

zenius

Kampus
Merdeka
INDONESIA JAYA

Final Project Presentation

Nomor Kelompok: 3

Nama Mentor: Erwin Fernanda

Nama:

- Nailia Ziyada
- Firdaus

Machine Learning Class

Program Studi Independen Bersertifikat
Zenius Bersama Kampus Merdeka



1. Latar Belakang
2. Explorasi Data dan Visualisasi
3. Modelling
4. Kesimpulan

Latar Belakang

Latar Belakang Project

Sumber Data: <https://www.kaggle.com/datasets/barun2104/telecom-churn?datasetId=567482>

Problem: **Classification**

Tujuan:

- Menggali insight untuk mengetahui faktor yang mempengaruhi *customer churn*
- Membuat saran untuk mengatasi *customer churn*
- Membuat model machine learning untuk memprediksi *customer churn*

Explorasi Data dan Visualisasi

Business Understanding

Customer Churn in Telecommunication Industry

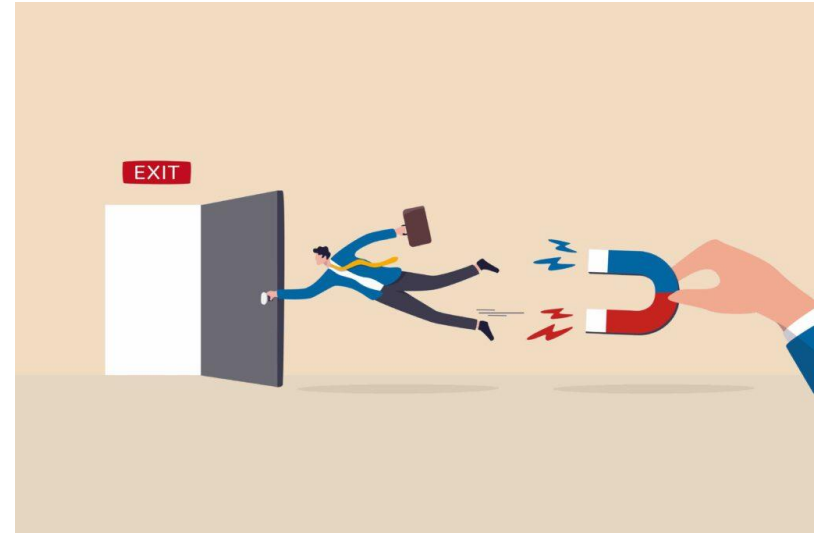
Definisi

Churning is when a customer stops using a company's service thereby opting for the next available service provider.



Why It Matters?

- Churn dapat mempengaruhi CAC (*Customer Acquisition Cost*) dan mengurangi pendapatan perusahaan
- Biaya untuk mempertahankan customer lebih rendah dibanding menggaet customer baru
- Berpengaruh langsung terhadap biaya operasional dan marketing perusahaan
- Jika churn disebabkan layanan buruk, maka reputasi perusahaan menjadi taruhan



Faktor	Macam Perilaku Churn
<ul style="list-style-type: none">- Kurang atau rendahnya kualitas dari <i>customer support</i>- Pengalaman customer yang negatif- Adanya kompetitor yang mempunyai layanan baik dengan harga yang bersaing- perubahan prioritas customer- customer tidak puas dengan layanan dalam waktu yang cukup lama- ekspektasi customer tidak terpenuhi oleh layanan yang diberikan	<ul style="list-style-type: none">- Pembatalan kontrak langganan- Berganti jasa penyedia layanan (kompetitor)- Tidak memperbarui kontrak langganan- Penutupan akun

Konteks Dataset

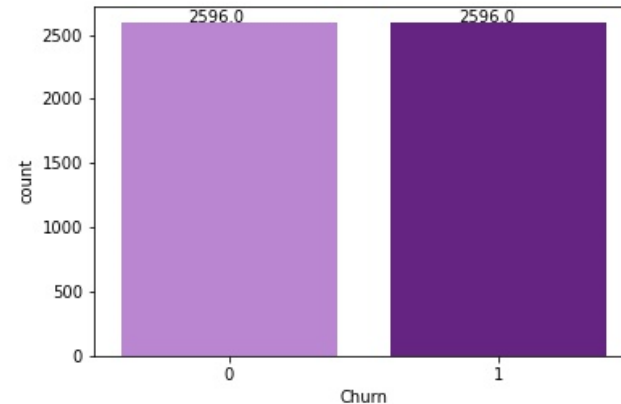
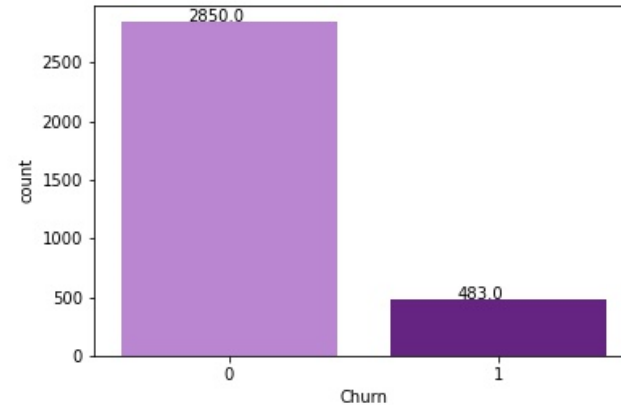
Dataset churn yang digunakan adalah dataset milik suatu perusahaan penyedia layanan di bidang telekomunikasi. Kolom yang tersedia adalah sebagai berikut:

