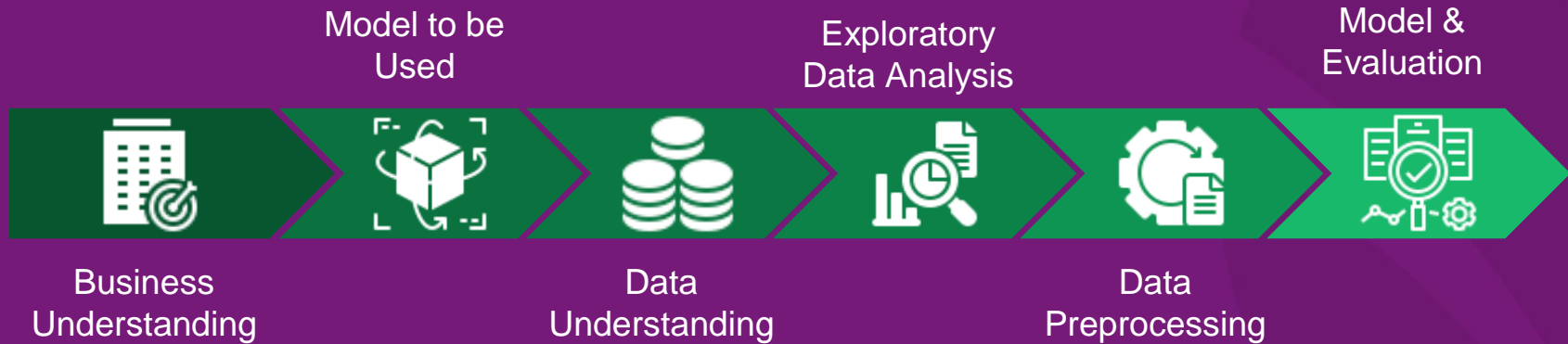


Predicted Customer Churn

Achmad Fathony

Data Science

Outline



BUSINESS UNDERSTANDING

Perkembangan industri telekomunikasi yang sangat cepat menyebabkan banyaknya perusahaan telekomunikasi dan meningkatnya persaingan antar provider



Oleh karena itu perusahaan diharuskan mampu untuk mempertahankan pelanggannya



Dimana perusahaan harus dapat mengetahui faktor-faktor yang menyebabkan pelanggan akan churn atau tidak, agar perusahaan dapat memetakan strategi bisnis untuk mempertahankan pelanggan

Pelanggan memiliki hak dalam memilih provider yang sesuai dan dapat beralih dari provider sebelumnya yang diartikan sebagai **Customer Churn**. Peralihan ini dapat menyebabkan berkurangnya pendapatan bagi perusahaan telekomunikasi sehingga penting untuk ditangani.

Goal

- Melakukan prediksi customer churn agar perusahaan dapat memetakan strategi bisnis untuk melakukan prediksi customer churn

Objective

- Mengetahui insight yang ada pada parameter customer churn
- Membuat model machine learning yang dapat memprediksi pelanggan yang kemungkinan besar akan churn

Metric

- Churn Rate

Model to be Used

Logistic Regression:

- Melakukan prediksi customer churn agar perusahaan dapat memetakan strategi bisnis untuk melakukan prediksi customer churn

Random Forest:

- Mengetahui insight yang ada pada parameter customer churn
- Membuat model machine learning yang dapat memprediksi pelanggan yang kemungkinan besar akan churn

K-Nearest Neighbors (KNN):

- Melakukan prediksi customer churn agar perusahaan dapat memetakan strategi bisnis untuk melakukan prediksi customer churn

XGBOOST:

- Mengetahui insight yang ada pada parameter customer churn
- Membuat model machine learning yang dapat memprediksi pelanggan yang kemungkinan besar akan churn

SVC (Support Vector Machine):

- Mengetahui insight yang ada pada parameter customer churn
- Membuat model machine learning yang dapat memprediksi pelanggan yang kemungkinan besar akan churn

Data Understanding

Nama Fitur	Deskripsi
state	Status pelanggan (aktif/tidak aktif)
account_length	Lama kepemilikan akun (dalam bulan)
area_code	Kode area nomor telepon pelanggan
international_plan	Apakah pelanggan memiliki paket telepon internasional (ya/tidak)
voice_mail_plan	Apakah pelanggan memiliki paket voicemail (ya/tidak)
number_vmail_messages	Jumlah pesan voicemail yang diterima
total_day_minutes	Jumlah menit panggilan yang dilakukan pada siang hari
total_day_calls	Jumlah panggilan yang dilakukan pada siang hari
total_day_charge	Biaya total panggilan siang hari
total_eve_minutes	Jumlah menit panggilan yang dilakukan pada malam hari
total_eve_calls	Jumlah panggilan yang dilakukan pada malam hari
total_eve_charge	Biaya total panggilan malam hari
total_night_minutes	Jumlah menit panggilan yang dilakukan pada dini hari
total_night_calls	Jumlah panggilan yang dilakukan pada dini hari
total_night_charge	Biaya total panggilan dini hari
total_intl_minutes	Jumlah menit panggilan internasional
total_intl_calls	Jumlah panggilan internasional
total_intl_charge	Biaya total panggilan internasional
number_customer_service_calls	Jumlah panggilan ke layanan pelanggan
churn	Apakah pelanggan berhenti berlangganan (ya/tidak)

Dataset dibagi menjadi data train dan data test untuk dilakukan modeling

Informasi Data Train:

- Data Train terdiri dari 4250 baris dan 20 kolom
- Data Train terdiri dari 15 fitur numerikal dan 5 fitur kategorikal
- Data Train tidak memiliki nilai null atau missing value
- Data Train tidak memiliki baris data duplikat

