Final Pitching

Intern Data Science | Basic Computing Community

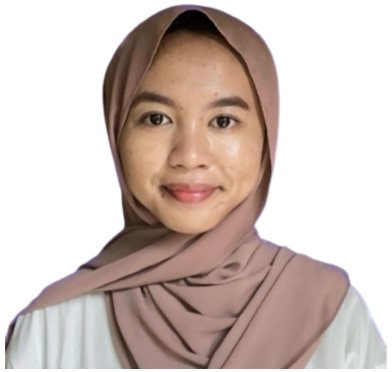
by Kelompok 2

tim

Bismillah Lolos 🥺

a.k.a. Kelompok 2

Our Team



Ervi N.F.

Sistem Informasi ’21
as Datactive

Kak Cesil

Teknik Informatika’20
as Mentor

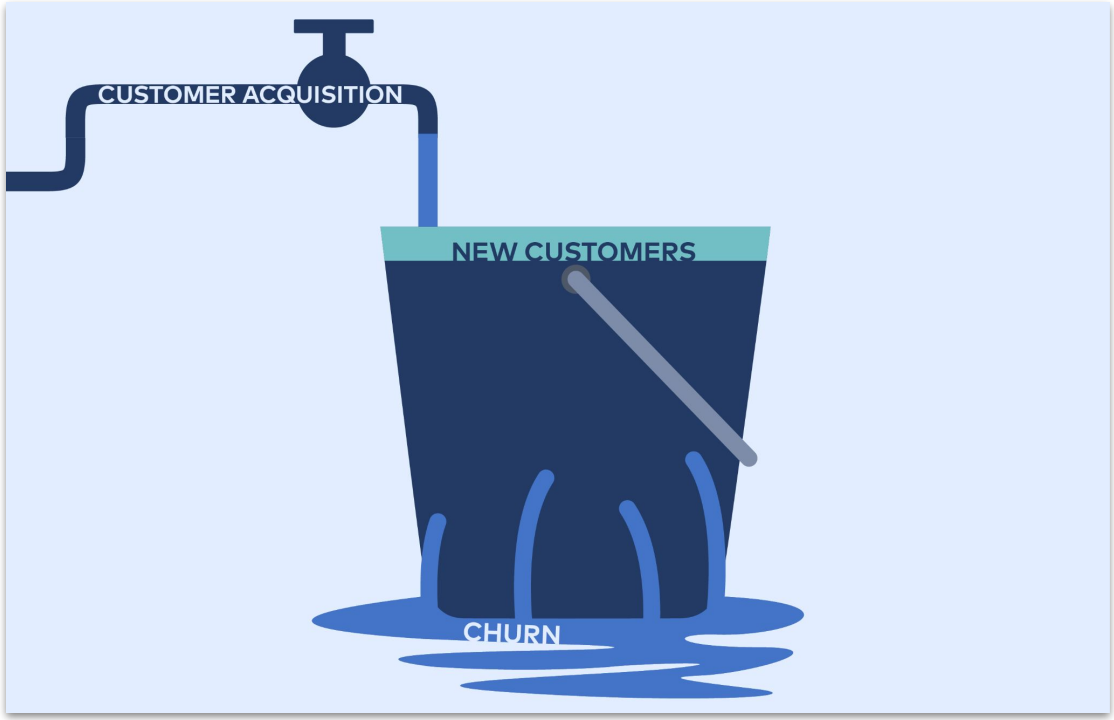
M Husain F

Teknik Informatika ’22
as Datactive

Mission?

Telkomsel Cabang Malang

Churn?



Churn = Pelanggan yang keluar dalam sebulan terakhir


HERE
WE
GO
SOCIOS.com

the dataset we choose is

Telco Customer Churn

Dataset ini mencakup 21 variabel atau kolom yang berisi informasi tentang karakteristik pelanggan seperti demografi, jenis kontrak, jenis layanan yang digunakan, durasi penggunaan layanan, tagihan bulanan, dan apakah pelanggan tersebut berhenti atau tetap menjadi pelanggan.

link kaggle https://www.kaggle.com/datasets/blastchar/telco-customer-churn

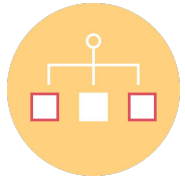
Our Today's Topics??

Objectives

Business Understanding

Data Understanding

Data Preprocessing



Modelling

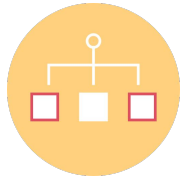
Evaluation

Objectives

Business Understanding

Data Understanding

Data Preprocessing



Modelling

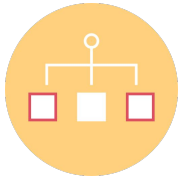
Evaluation

Objectives

Business Understanding

Data Understanding

Data Preprocessing



Modelling

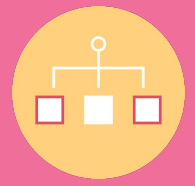
Evaluation

Objectives

Business Understanding

Data Understanding

Data Preprocessing



Modelling

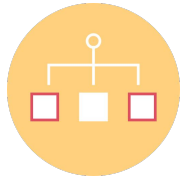
Evaluation

Objectives

Business Understanding

Data Understanding

Data Preprocessing



Modelling

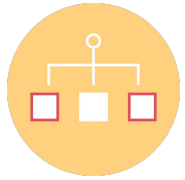
Evaluation

Objectives

Business Understanding

Data Understanding

Data Preprocessing



Modelling

Evaluation

Business Understanding

a. Problem formulation

Perkembangan teknologi informasi telah membawa perubahan dalam dunia bisnis dan meningkatkan ketersediaan data dan informasi. Dalam hal ini, machine learning dianggap sebagai solusi untuk memproses data tersebut. Dalam model bisnis berlangganan, keberhasilannya tergantung pada tingkat churn, yaitu jumlah pelanggan yang berhenti berlangganan. Oleh karena itu, digitalisasi dan strategi Customer relationship management (CRM) menjadi penting untuk meningkatkan aktivitas pemrosesan data dan meminimalkan churn.

