku pelanggan terbatas → insight berkurang.
- Model linear (LogReg) tidak tangkap hubungan non-linear dengan baik.
- Data tidak seimbang tetap bisa memengaruhi hasil meski sudah ditangani.



07

**Business
Implementation**

BUSINESS IMPLEMENTATION

... Analisis Dampak Finansial: Model Prediksi Churn vs Tanpa Intervensi

ASUMSI DASAR

Metric	Nilai
Total pelanggan (data test)	1.407
Pendapatan per pelanggan/bulan	\$64.76
Biaya akuisisi pelanggan baru	\$140
Biaya retensi per pelanggan	\$7

Skenario 1: Tanpa Model Prediksi

Churn Rate: 27% (≈380 pelanggan)



Konsekuensi	Perhitungan	Total
Kehilangan pendapatan tahunan	$380 \times \$64.76 \times 12$	\$295,305
Biaya akuisisi pengganti	$380 \times \$140$	\$53,200
Total kerugian tahunan		\$348,505

Skenario 2: Dengan Model Prediksi

- FN: 72 pelanggan churn (tdk terdeteksi)
- FP: 316 pelanggan loyal (terdeteksi salah)



Biaya	Perhitungan	Total
Kehilangan pendapatan (FN)	$72 \times \$64.76 \times 12$	\$55,952
Biaya akuisisi (FN)	$72 \times \$140$	\$10,080
Biaya retensi (TP + FP)	$(302 + 316) \times \$7$	\$4,326
Total biaya dengan model		\$70,358

Dengan menggunakan model prediksi churn, perusahaan dapat:

1. Efisiensi Finansial

- **Penghematan biaya hingga \$278.147 per tahun (80%)**
- Churn berkurang drastis: dari 380 jadi hanya 72 pelanggan
- Retensi lebih tepat sasaran: hanya 618 pelanggan diberi promosi

2. Biaya Salah Prediksi

- Ada **316 pelanggan loyal yang salah sasaran** → pemborosan \$2.212
- Namun, ini jauh lebih kecil dibanding kerugian tanpa model

...

“Berdasarkan hasil analisa, modelling dan evaluation serta penerapan business implementation, dapat disimpulkan bahwa **model terbukti strategis dan efektif, membantu menurunkan biaya akuisisi & meningkatkan kualitas keputusan berbasis data.**”

...





08

References



REFERENCES

1. Wang, Y., & Ni, X. S. (2019).
 - [Predicting Class-Imbalanced Business Risk Using Resampling, Regularization, and Model Ensembling Algorithms.](#)
2. SAS Institute Inc. (2002).
 - [A Case Study in Data Mining to Improve Customer Retention: The Application of Predictive Modeling and Profiling.](#)
3. ResearchGate. (2020).
 - [Analysis of Churn in Mobile Telecommunications: Predicting the Timing of Customer Churn.](#)
4. Saleh, S., & Saha, S. (2023).
 - [Customer Retention and Churn Prediction in the Telecommunication Industry: A Case Study on a Danish University.](#)
5. Mohanty, N., Behera, B. K., Ferrie, C., & Dash, P. (2024).
 - [A Quantum Approach to Synthetic Minority Oversampling Technique \(SMOTE\).](#)
6. Imani, M. (2024).
 - [Customer Churn Prediction in Telecommunication Industry: A Literature Review.](#)
7. Zdziebko, T., Sulikowski, P., Sałabun, W., Przybyła-Kasperek, M., & Bąk, I. (2024).
 - [Optimizing Customer Retention in the Telecom Industry: A Fuzzy-Based Churn Modeling with Usage Data](#)

Thanks!

Notebook Python Code: [Here](#)
All final project include dataset: [Here](#)



[rofirosdiani](#)



[rofirosdiani](#)



rosdianirofi@gmail.com

CREDITS: This presentation template was created by [Slidesgo](#),
including icons by [Flaticon](#), infographics & images by [Freepik](#)
and illustrations by [Stories](#)

Please keep this slide for attribution