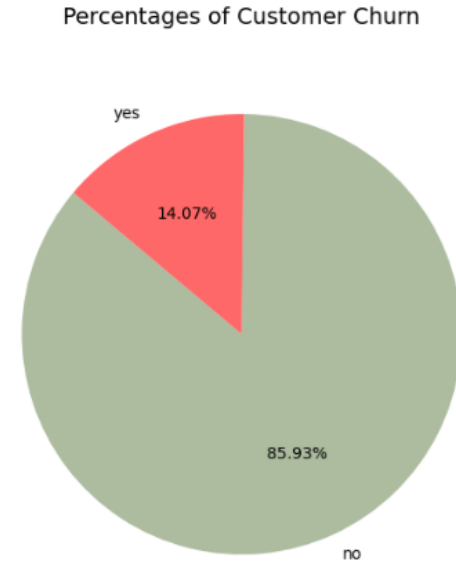
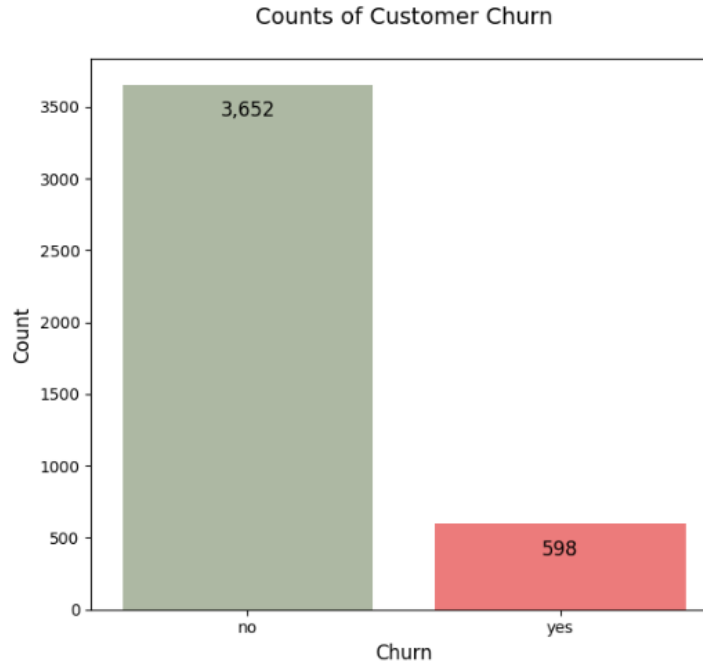
	columns	datatype	null_values	unique_values
0	id	int64	0	750
1	state	object	0	51
2	account_length	int64	0	175
3	area_code	object	0	3
4	international_plan	object	0	2
5	voice_mail_plan	object	0	2
6	number_vmail_messages	int64	0	39
7	total_day_minutes	float64	0	619
8	total_day_calls	int64	0	100
9	total_day_charge	float64	0	619
10	total_eve_minutes	float64	0	611
11	total_eve_calls	int64	0	102
12	total_eve_charge	float64	0	584
13	total_night_minutes	float64	0	628
14	total_night_calls	int64	0	97
15	total_night_charge	float64	0	502
16	total_intl_minutes	float64	0	135
17	total_intl_calls	int64	0	17
18	total_intl_charge	float64	0	135
19	number_customer_service_calls	int64	0	7

Informasi Data Test:

- Data Test terdiri dari 750 baris dan 20 kolom
- Terdapat perbedaan antara data train dan data test yaitu pada column id
- Data Test terdiri dari 16 fitur numerikal dan 4 fitur kategorikal
- Data Test tidak memiliki nilai null atau missing value
- Data Test tidak memiliki baris data duplikat

Exploratory Data analysis

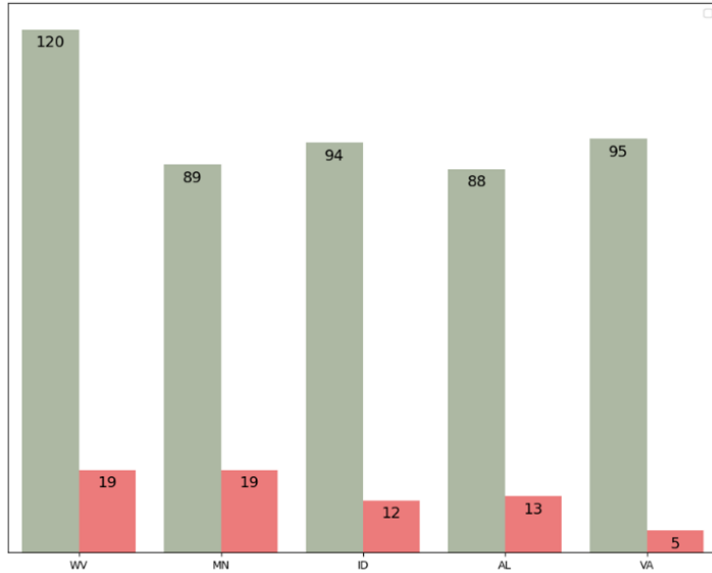
Total Churn



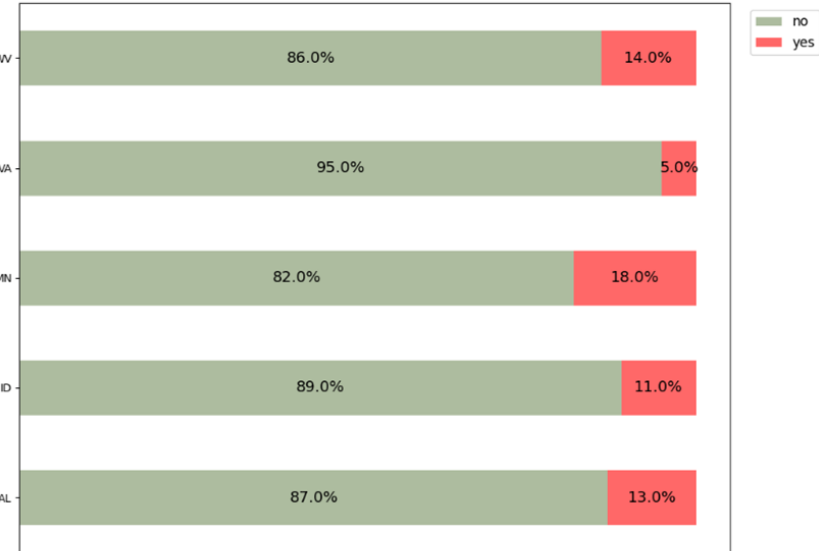
Pada grafik diatas terlihat bahwa pelanggan yang berhenti berlangganan sebanyak 598 orang dari 4250 orang atau sekitar 14.7%. Ini menunjukkan bahwa 1 dari 7 pelanggan berhenti berlangganan.