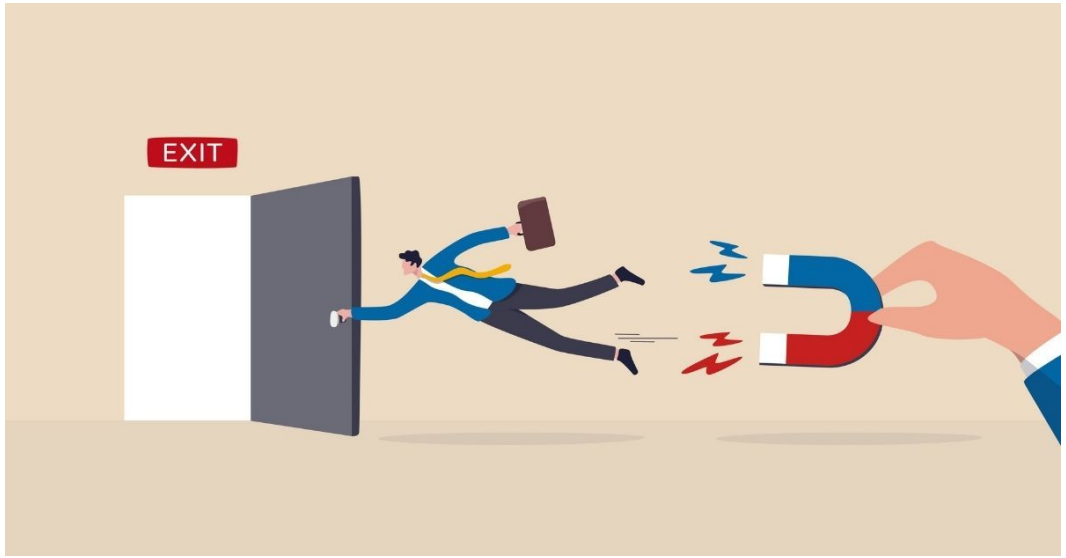
Business Understanding

CRM (Customer Realitionship Management)

RM merujuk pada proses dan strategi yang digunakan oleh perusahaan untuk mengelola dan meningkatkan hubungan dengan pelanggan. Tujuannya adalah untuk memahami kebutuhan pelanggan, meningkatkan pengalaman pelanggan, dan membangun loyalitas pelanggan yang lebih kuat. CRM dapat melibatkan berbagai taktik, seperti pengumpulan data pelanggan, analisis data, personalisasi komunikasi, dan layanan pelanggan yang lebih baik.

Business Understanding

b. Purpose



Business Objective:

- **Maximize** : Keuntungan perusahaan dengan mempertahankan pelanggan.
- **Minimize** : Pelanggan churn dengan mengidentifikasi penyebab utama masalah.

Business Understanding

b. Purpose

Main Objective of Our Project:

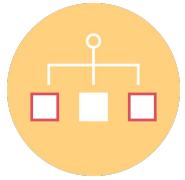
- Menemukan faktor dan penyebab yang mempengaruhi pelanggan untuk churn.
- Memprediksi apakah seorang pelanggan akan churn atau tidak.
- Mengembangkan strategi untuk meminimalkan churn rate.

Objectives

Business Understanding

Data Understanding

Data Preprocessing



Modelling

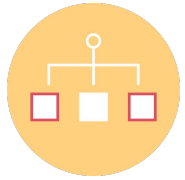
Evaluation

Objectives

Business Understanding

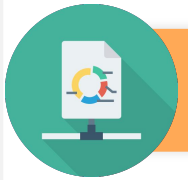
Data Understanding

Data Preprocessing



Modelling

Evaluation



Data Understanding

a. Data Description

(1) Demographic Information

- `gender` : Whether the client is a female or a male (Female, Male).
- `SeniorCitizen` : Indicates if the customer is 65 or older (0, 1).
- `Partner` : Whether the client has a partner (married) or not (Yes, No).
- `Dependents` : Indicates if the customer lives with any dependents (Yes, No). Dependents could be children, parents, grandparents, etc.

(2) Customer Account Information

- `tenure` : Number of months the customer has stayed with the company (Multiple different numeric values).
- `Contract` : Indicates the customer's current contract type (Month-to-Month, One year, Two year).
- `PaperlessBilling` : Whether the client has paperless billing or not (Yes, No).
- `PaymentMethod` : The customer's payment method (Electronic check,
- Mailed check, Bank transfer (automatic), Credit Card (automatic)).
- `MontlyCharges` : The amount charged to the customer monthly (Multiple different numeric values).
- `TotalCharges` : The total amount charged to the customer (Multiple different numeric values).

(3) Services Information

- `PhoneService` : Whether the client has a phone service or not (Yes, No).
- `MultipleLines` : Whether the client has multiple lines or not (No phone service, No, Yes).
- `InternetServices` : Whether the client is subscribed to Internet service with the company (DSL, Fiber, optic, No)
- `OnlineSecurity` : Whether the client has online security or not (No internet service, No, Yes).
- `OnlineBackup` : Whether the client has online backup or not (No internet service, No, Yes).
- `DeviceProtection` : Whether the client has device protection or not (No internet service, No, Yes).
- `TechSupport` : Whether the client has tech support or not (No internet service, No, Yes).
- `StreamingTV` : Whether the client has streaming TV or not (No internet service, No, Yes).
- `StreamingMovies` : Whether the client has streaming movies or not (No internet service, No, Yes).