- | | |
|--|---|
| 1. Churn
Apakah customer churn (1) atau tidak (0) | 7. DayMins
Rata-rata menit panggilan <i>daytime</i> per bulan |
| 1. AccountWeeks
Jumlah minggu akun customer aktif | 8. DayCalls
Rata-rata menit panggilan <i>daytime</i> |
| 1. Contract Renewal
Apakah customer memperbarui kontrak (1) atau tidak (0) | 9. MonthlyCharge
Rata-rata biaya bulanan |
| 1. DataPlan
Apakah customer memiliki data plan (1) atau tidak (0) | 10. OverageFee
<i>Overage fee</i> terbesar 12 tahun terakhir |
| 1. DataUsage
Penggunaan data bulanan (Gb) | 11. RoamMins
Rata-rata jumlah menit <i>roaming</i> |
| 1. CustServCalls
Jumlah panggilan ke <i>customer service</i> | |

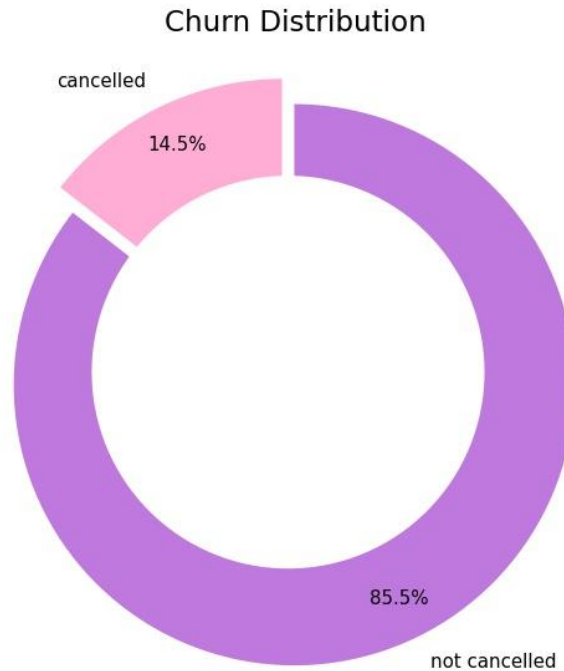
Data Cleansing

- Data tidak memiliki missing value
- Kolom 'DataPlan' dan 'DataUsage' memiliki korelasi sangat tinggi sehingga dihilangkan salah satu
- Terdapat outlier sehingga perlu dibersihkan Kolom yang dibersihkan:
'AccountWeeks', 'CustServCalls', 'DayMins', 'DayCalls', 'MonthlyCharge', 'OverageFee', 'RoamMins'
- Data masih tidak seimbang maka perlu dilakukan resampling