Churn by State

Top 5 Churn by State

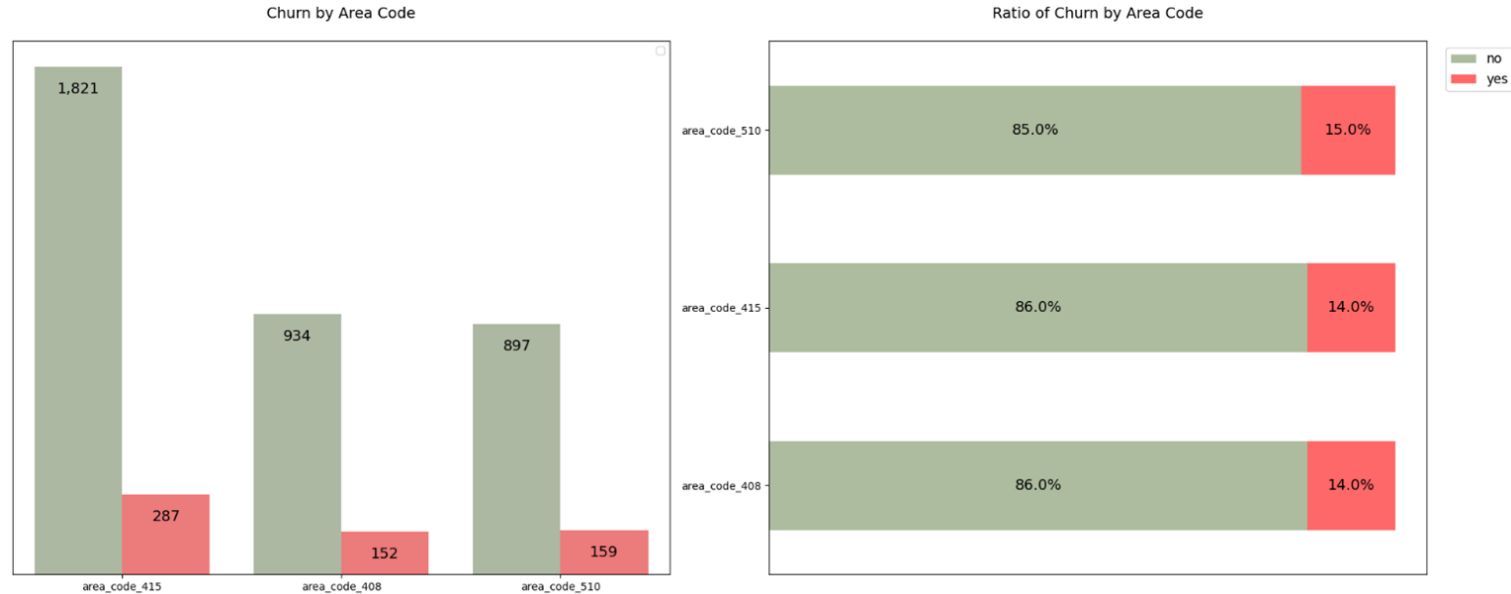


Top 5 Ratio of Churn by State



Grafik menunjukkan bahwa lima negara bagian AS: WV, MN, ID, AL, dan VA memiliki pelanggan terbanyak dengan tingkat churn minimal 5%.