Data Understanding

b. Data Profile

Dataset Shape: (7043, 21)

	0	1	2	3	4	5	6	7	8	9	...	11	12	13	14	15	16	17	18	19	20
Name	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn
dtypes	object	object	int64	object	object	int64	object	object	object	object	...	object	object	object	object	object	object	object	float64	object	object
Missing	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
Uniques	7043	2	2	2	2	73	2	3	3	3	...	3	3	3	3	3	2	4	1585	6531	2
Sample Value	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...	No	No	No	No	Month-to-month	Yes	Electronic check	29.85	29.85	No

5 rows × 21 columns

Observation:

melihat data profile seperti feature, dtypes, Missing, Uniques dan Sample Value

Action:

- Menghapus customerID dari kumpulan data karena ID tersebut unik untuk setiap catatan dan oleh karena itu kami tidak akan menggunakannya sebagai variabel prediktor.
- Melakukan Label Encoding semua kolom kategorikal yang memiliki 2 nilai.
- Mengonversi Total Biaya menjadi tipe data numerik.

Data Understanding

b. Data Profile

Dataset Shape: (7043, 21)

	0	1	2	3	4	5	6	7	8	9	...	11	12	13	14	15	16	17	18	19	20
Name	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn
dtypes	object	object	int64	object	object	int64	object	object	object	object	...	object	object	object	object	object	object	object	float64	object	object
Missing	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
Uniques	7043	2	2	2	2	73	2	3	3	3	...	3	3	3	3	3	2	4	1585	6531	2
Sample Value	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...	No	No	No	No	Month-to-month	Yes	Electronic check	29.85	29.85	No

5 rows × 21 columns

Observation:

melihat data profile seperti feature, dtypes, Missing, Uniques dan Sample Value

Action:

- Menghapus customerID dari kumpulan data karena ID tersebut unik untuk setiap catatan dan oleh karena itu kami tidak akan menggunakannya sebagai variabel prediktor.
- Melakukan Label Encoding semua kolom kategorikal yang memiliki 2 nilai.
- Mengonversi Total Biaya menjadi tipe data numerik.

Data Understanding

c. Descriptive Statistics

 # statistics of numerical columns

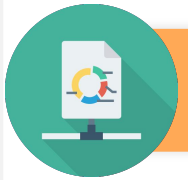
	count	mean	std	min	25%	50%	75%	max
SeniorCitizen	7043.0	0.162147	0.368612	0.00	0.0	0.00	0.00	1.00
tenure	7043.0	32.371149	24.559481	0.00	9.0	29.00	55.00	72.00
MonthlyCharges	7043.0	64.761692	30.090047	18.25	35.5	70.35	89.85	118.75

Dari statistik kolom numerik dapat disimpulkan beberapa hal sebagai berikut:

- Pada kolom ***SeniorCitizen*** Terdapat nilai 0 karena values hanya berisi 1 dan 0.
- pada kolom ***tenure*** terdapat nilai min 0, karena masa ***tenure*** minimalnya 0.

 # statistics of categorical columns

	customerID	gender	Partner	Dependents	PhoneService	MultipleLines	InternetService	OnlineSecurity	Onl
count	7043	7043	7043	7043	7043	7043	7043	7043	
unique	7043	2	2	2	2	3	3	3	
top	7590-VHVEG	Male	No	No	Yes	No	Fiber optic	No	
freq	1	3555	3641	4933	6361	3390	3096	3498	



Data Understanding

d. Duplicate

```
[ ]  print('Apakah terdapat baris yang duplikat ?', df.duplicated().any())
```

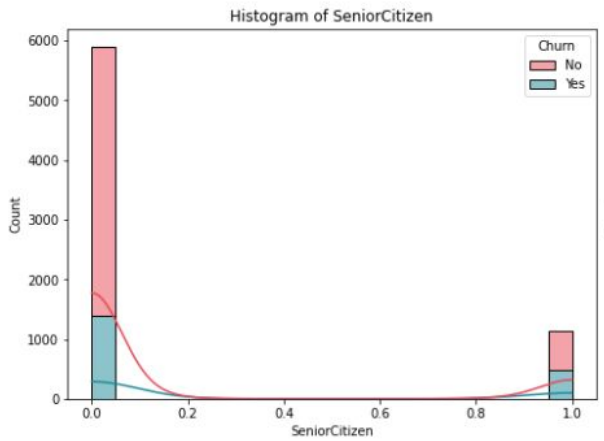
```
Apakah terdapat baris yang duplikat ? False
```

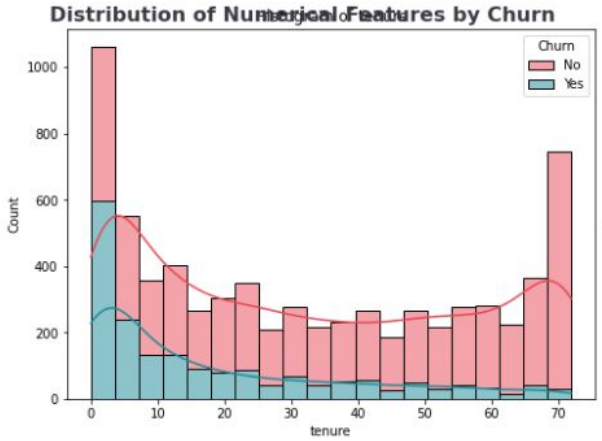
tidak ada baris yang berisi duplicate value