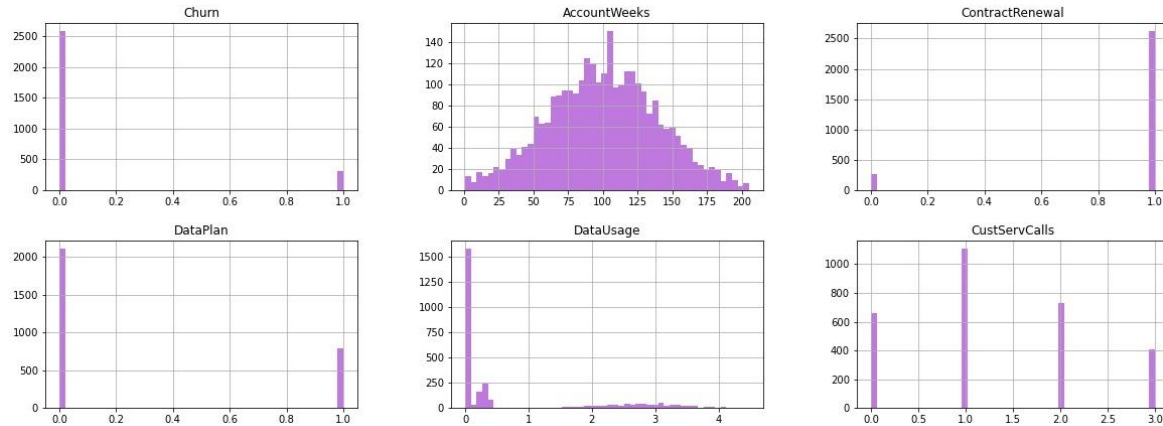
Total data : 5192 rows × 10 columns



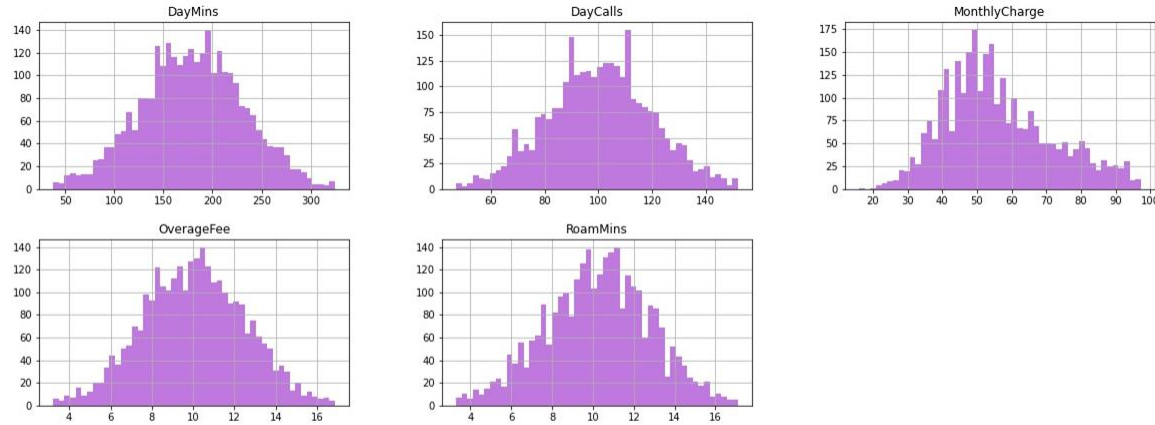
Exploratory Data Analysis



Exploratory Data Analysis



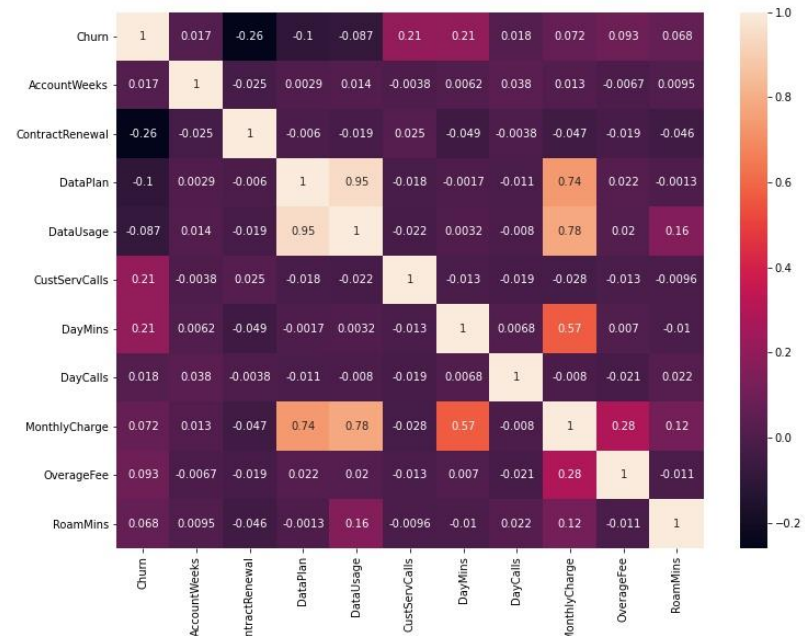
Exploratory Data Analysis



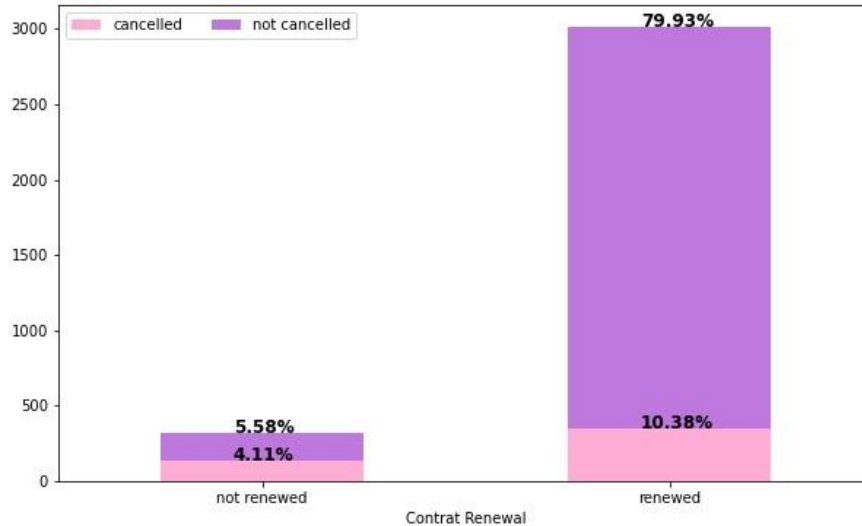
Exploratory Data Analysis

Terdapat beberapa korelasi tinggi di antara 2 variabel di antaranya:

1. MonthlyCharge dengan DataPlan
2. MonthlyCharge dengan DataUsage
3. DataUsage dengan DataPlan
4. MonthlyCharge dengan DayMins



Exploratory Data Analysis

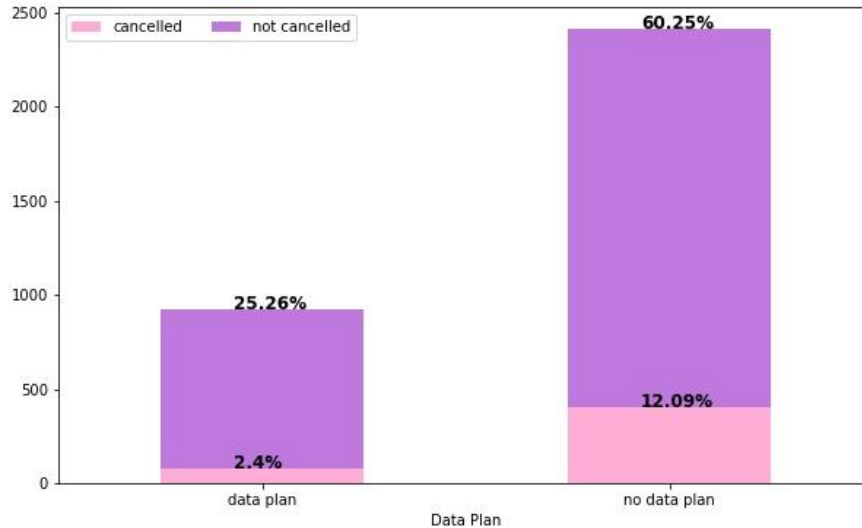


	Churn	cancelled	not cancelled
ContractRenewal			
not renewed	42.414861	57.585139	
renewed	11.495017	88.504983	