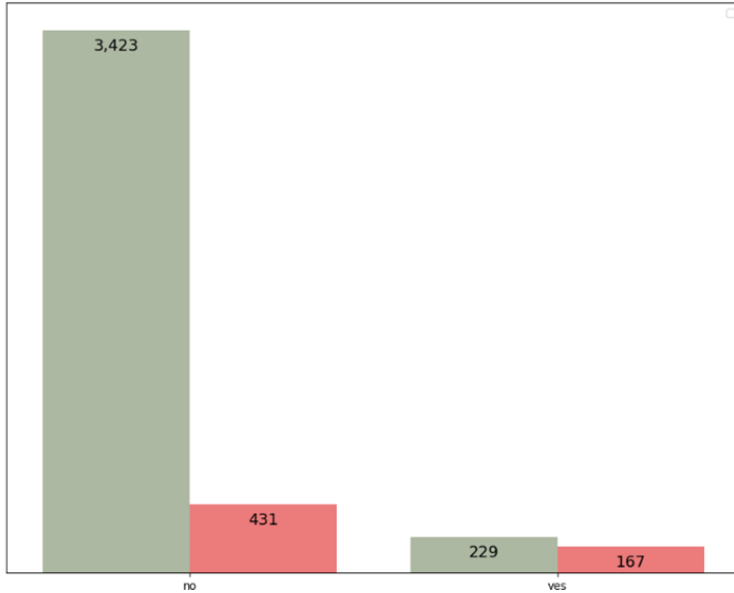
Churn by Area code



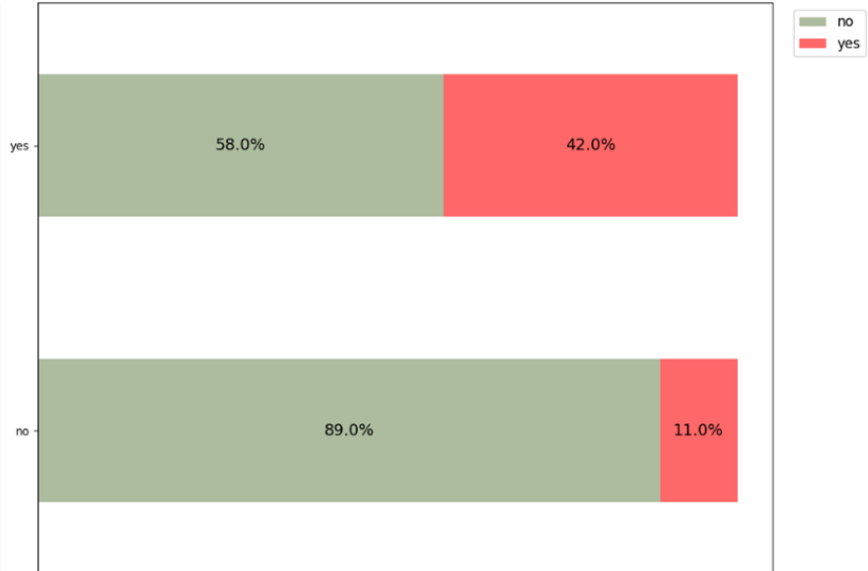
Grafik menunjukkan bahwa Area 415 memiliki jumlah pelanggan paling tinggi dan secara proporsi, persentase churn di setiap area relatif sama, yaitu sekitar 14-15%.

Churn by International Plan

Churn by International Plan



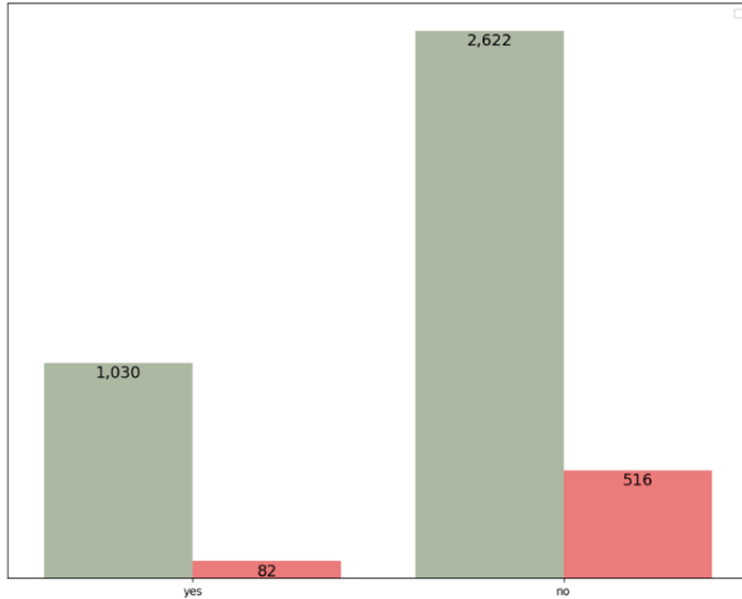
Ratio of Churn by International Plan



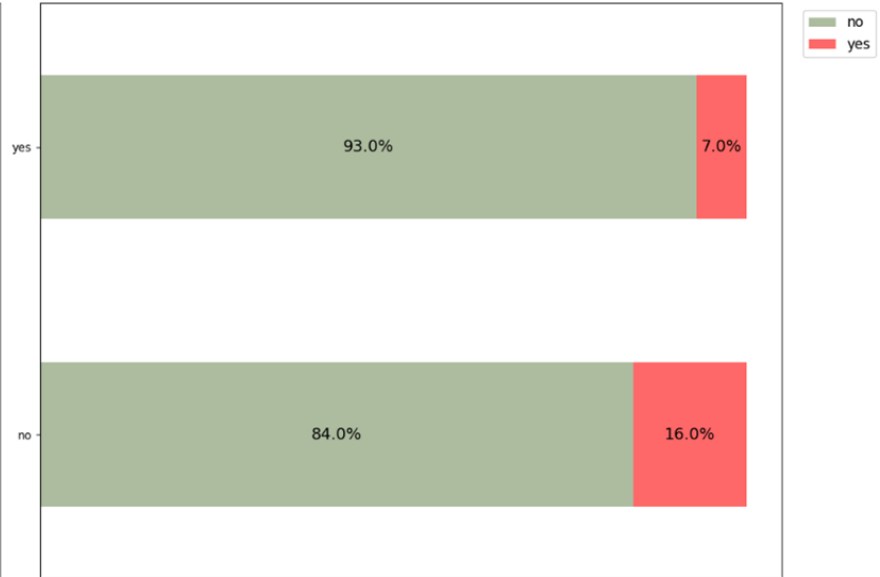
Grafik menunjukkan bahwa meskipun jumlahnya rendah, pelanggan dengan International Plan memiliki tingkat churn yang tinggi, kemungkinan karena biaya roaming yang mahal atau masalah pada kualitas jaringan.