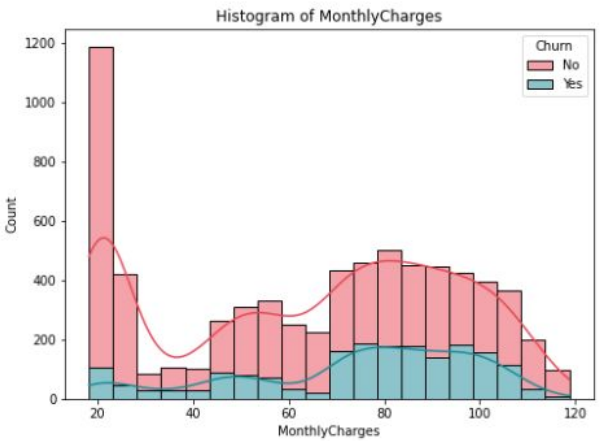
Data Understanding

h. Features by Target

EDA Numerical Features by Target





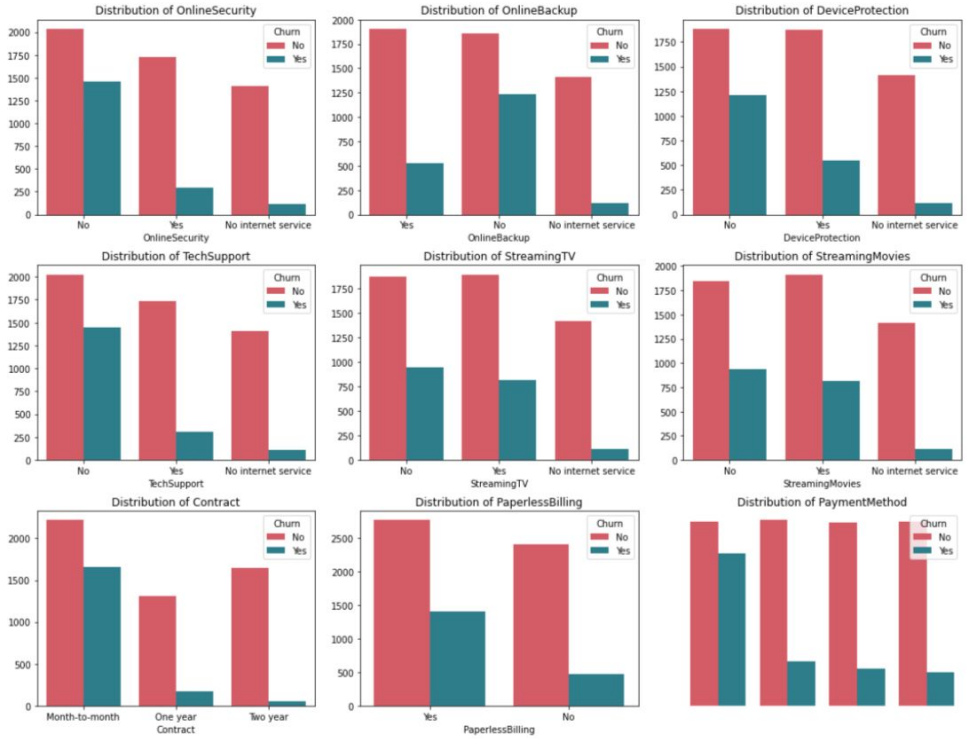


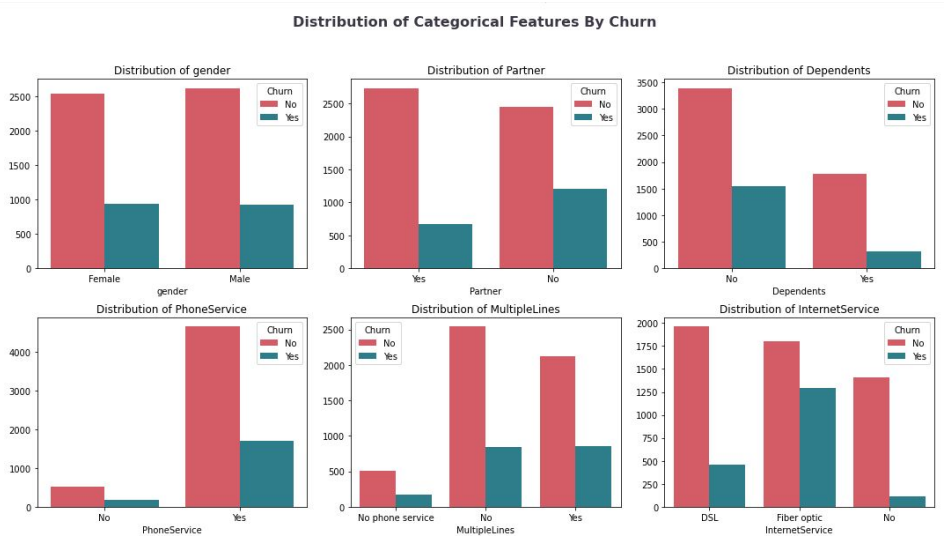
- SeniorCitizen : pelanggan yang lebih tua cenderung lebih cenderung untuk churn.
- tenure : pelanggan dengan tenure rendah (kurang dari sekitar 20 bulan) cenderung lebih cenderung untuk churn.
- MonthlyCharges : pelanggan dengan biaya bulanan tinggi cenderung lebih cenderung untuk churn.

