



# Fighting **Churn** with Machine Learning

**From Data to Decisions**

Where the data come from?

	Customer ID	Full Name	Date of Purchase	Age	State	MTN Device	Gender	Satisfaction Rate	Customer Review	Customer Tenure in months	Subscription Plan	Unit Price	Number of Times Purchased	Total Revenue	Data Usage	Customer Churn Status	Reasons for Churn
0	CUST0001	Ngozi Berry	Jan-25	27	Kwara	4G Router	Male	2	Fair	2	165GB Monthly Plan	35000	19	665000	44.48	Yes	Relocation
1	CUST0002	Zainab Baker	Mar-25	16	Abuja (FCT)	Mobile SIM Card	Female	2	Fair	22	12.5GB Monthly Plan	5500	12	66000	19.79	Yes	Better Offers from Competitors
2	CUST0003	Saidu Evans	Mar-25	21	Sokoto	5G Broadband Router	Male	1	Poor	60	150GB FUP Monthly Unlimited	20000	8	160000	9.64	No	NaN
3	CUST0003	Saidu Evans	Mar-25	21	Sokoto	Mobile SIM Card	Male	1	Poor	60	1GB+1.5mins Daily Plan	500	8	4000	197.05	No	NaN
4	CUST0003	Saidu Evans	Mar-25	21	Sokoto	Broadband MiFi	Male	1	Poor	60	30GB Monthly Broadband Plan	9000	15	135000	76.34	No	NaN

Saya mendapatkan data tersebut melalui Kaggle ([Click in here if you want to check the details](#)).

Data tersebut terdapat:

17 Kolom dan 974 Data

Penjelasan beberapa kolom:

- MTN Device – Jenis Perangkat
  - Satisfaction Rate – Tingkat Kepuasan
  - Customer Review – Ulasan Pelanggan
  - Customer Tenure – Lama waktu pelanggan menggunakan layanan
- State – Negara Bagian
  - Subscription Plan – Jenis paket langganan
  - Data Usage – Jumlah data internet yang digunakan pelanggan
  - Customer Churn Status – Status pelanggan masih aktif atau sudah berhenti
  - Reasons for Churn – Alasan mengapa pelanggan berhenti

## Problem Understanding

We need to understand what the problem that we want to solve. Disini karena memang terdapat data terkait **Churn Rate**, maka kita akan fokus kepada **Churn Rate**.

Setelah dihitung, ternyata sekitar **29,13%** dari total pelanggan udah berhenti langganan alias churn. Sebagai Informasi tambahan, **benchmark Industri Telekomunikasi di United States sebesar 21%**. Ini menunjukkan bahwa pada kuartal pertama, **MTN memiliki churn rate yang tergolong besar**.