Insight:

- Dapat diketahui bahwa customer cenderung Churn ketika mereka tidak memperbarui kontrak dengan persentase 42% dari total customer yang tidak memperbarui kontrak.
- Hampir 80% customer yang memperbarui kontrak tetap loyal

Exploratory Data Analysis

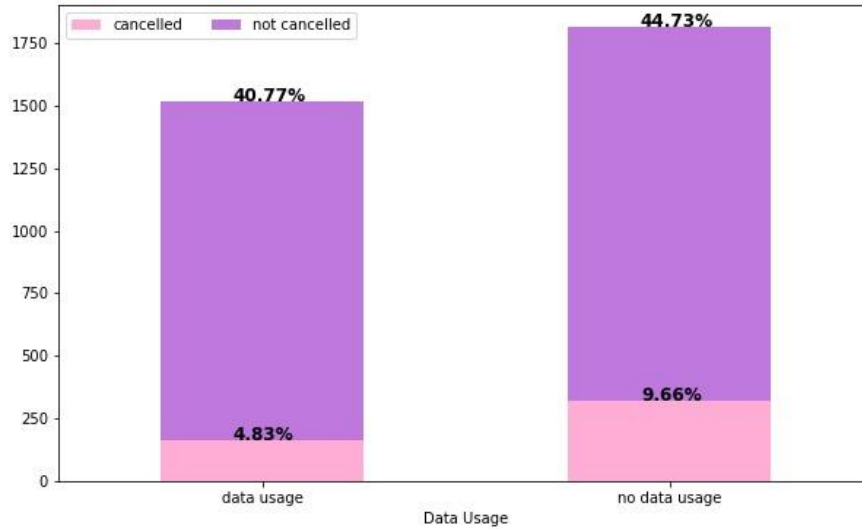


Churn	cancelled	not cancelled
DataPlan		
data plan	8.676790	91.323210
no data plan	16.715056	83.284944

Insight:

- Sekitar 72% dari total customer tidak memiliki data plan, dan 16% di antaranya adalah customer churn.
- Dari total semua customer, 12% customer churn tidak memiliki data plan. Persentase ini terlihat cukup signifikan dibandingkan dengan persentase customer churn yang memiliki data plan yaitu hanya 2% dari total customer keseluruhan.

Exploratory Data Analysis

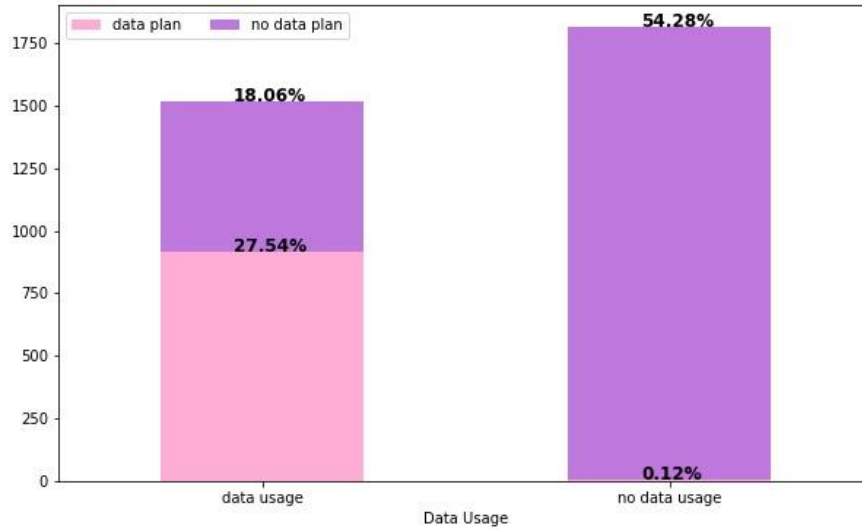


Churn	cancelled	not cancelled
DataUse		
data usage	10.592105	89.407895
no data usage	17.760618	82.239382

Insight:

- Dari keseluruhan customer yang churn, lebih banyak dari mereka yang tidak memiliki data usage.
- Lebih banyak customer yang tidak memiliki data usage yaitu dengan persentase 54%.

Exploratory Data Analysis



DataPlan data plan no data plan

DataUse

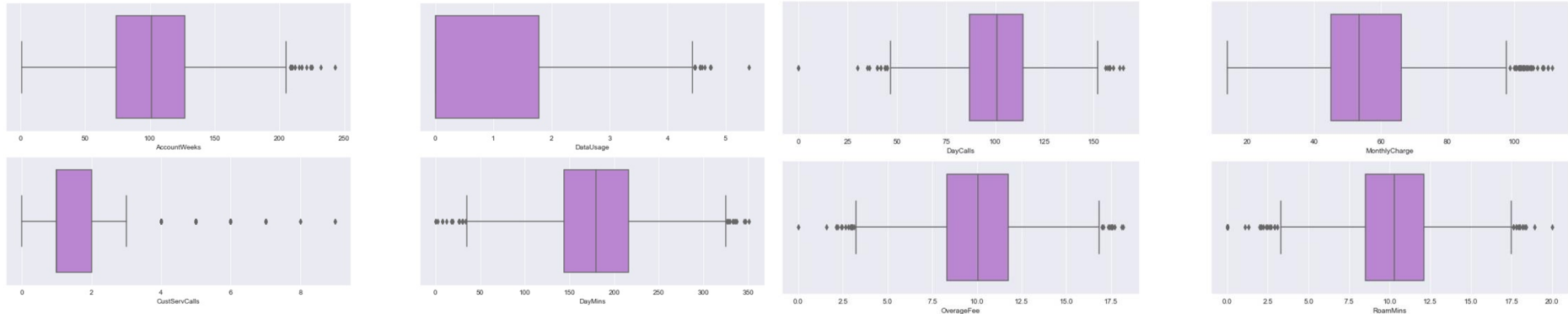
	data plan	no data plan
data usage	60.394737	39.605263
no data usage	0.220629	99.779371