Data Understanding

h. Features by Target

EDA Categorical Features by Target

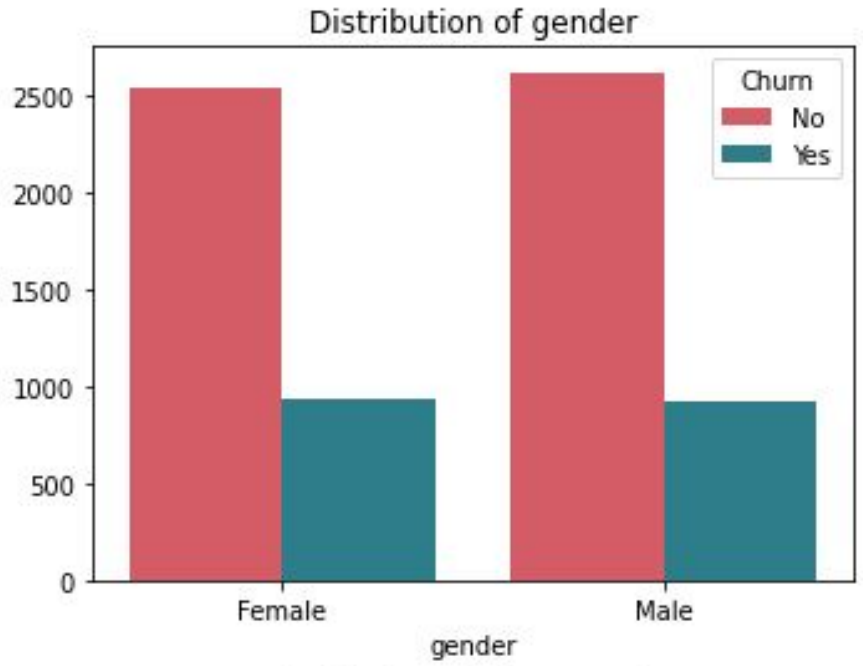


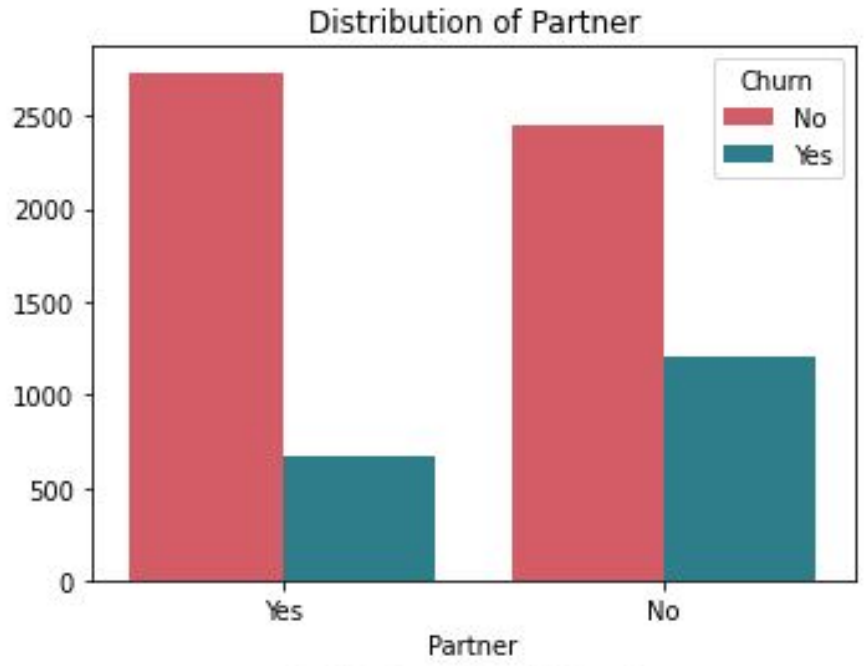


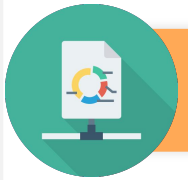
Data Understanding

h. Features by Target

EDA Categorical Features by Target







Data Understanding

i. EDA Conclusion

Berdasarkan EDA yang telah dilakukan, dapat ditarik kesimpulan sebagai berikut:

1. Dataset overview:

- Dataset berisi 7.043 baris dan 21 kolom.
- Tidak ada missing values dalam dataset.
- Terdapat 4 kolom biner, 9 kolom kategorikal, dan 8 kolom numerik pada dataset.
- Variabel targetnya adalah "Churn", yang memiliki 2 kelas: "Yes" dan "No".
- Proporsi "No" dalam variabel target lebih tinggi daripada "Yes".

2. Univariate Analysis of Numerical Columns:

- Kolom "tenure" dan "MonthlyCharges" memiliki distribusi bimodal.
- Kolom "SeniorCitizen" memiliki highly skewed distribution.
- Sebagian besar kolom numerik tidak memiliki outliers yang ekstrim.

3. Univariate Analysis of Categorical Columns:

- Mayoritas pelanggan adalah laki-laki (50,5%) dan tidak memiliki partners (51,7%).
- Sebagian besar pelanggan memiliki phone service (90,3%), tetapi hanya sedikit pelanggan yang memiliki multiple lines (42,7%).
- Mayoritas pelanggan menggunakan fiber optic (43,96%) dan tidak memiliki online security (49,7%).
- Sebagian besar pelanggan membayar menggunakan electronic check (33,6%).

Data Understanding

i. EDA Conclusion

4. EDA Numerical Features by Target:

- Pelanggan yang melakukan churning cenderung memiliki "MonthlyCharges" yang lebih tinggi dibandingkan pelanggan yang tidak melakukan churning.
- Pelanggan yang melakukan churning cenderung memiliki "tenure" yang lebih pendek dibandingkan dengan pelanggan yang tidak melakukan churning.
- "SeniorCitizen" tampaknya tidak memiliki pengaruh yang signifikan terhadap churn.

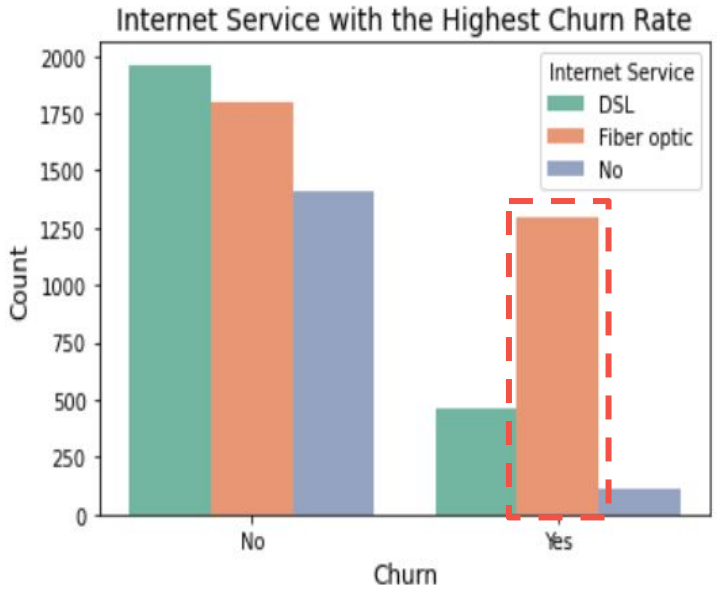
5. EDA Categorical Features by Target:

- Pelanggan wanita memiliki tingkat churn yang sedikit lebih tinggi daripada pelanggan pria.
- Pelanggan yang tidak memiliki partners atau dependents memiliki tingkat churn yang lebih tinggi.
- Pelanggan yang memiliki layanan internet iber optic internet service, online security, online backup, device protection, atau tech support memiliki tingkat churn yang lebih tinggi.
- Pelanggan yang memiliki month-to-month contract, paperless billing, atau pay using electronic check memiliki tingkat churn yang lebih tinggi.