**MRR**

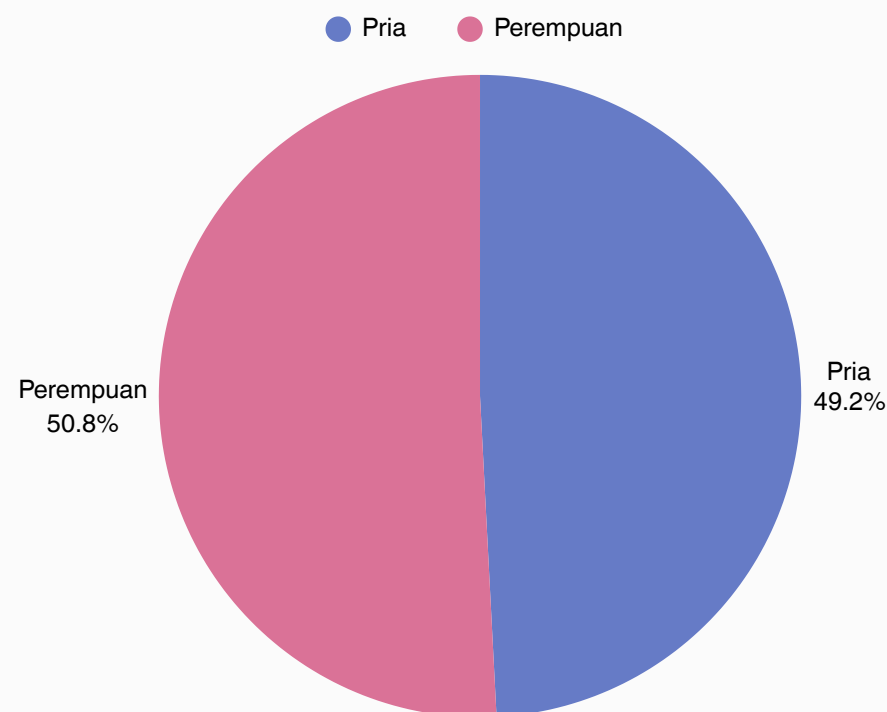
**29.09%**

**CSAT**

**38.81%**

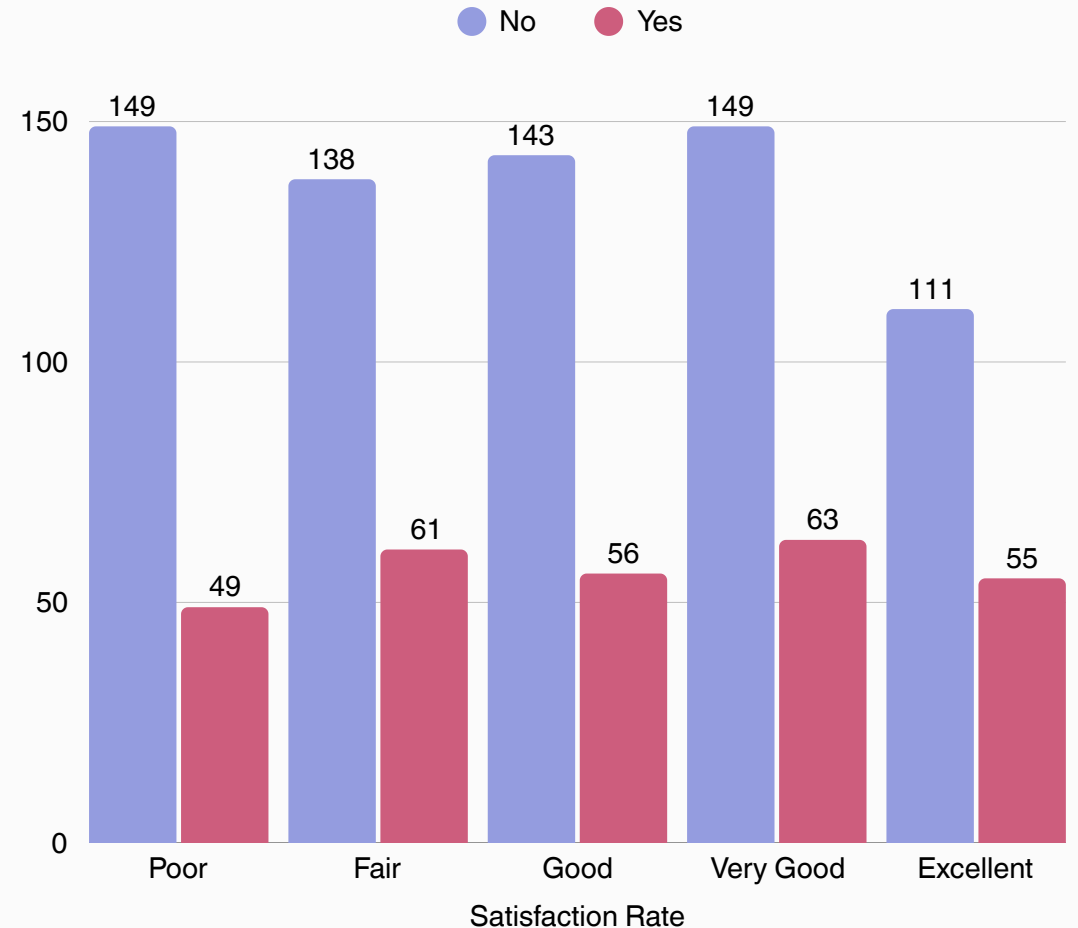
Ditambah, **MRR** dan **CSAT** menunjukkan angka yang rendah. Ini menjadi sinyal yang kuat bahwa kita perlu serius memikirkan strategi untuk menurunkan **Churn Rate**.