Insight:

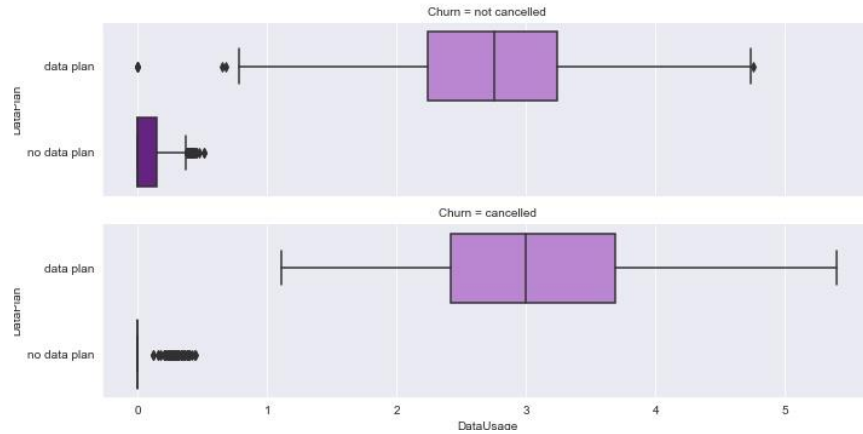
- Hampir 100% customer yang tidak memiliki data usage juga tidak memiliki data plan.
- Customer yang tidak memiliki data usage perlu ditandai sebagai customer yang memiliki kecenderungan untuk churn dibanding yang memiliki data usage.

Exploratory Data Analysis

Boxplot

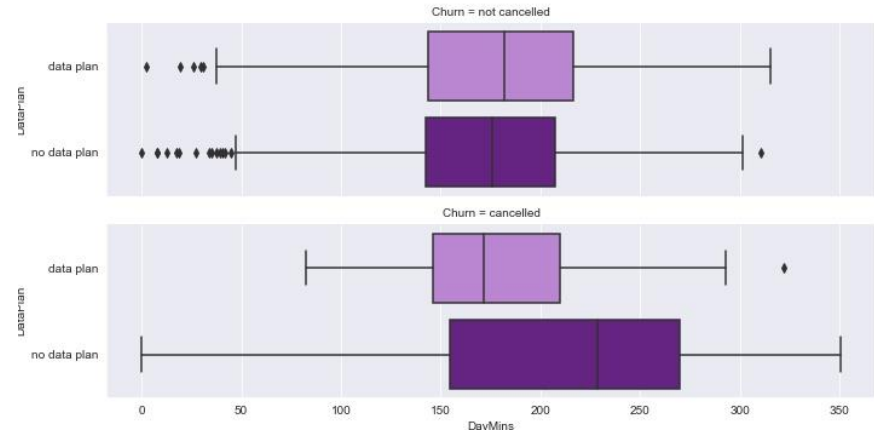


Exploratory Data Analysis



Insight:

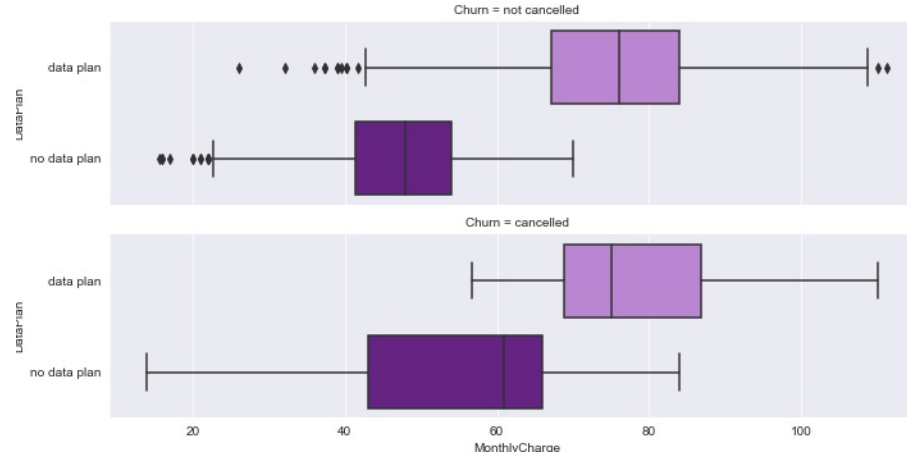
- customer yang churn menggunakan data lebih banyak dari rata-rata dibandingkan dengan yang tidak
- separuh dari pengguna tanpa data plan yang tidak churn memiliki data usage, sedangkan pada customer yang churn hanya outlier saja yang memiliki data usage.



Insight:

- Rata-rata customer churn yang tidak memiliki data plan lebih tinggi dibandingkan dengan yang memiliki data plan.