Churn by Voicemail Plan

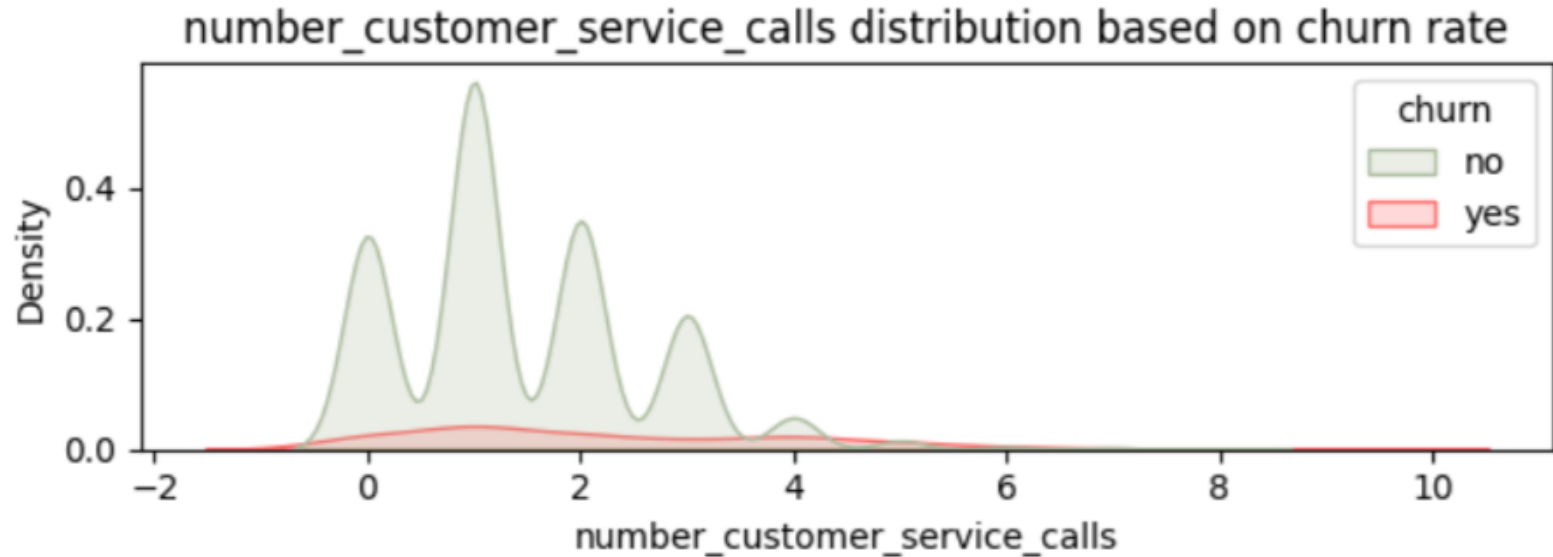
Churn by Voice Mail Plan



Ratio of Churn by Voice Mail Plan

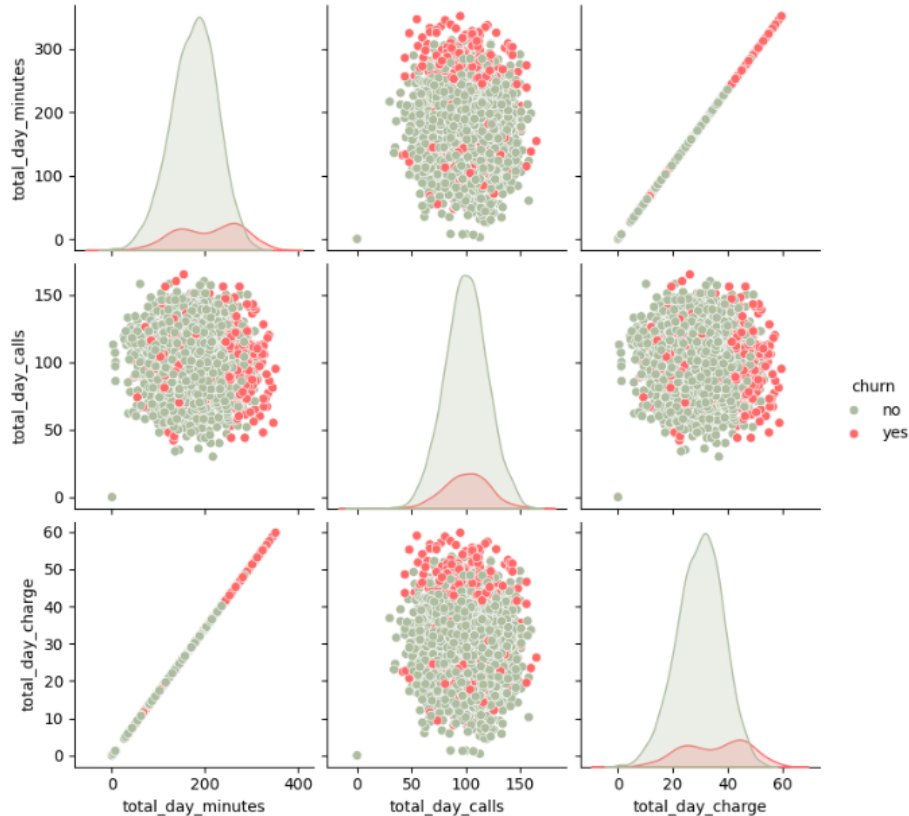


Grafik menunjukkan bahwa pelanggan tanpa voice mail plan lebih berisiko churn karena tidak dapat menerima pesan suara, yang dapat menyebabkan ketidaknyamanan.