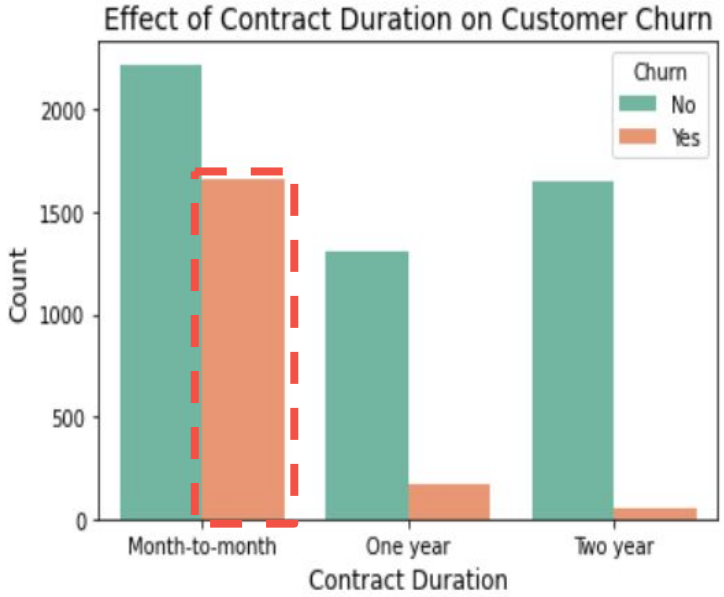
Data Understanding

j. Business Insight

1. Apakah jenis layanan yang memiliki tingkat churn paling tinggi?



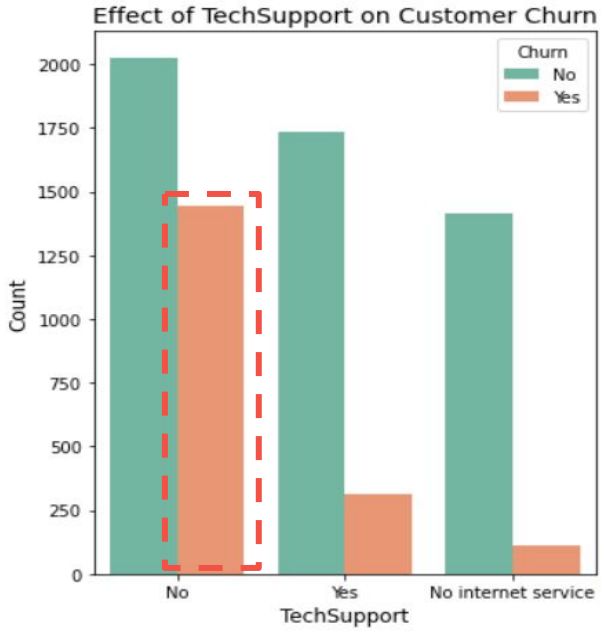
2. Apakah durasi kontrak berpengaruh terhadap keputusan pelanggan untuk berhenti langganan?

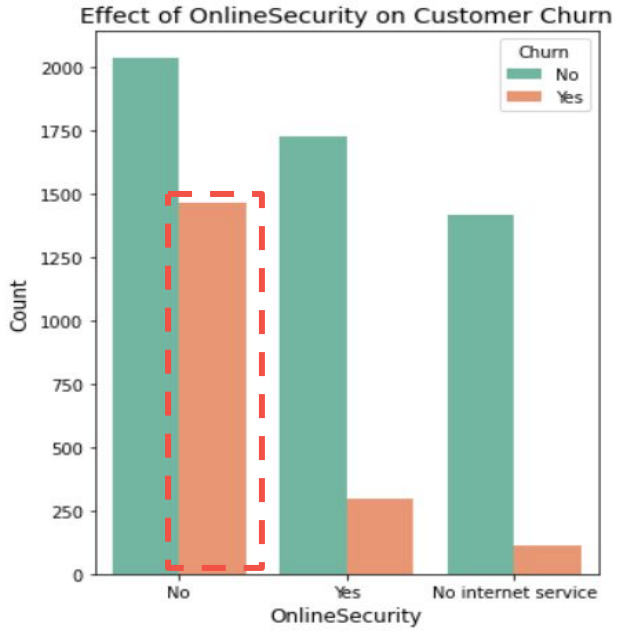


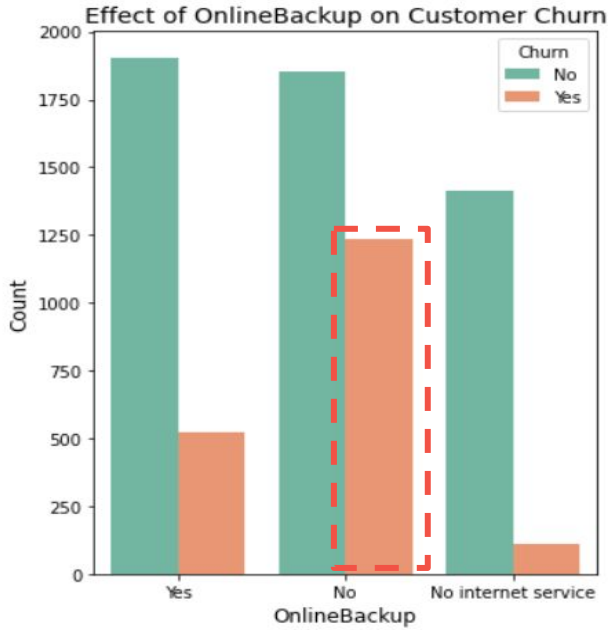
Data Understanding

j. Business Insight

3. Apakah ada perbedaan dalam tingkat churn antara pelanggan yang menggunakan layanan tambahan seperti tech support, online security, dan backup?