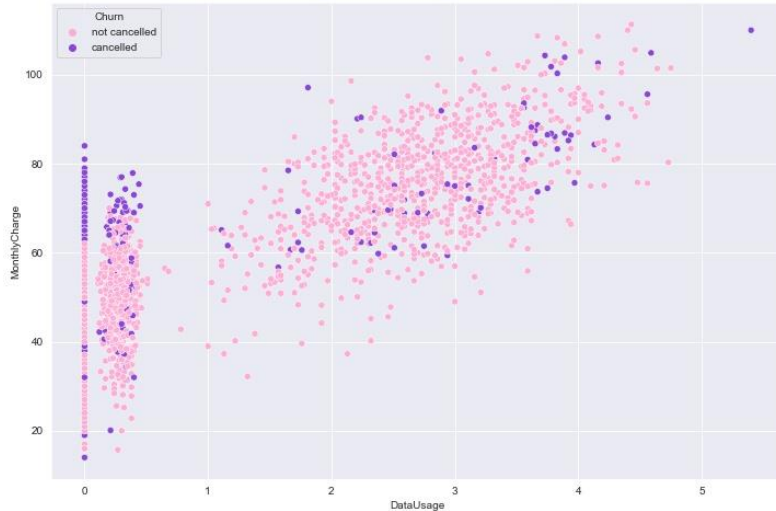
Exploratory Data Analysis



Insight:

- Rata-rata customer churn yang tidak memiliki data plan lebih tinggi dibandingkan dengan customer yang tidak churn dan tidak memiliki data plan.

Exploratory Data Analysis



Insight:

- Semakin tinggi data usage customer maka semakin besar pula monthly charge-nya.
- Kebanyakan customer yang churn adalah yang memiliki monthly charge tinggi namun data usage kecil (kurang dari 1 sampai 0)

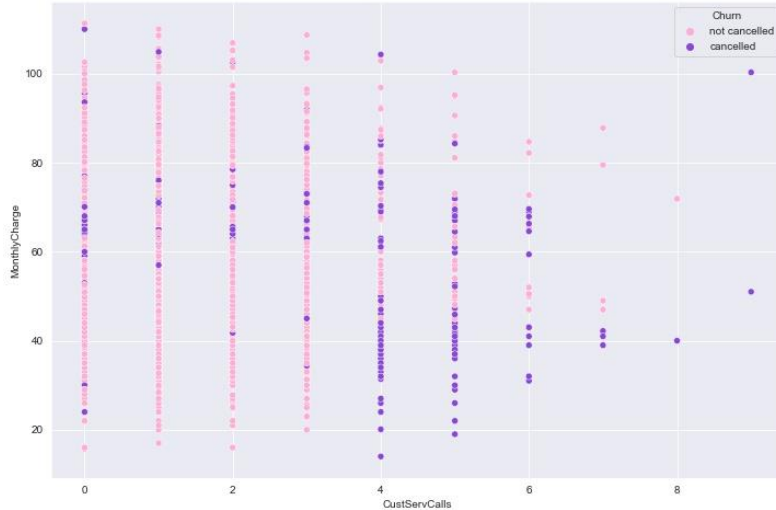
Exploratory Data Analysis



Insight:

- Semakin tinggi DayMins maka semakin tinggi pula MonthlyCharge
- Customer churn lebih banyak terjadi untuk DayMins dan MonthlyCharge yang tinggi di atas rata-rata

Exploratory Data Analysis



Insight:

- Customer yang cenderung churn banyak yang melakukan customer service calls lebih dari rata-rata dan memiliki monthly charge kurang lebih di bawah rata-rata.