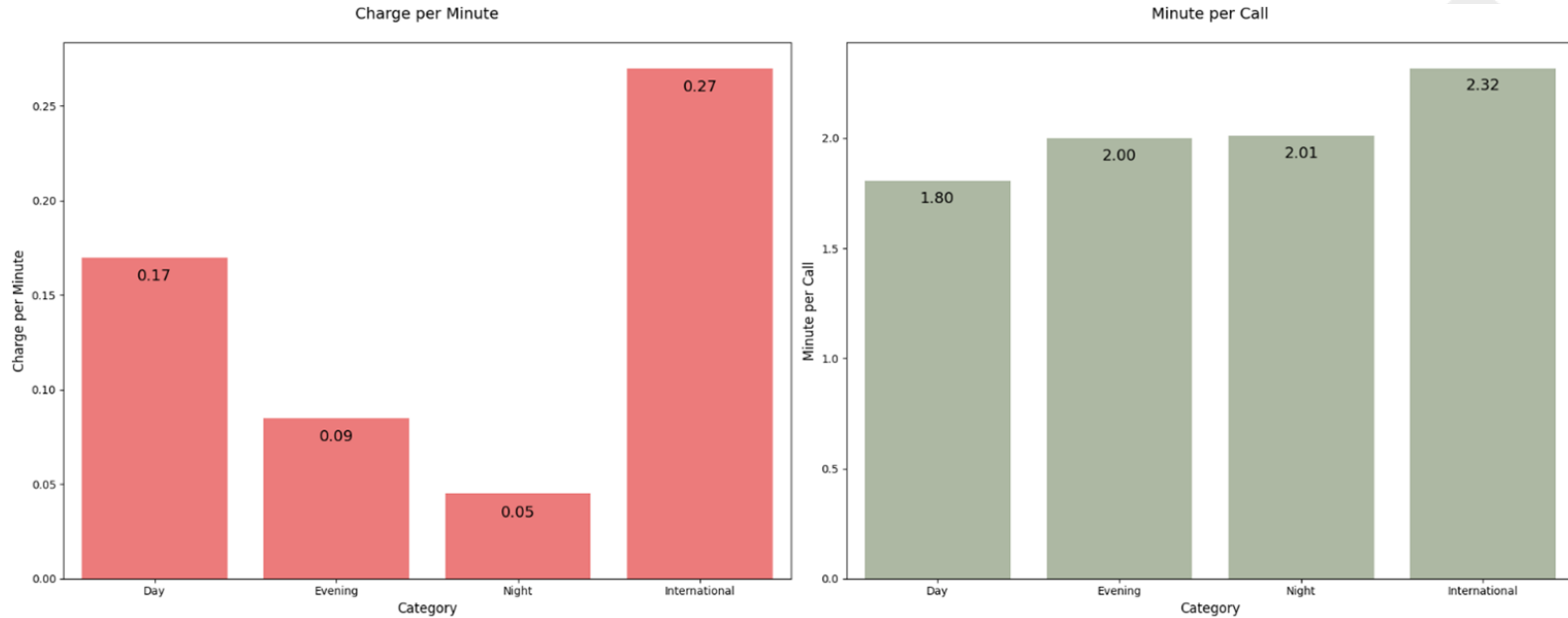
Grafik menunjukkan bahwa pelanggan yang sering menghubungi customer service (lebih dari 4 kali) memiliki potensi churn tinggi, kemungkinan karena ketidakpuasan terhadap layanan atau solusi yang diberikan.

Total Day Charge



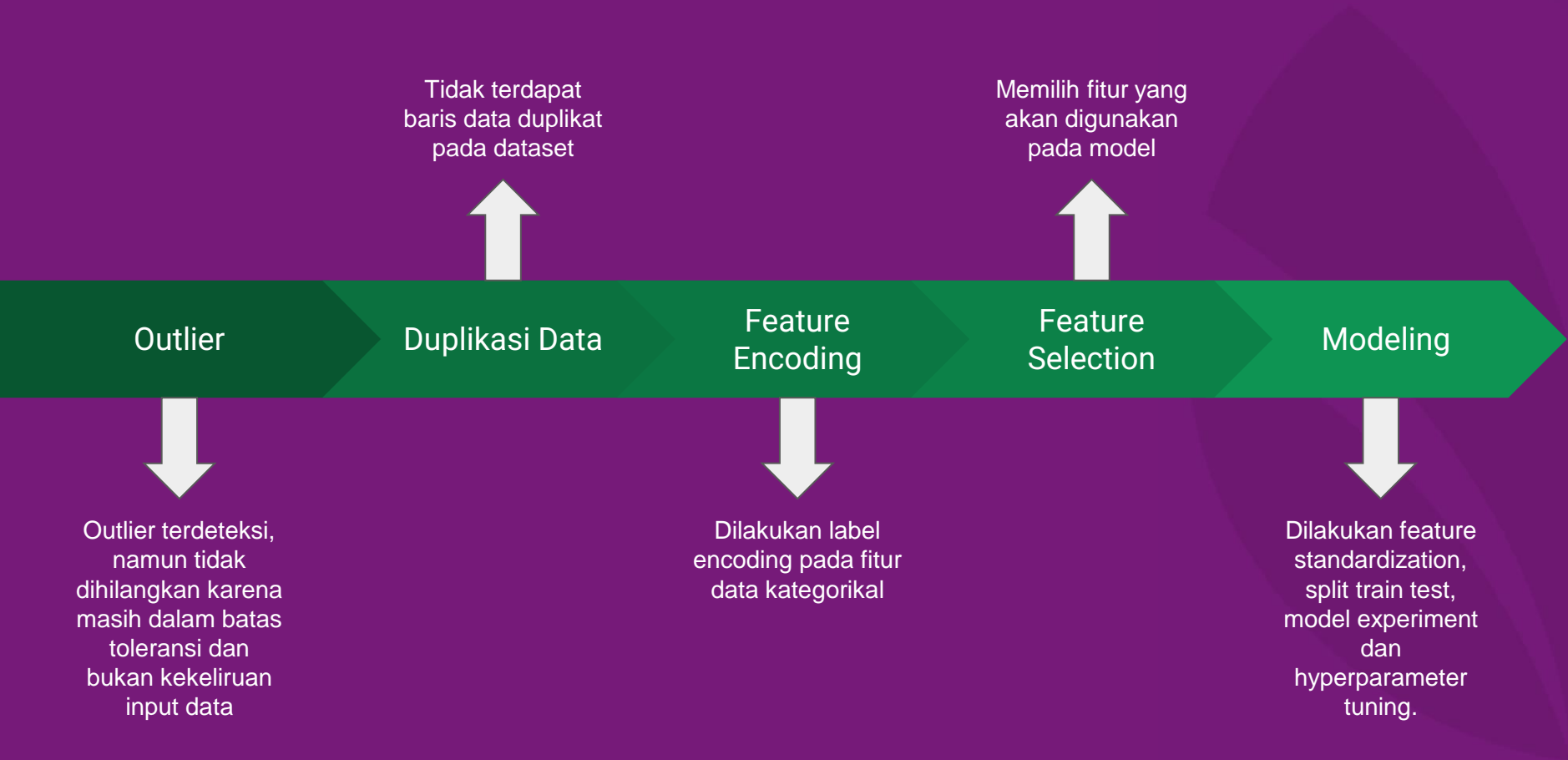
Grafik menunjukkan bahwa harga panggilan pagi hari, terutama untuk durasi lama, memiliki pengaruh besar pada churn pelanggan. Pelanggan yang melakukan panggilan lama di pagi hari kemungkinan churn karena merasa harga yang mereka bayar terlalu mahal.