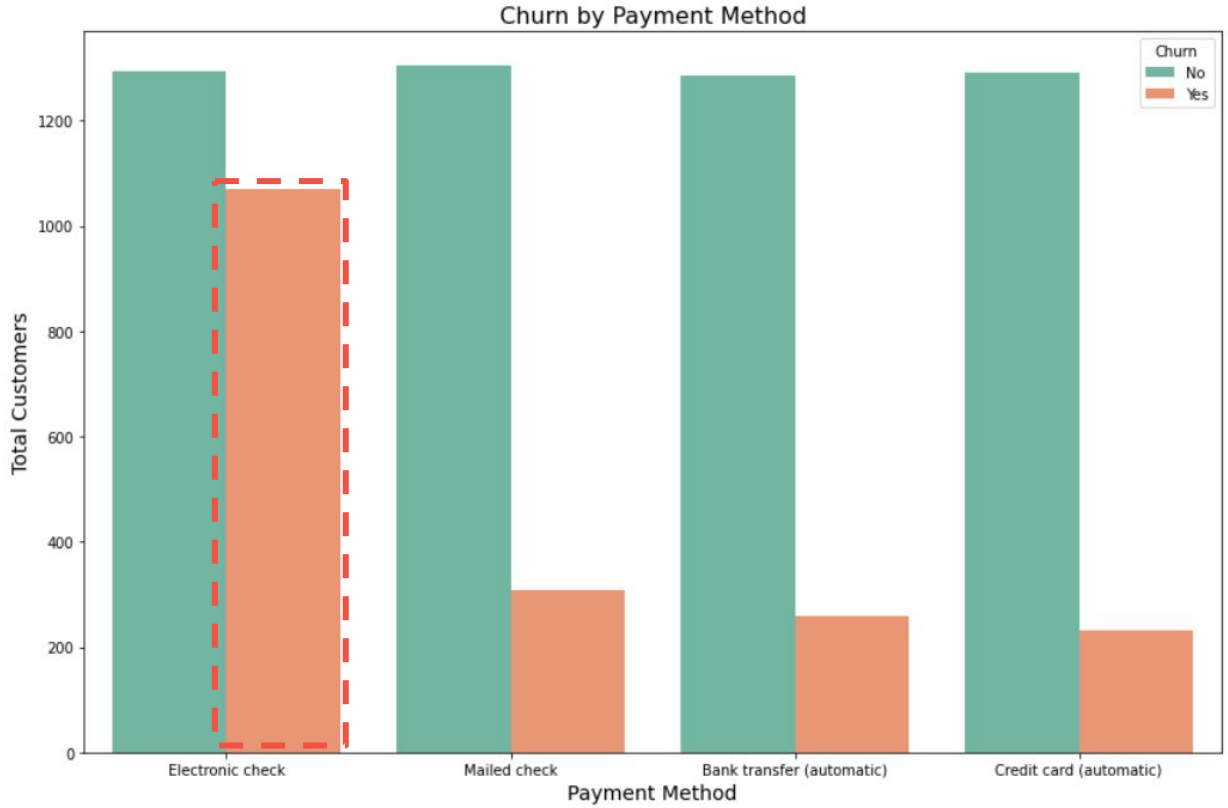




Data Understanding

j. Business Insight

4. Apakah metode pembayaran yang dipilih oleh pelanggan berpengaruh terhadap kecenderungan churn?

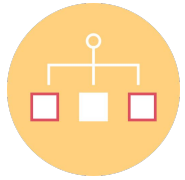


Objectives

Business Understanding

Data Understanding

Data Preprocessing



Modelling

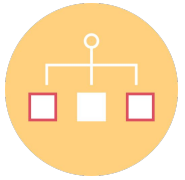
Evaluation

Objectives

Business Understanding

Data Understanding

Data Preprocessing



Modelling

Evaluation

Data Prepocessing

Data Cleaning

a. Handling Irrelevant Feature

```
df_prep = df.drop(['customerID', 'gender', 'PhoneService','MultipleLines'], axis=1)
```

- kolom `**Customer_id**` dihapus karena tidak memiliki efek pada indikasi churn,
- kolom `**gender** dihapus karena nilai churn untuk laki-laki dan perempuan hampir sama.
- kolom `**PhoneService** dan `**Multiple Line**` dihapus karena tidak memberikan informasi apa pun tentang churn.

b. Handling Typo Error Data Types

```
df_prep['TotalCharges'] = pd.to_numeric(df_prep['TotalCharges'], errors='coerce')
```

c. Handling Missing Value

```
df_prep.TotalCharges.isna().sum()
```

```
11
```

Data Prepocessing

Data Cleaning

d. Drop Missing Value

```
df_prep.dropna(inplace=True)
```

e. Handling Duplicate Value

```
df_prep.duplicated().sum()
```

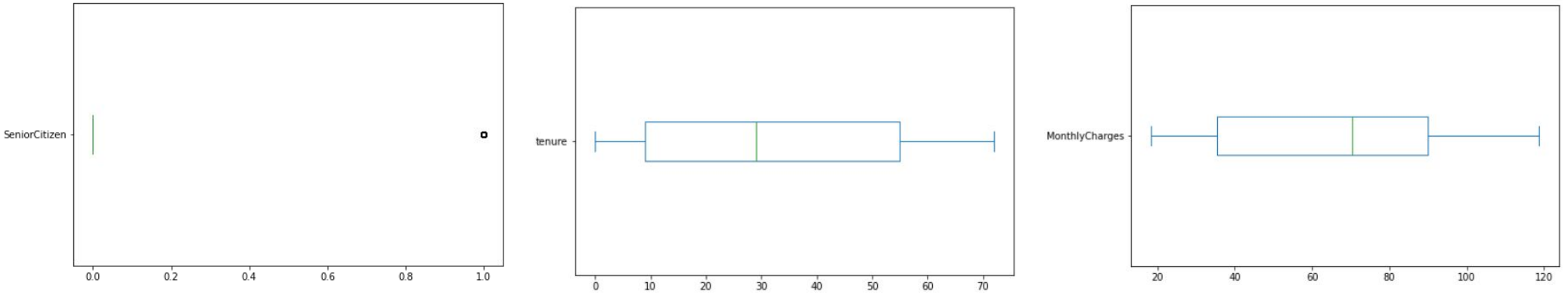
```
41
```

hal tersebut tidak di hapus karena ….