Modelling

Perbandingan Model

Random Forest & Logistic Regression

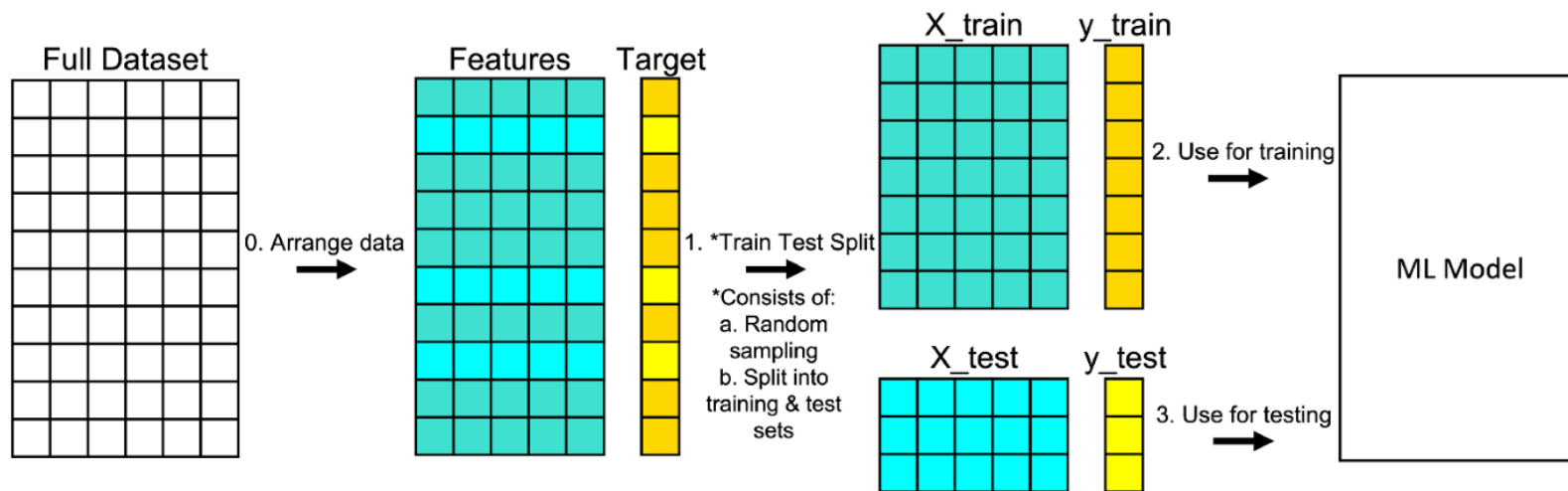
Prediktor

1. 'AccountWeeks'
2. 'CustServCalls'
3. 'DayMins'
4. 'DayCalls'
5. 'MonthlyCharge'
6. 'OverageFee'
7. 'RoamMins'
8. 'ContractRenewal'
9. 'DataPlan'

Target

1. 'Churn'

Train Test Split Method

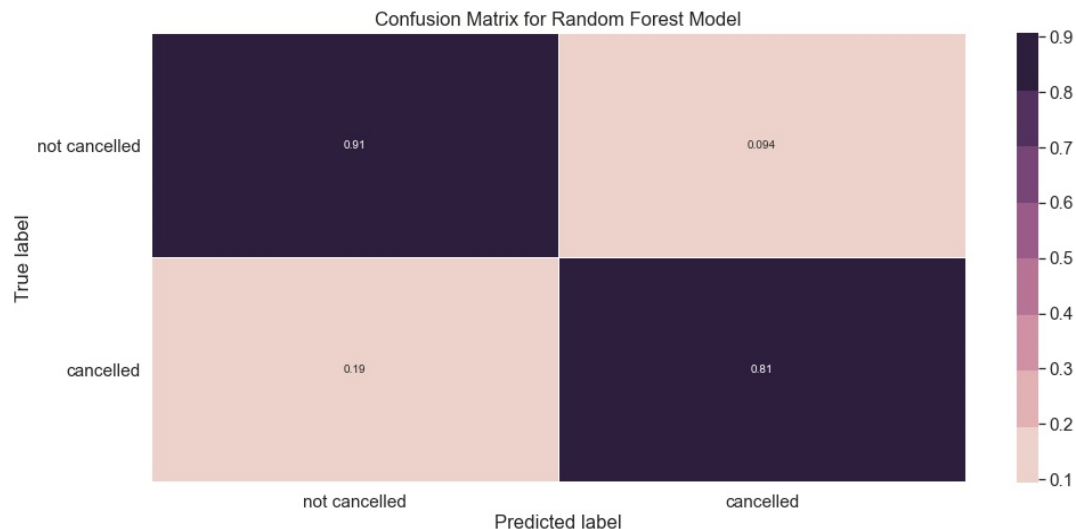


Sumber: <https://towardsdatascience.com/understanding-train-test-split-scikit-learn-python-ea676d5e3d1>

Random Forest

(Model 1: Original Model)

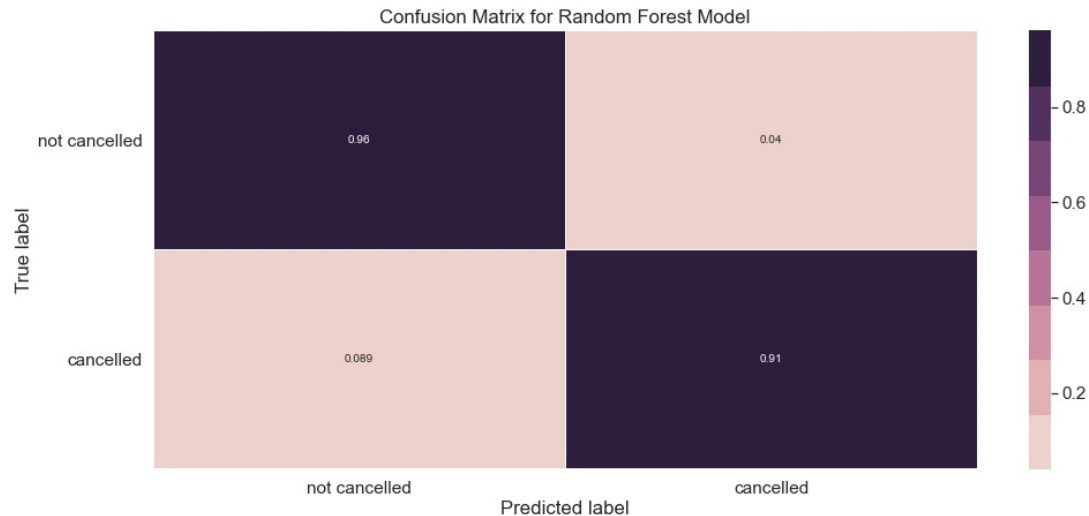
	Precision	recall	f1-score	support
0	0.82	0.91	0.86	502
1	0.90	0.81	0.86	537
accuracy			0.86	1039
macro avg	0.86	0.86	0.86	1039
weighted avg	0.86	0.86	0.86	1039



Random Forest

(Model 2: Hyperparameter Tuning)