## Data Insight

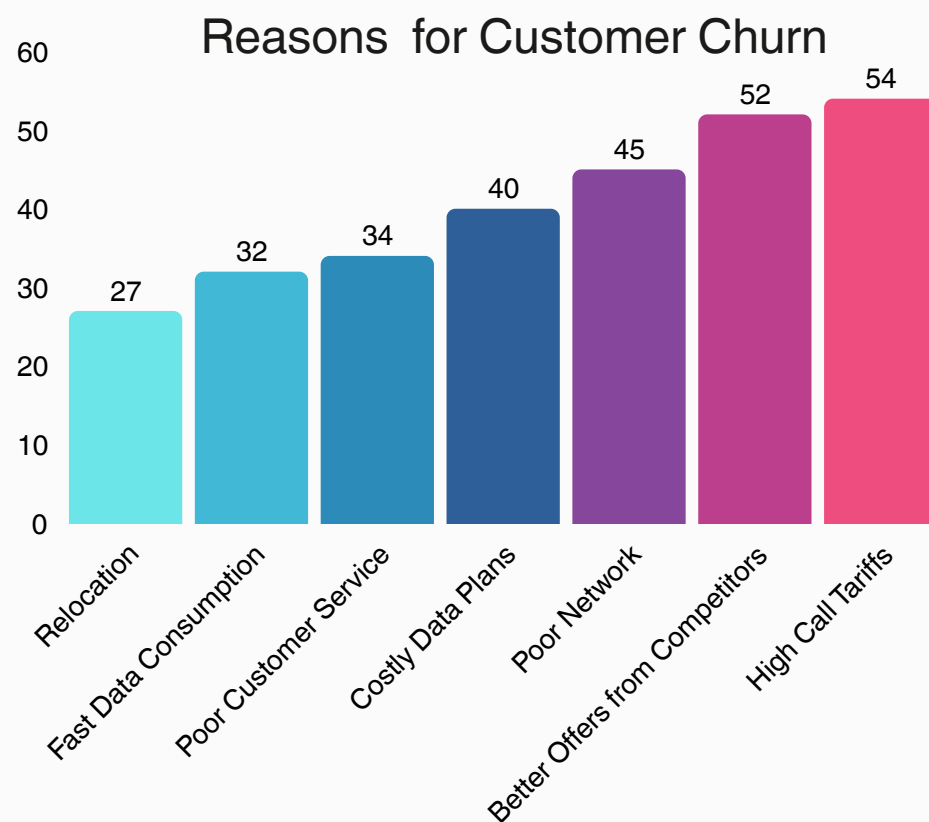


Tidak ada perbedaan signifikan pada Gender, ini bisa menjadi bukti bahwa **Gender tidak berpengaruh besar pada Churn Rate.**

Churn terjadi di semua tingkat kepuasan, bahkan churn terjadi dari pelanggan "**Very Good**", hal ini menunjukkan bahwa kepuasan saja tak cukup. **Faktor seperti harga, promo, dan branding juga berpengaruh.**