Charge vs Minute

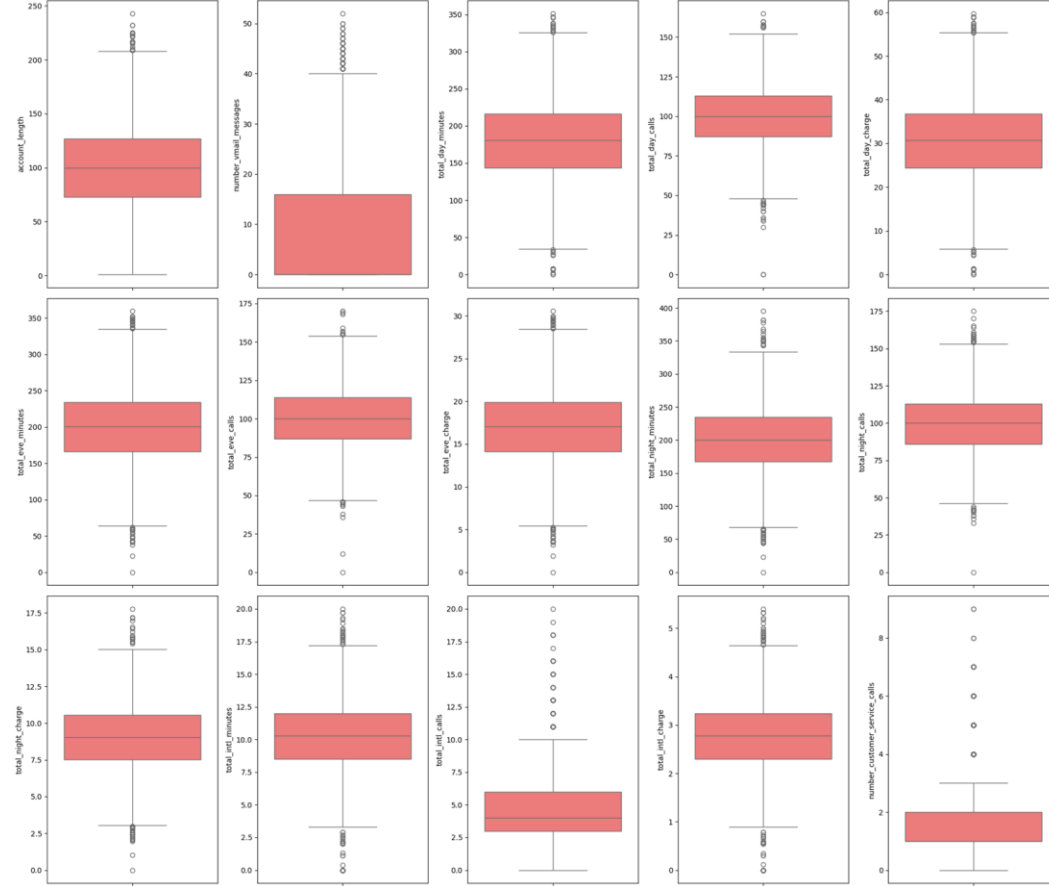


Pelanggan yang melakukan panggilan internasional cenderung berbicara lebih lama, meskipun biayanya paling mahal. Biaya murah di kategori Night tidak berpengaruh signifikan pada durasi panggilan.

Data Preprocessing



Outlier



```
churn                1.000000
international_plan   0.259053
number_customer_service_calls 0.221220
total_day_minutes    0.215272
total_day_charge     0.215263
voice_mail_plan      0.114643
number_vmail_messages 0.100347
total_eve_minutes    0.078855
total_eve_charge     0.078852
total_intl_minutes   0.055186
total_intl_charge    0.055177
total_night_minutes  0.046647
total_night_charge   0.046641
total_intl_calls     0.034334
account_length       0.019378
total_night_calls    0.012699
total_day_calls      0.011640
area_code            0.010696
state                0.007181
total_eve_calls      0.006817
Name: churn, dtype: float64
```

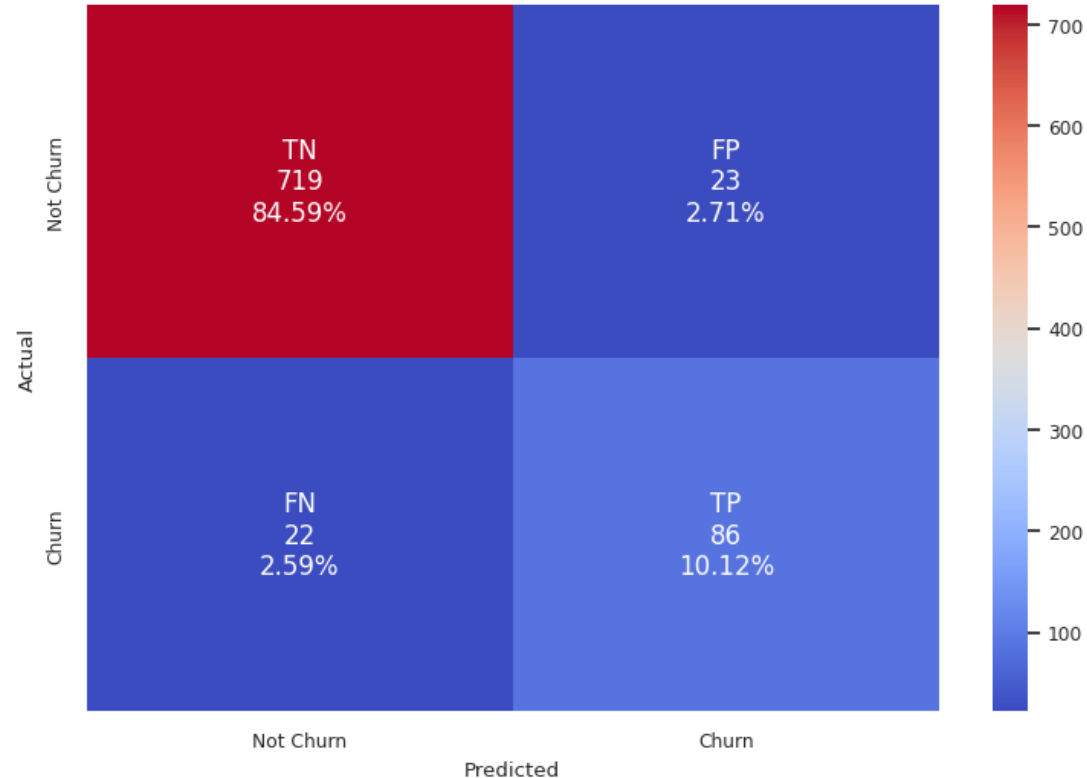
Kami menggunakan semua fitur untuk memungkinkan eksplorasi menyeluruh terhadap hubungan potensial dengan variabel target.