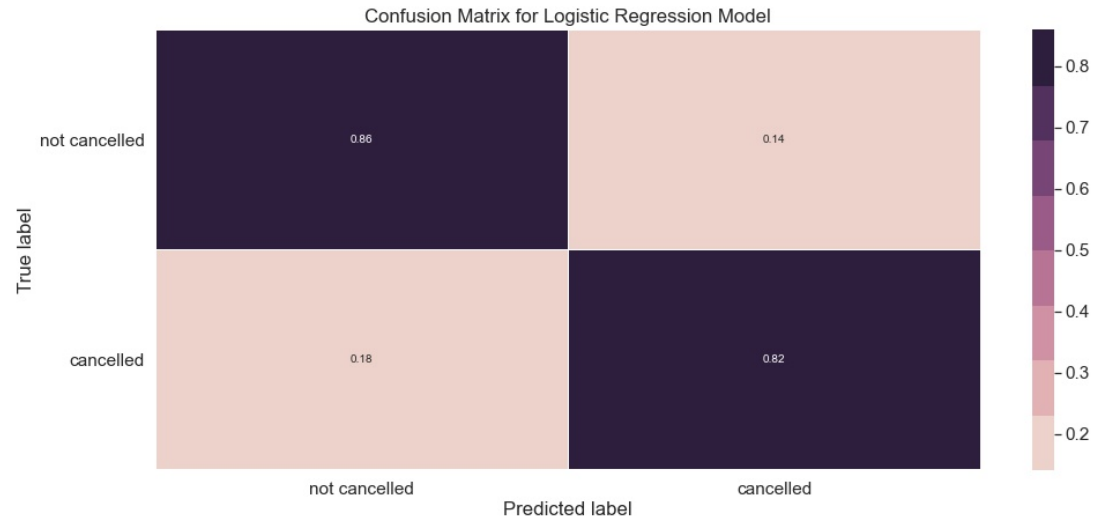
	Precision	recall	f1-score	support
0	0.91	0.96	0.93	502
1	0.96	0.91	0.93	537
accuracy			0.93	1039
macro avg	0.94	0.94	0.93	1039
weighted avg	0.94	0.93	0.93	1039



Logistic Regression

(Model 1: Original Model)

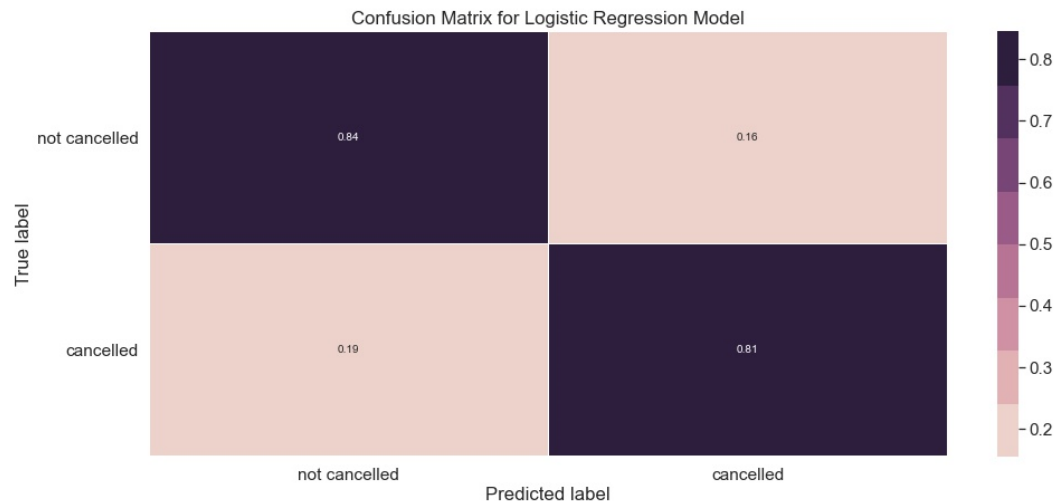
	Precision	recall	f1-score	support
0	0.81	0.86	0.84	502
1	0.86	0.82	0.84	537
accuracy			0.84	1039
macro avg	0.84	0.84	0.84	1039
weighted avg	0.84	0.84	0.84	1039



Logistic Regression

(Model 2: Hyperparameter Tuning)

	Precision	recall	f1-score	support
0	0.81	0.84	0.83	502
1	0.85	0.81	0.83	537
accuracy			0.83	1039
macro avg	0.83	0.83	0.83	1039
weighted avg	0.83	0.83	0.83	1039



Evaluation

Metrik	Model	
	Random Forest Model 2	Logistic Regression Model 1
Area Under the Curve (AUC)	0.94	0.84
Accuracy	0.93	0.84
Precision	0.94	0.84
Recall	0.94	0.84

Conclusion

Recap Insight

- Customer yang memperbarui kontrak cenderung tetap loyal
- Customer yang tidak memiliki data plan cenderung untuk churn
- Sekitar 1 dari 5 customer yang tidak memiliki data usage berpotensi churn
- Monthly charge tinggi dengan penggunaan data kecil dapat mempengaruhi customer churn
- Customer churn lebih banyak terjadi untuk DayMins dan MonthlyCharge yang tinggi di atas rata-rata
- Customer yang banyak melakukan Customer Service Call dengan MonthlyCharge kurang lebih di bawah rata-rata cenderung untuk Churn

Saran

- Mengevaluasi pelayanan dan membandingkannya dengan perusahaan kompetitor
- Memperbarui strategi dengan membuat paket harga layanan yang tidak hanya menguntungkan perusahaan tapi juga customer
- Meningkatkan pelayanan customer service sehingga kepuasan pelanggan lebih dapat tercapai
- Membuat promo menarik sesuai dengan kebutuhan pelanggan, misal paket data/panggilan dengan harga murah atau memberi bonus.

Terima kasih!

Ada pertanyaan?

zenius



Kampus
Merdeka
INDONESIA JAYA