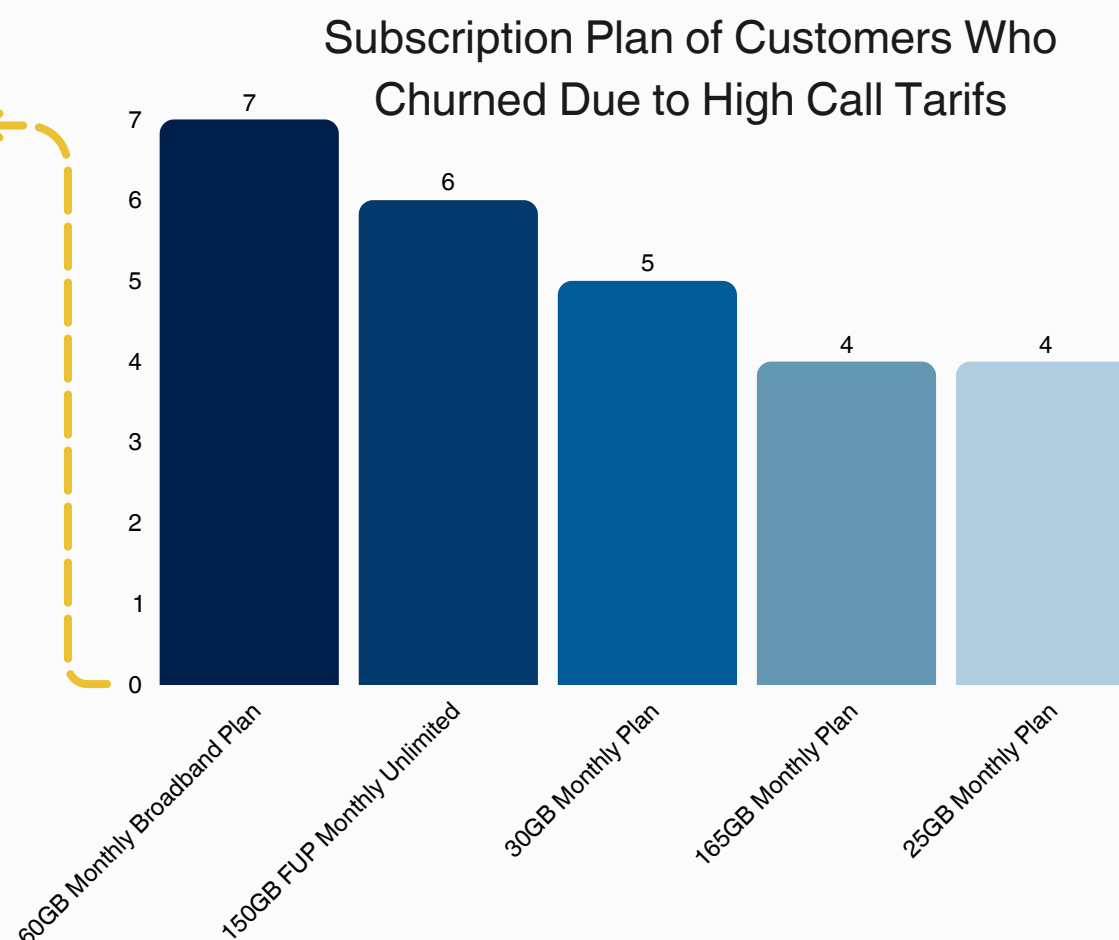


## Data Insight

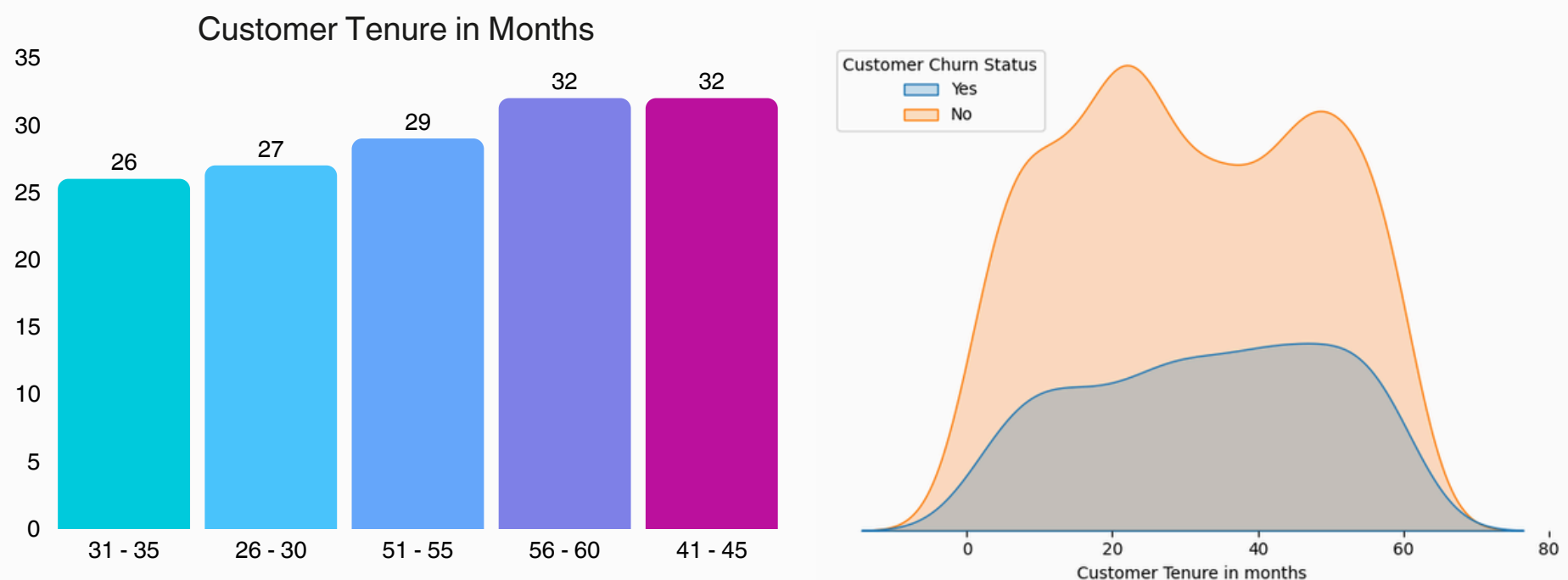


Visualisasi menunjukkan tiga alasan utama meningkatnya churn rate: **tarif mahal, penawaran kompetitor yang lebih menarik, dan masalah jaringan.** Ketiganya perlu jadi fokus perusahaan untuk menekan churn rate.

Terlihat bahwa paket **60GB Monthly Broadband** dan **150GB FUP Monthly Unlimited** jadi yang paling banyak ditinggalkan karena **tarif nelson yang mahal.** Ini menunjukkan bahwa pelanggan masih merasa tarif komunikasi tidak sepadan. Artinya, **perlu evaluasi khusus pada paket-paket ini.** Bisa dengan menambahkan bonus nelson atau mengkaji ulang harga agar lebih sesuai ekspektasi pelanggan.