Modeling & Evaluation

Model	Akurasi	Recall	AUC Train	AUC Test
Logistic Regression	88%	60%	82%	84%
Random Forest Classification	95%	89%	94%	92%
K-Nearest Neighbors	90%	68%	96%	83%
XGBOOST	79%	81%	89%	86%
SVC	84%	85%	91%	91%

Model Evaluation

Recall berfokus pada kemampuan model untuk mengidentifikasi pelanggan yang benar-benar churn (TP). Model dapat mengidentifikasi sebagian besar pelanggan yang berpotensi berhenti menggunakan layanan (TP + FP), memungkinkan perusahaan untuk mengambil tindakan pencegahan dan mempertahankan mereka. Kesalahan prediksi pada pelanggan yang tidak churn (FN) juga memiliki nilai paling rendah, menunjukkan ketepatan model yang tinggi dalam mengidentifikasi pelanggan yang tidak akan churn.



s	total_day_calls	total_day_charge	...	total_eve_calls	total_eve_charge	total_night_minutes	total_night_calls	total_night_charge	total_intl_minutes	total_intl_calls	total_intl_charge	number_customer_service_calls	predicted_Churn
1	110	45.07	...	99	16.78	244.7	91	11.01	10.0	3	2.70	1	0
4	98	37.98	...	101	18.75	203.9	118	9.18	6.3	6	1.70	0	1
7	70	20.52	...	76	26.11	203.0	99	9.14	13.1	6	3.54	4	1
7	114	32.42	...	111	18.55	129.6	121	5.83	8.1	3	2.19	3	1
3	76	21.13	...	112	23.55	250.7	115	11.28	15.5	5	4.19	3	1
..
4	99	20.30	...	97	19.24	202.7	111	9.12	11.3	7	3.05	0	1
2	118	30.12	...	84	22.99	241.8	112	10.88	12.3	2	3.32	3	1
2	90	31.31	...	73	21.83	213.6	113	9.61	14.7	2	3.97	3	1
6	89	23.90	...	128	14.69	212.4	97	9.56	13.6	4	3.67	1	1
8	67	32.10	...	92	14.59	224.4	89	10.10	8.5	6	2.30	0	1

Berdasarkan hasil pengujian dari data test, prediksi customer melakukan churn sebanyak 539 orang, diprediksi customer yang tidak melakukan churn sebanyak 211 orang.

Business Recommendation

1. Segmentasi Pelanggan Berdasarkan Lokasi dan Tingkat Churn:

- Prioritaskan strategi retensi pelanggan di lima negara bagian dengan tingkat churn minimal 5%, yaitu WV, MN, ID, AL, dan VA.
- Melakukan analisis mendalam untuk memahami penyebab churn di negara-negara bagian tersebut dan menyesuaikan strategi retensi secara lokal.
- Membangun program loyalitas atau penawaran khusus untuk pelanggan di area-area tersebut untuk meningkatkan retensi.

2. Optimalisasi Penawaran dan Layanan Berdasarkan Kebiasaan Penggunaan:**

- Meninjau kembali biaya roaming dan kualitas jaringan untuk pelanggan dengan International Plan dan melakukan perbaikan yang diperlukan.
- Menawarkan paket roaming internasional yang lebih terjangkau atau meningkatkan kualitas jaringan di wilayah-wilayah tertentu.
- Mempertimbangkan penawaran khusus atau insentif untuk pelanggan yang melakukan panggilan internasional, mengingat mereka cenderung berbicara lebih lama.

3. Peningkatan Kualitas Layanan dan Pengalaman Pelanggan:

- Memberikan insentif atau promosi untuk pelanggan yang mengaktifkan voice mail plan, meningkatkan kesadaran akan manfaatnya.
- Mengidentifikasi dan memperbaiki masalah yang sering dilaporkan oleh pelanggan yang menghubungi customer service lebih dari 4 kali, untuk meningkatkan kepuasan pelanggan.

4. Penyesuaian Strategi Harga Berdasarkan Kebutuhan Pelanggan:

- Meninjau kembali struktur harga panggilan pagi hari, terutama untuk panggilan dengan durasi lama, untuk mengurangi kemungkinan churn.
- Memperkenalkan paket atau diskon khusus untuk panggilan pagi hari yang sesuai dengan kebutuhan pelanggan dan memberikan nilai tambah yang jelas.
- Menyediakan opsi harga yang kompetitif untuk panggilan internasional, sambil tetap mempertahankan profitabilitas perusahaan.

Thank You