## Data Insight



Ada beberapa insight yang kita dapatkan:

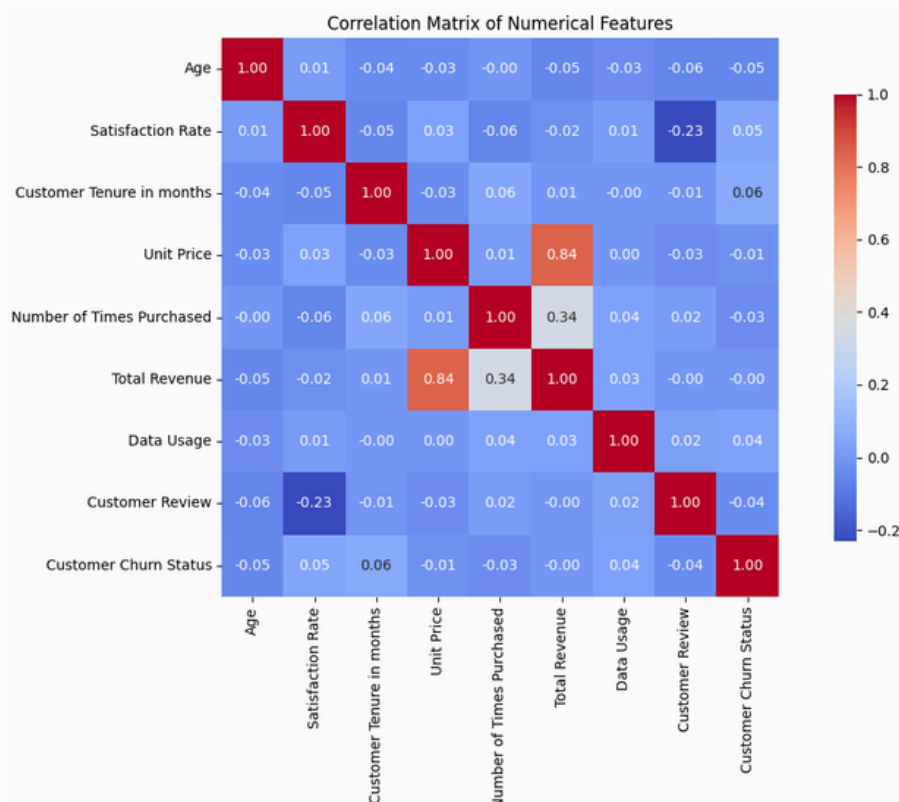
1. Pelanggan yang retention memiliki masa langganan lebih panjang dibandingkan **churn**.
2. Distribusi **Churn** cenderung meningkat ketika sudah mulai masuk ke rentang 30 – 60.
3. **Churn** paling banyak terjadi pada pelanggan dengan tenure 41 – 60.

Ini menunjukkan bahwa pelanggan lama mulai bosan atau tergoda penawaran kompetitor, menandakan program loyalitas belum efektif. Perusahaan perlu fokus pada R&D untuk inovasi produk dan memperkuat loyalitas agar pelanggan tetap setia.

# Machine Learning

Kita memiliki data dan tujuan kita ingin menekan **churn rate**. kita bisa memanfaatkan data tersebut untuk membuat **model prediksi** agar dapat memberikan **penawaran ataupun promo** kepada customer yang lebih **tepat dan personalisasi**.

Tapi sebelum itu kita perlu terlebih dahulu memilih **feature yang relevan**. Kita bisa coba cek dari korelasi terlebih dahulu, jika tidak memiliki korelasi yang signifikan maka bisa kita hapus dan **fokus kepada feature yang memang memiliki pengaruh terhadap keputusan**.



Dari korelasi terlihat bahwa **Unit Price** dan **Total Revenue** tidak berkorelasi dengan **churn**, sehingga bisa dihapus agar model dapat fokus pada fitur yang lebih relevan.



## Feature Engineering

Tetapi harus diingat selalu **kita perlu hapus beberapa kolom yang bisa memberikan potensi terjadinya Data Leakage**. Kita akan hapus kolom **Satisfaction Score** dan **Customer Review** karena berpotensi bisa membuat **Data Leakage**.

As an additional information, kita juga membuat dua fitur baru yaitu **Estimated\_spend** (Perkiraan Pengeluaran per data usage) dan **Spend\_per\_month** (Pengeluaran Perbulan).

You want to check details algorithm or code? [You can click here :D](#)

Hasil dari beberapa pelatihan model:

### Decision Tree

Accuracy: 69%  
Recall: 45%  
Precision: 44%  
F1 Score: 44%

### Random Forest

Accuracy: 73%  
Recall: 25%  
Precision: 52%  
F1 Score: 33%

### Gradient Boosting

Accuracy: 73%  
Recall: 23%  
Precision: 50%  
F1 Score: 31%



## Choose the Model

Kita akan memilih **Decision Tree** sebagai model dasar karena tujuan utama kita memang untuk melakukan identifikasi pengguna yang berpotensi **churn**. Meskipun akurasi tidak setinggi model lain, **Decision Tree** memiliki nilai **recall** yang tinggi yang berarti lebih mampu menangkap sebanyak mungkin pengguna yang benar-benar akan **churn**.

Kesalahan dalam mengklasifikasikan pengguna aktif sebagai **churner** dapat ditoleransi, karena tidak berdampak langsung pada kerugian, sementara manfaat utamanya adalah potensi untuk mengurangi **churn** secara signifikan.



## Conclusions

Tujuan utama dari proyek ini adalah mengidentifikasi pelanggan yang berpotensi **churn** sedini mungkin, agar perusahaan dapat melakukan tindakan pencegahan yang lebih proaktif dan terarah.

Model **Decision Tree** dipilih sebagai model dasar karena:

- **Memiliki recall tinggi**, yang memungkinkan model menangkap lebih banyak pelanggan yang benar-benar akan **churn**.
- Meski akurasi bukan yang tertinggi, model ini tetap efektif karena **false negative lebih berisiko dibanding false positive**, yang dampaknya relatif kecil.

Strategi ini kemungkinan akan **meningkatkan biaya pemasaran**, karena beberapa pelanggan aktif mungkin ikut ditargetkan. Namun, hal ini masih dapat diterima karena **tujuan utamanya adalah untuk mengurangi churn rate secara signifikan**.

Dengan pendekatan ini, **target kita adalah menurunkan churn rate dari 29% menjadi di bawah benchmark industri sebesar 21%**, hal tersebut dapat dilakukan melalui tindakan pencegahan berbasis model prediktif.

## Additional Information



### Tentang Saya

Saya Abdul Aziz, mahasiswa Informatika di Universitas Sultan Ageng Tirtayasa dengan minat kuat di bidang Data Science. Saat ini saya sedang mencari kesempatan magang atau Studi Independen untuk mengembangkan keterampilan analisis data, machine learning, dan AI, serta memperoleh pengalaman nyata di dunia kerja.

## Resource:



| [Click here to see the detailed code](#)



| [Click here to see the dataset](#)

## Reference:

"Customer Retention Rates: 2023 Benchmarks by Industry." Exploding Topics, 2023, <https://explodingtopics.com/blog/customer-retention-rates>. Diakses 5 Mei 2025.