tenure	InternetService	OnlineSecurity	OnlineBackup	MonthlyCharges	TotalCharges	Churn
1	DSL	No	No	45.70	45.70	Yes
1	No	No internet service	No internet service	20.15	20.15	Yes
1	No	No internet service	No internet service	19.55	19.55	No
1	Fiber optic	No	No	69.90	69.90	Yes
1	No	No internet service	No internet service	20.20	20.20	No
1	No	No internet service	No internet service	19.60	19.60	Yes
1	No	No internet service	No internet service	20.45	20.45	No

Data Prepocessing

e. Outlier



Observation : `SeniorCitizen` adalah sebuah kolom biner sehingga 1142 instansi bukanlah data outlier, jadi tidak ada outlier dalam dataset ini

Data Prepocessing

Feature Engineering

a. Feature Extraction

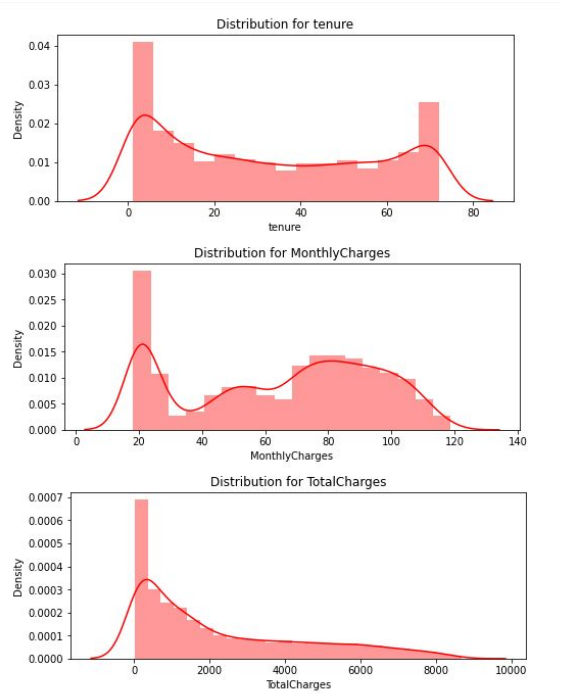
```
[32] # get correlation > +- 0.5
     df_corr = df_clean.corr().iloc[1:,1:]
     df_corr = df_corr.apply(lambda x: round(x, 3))
     dfcorr = []
     for idx in df_corr.index:
         for col in df_corr.columns:
             dfcorr.append([idx, col])
             if (np.abs(df_corr.loc[idx, col]) > 0.5) and (idx != col) and [col, idx] not in dfcorr:
                 print(f'{idx} with {col} has correlation : {df_corr.loc[idx, col]}')
```

```
tenure with TotalCharges has correlation : 0.826
MonthlyCharges with TotalCharges has correlation : 0.651
```

Data Prepocessing

Feature Engineering

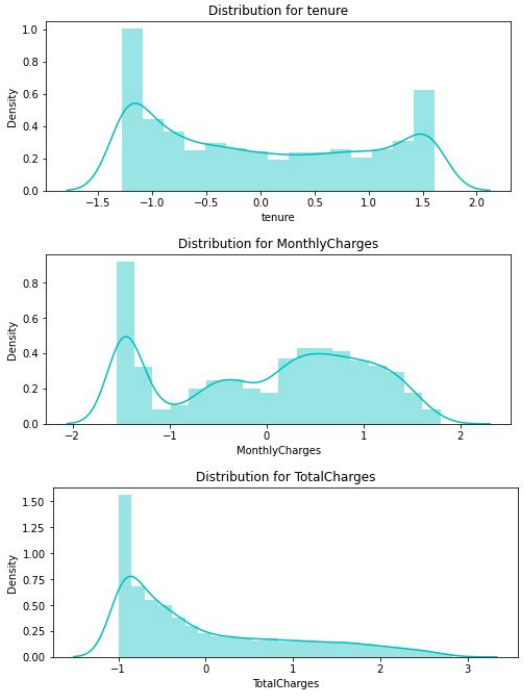
b. Feature Scaling



Berdasarkan hasil visualiasi tersebut, maka:

- Kolom **`tenure`** akan dilakukan Standard scaling karena memiliki distribusi right skewd
- Kolom **`MonthlyCharges`** dan **`TotalCharges`** akan dilakukan standard scaling untuk memastikan bahwa kedua kelompok memiliki skala yang sama.

Hasil Feature Scaling



Data Prepocessing

c. Feature Encoding

```
def object_to_int(dataframe_series):
    if dataframe_series.dtype=='object':
        dataframe_series = LabelEncoder().fit_transform(dataframe_series)
    return dataframe_series
```

```
df_clean = df_clean.apply(lambda x: object_to_int(x))
df_clean.head()
```

df.head()

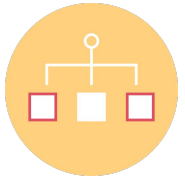
	SeniorCitizen	Partner	Dependents	tenure	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies
0	0	1	0	1	0	0	2	0	0	0	0
1	0	0	0	34	0	2	0	2	0	0	0
2	0	0	0	2	0	2	2	0	0	0	0
3	0	0	0	45	0	2	0	2	2	0	0
4	0	0	0	2	1	0	0	0	0	0	0

Objectives

Business Understanding

Data Understanding

Data Preprocessing



Modelling

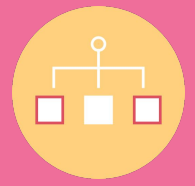
Evaluation

Objectives

Business Understanding

Data Understanding

Data Preprocessing



Modelling

Evaluation

Modelling

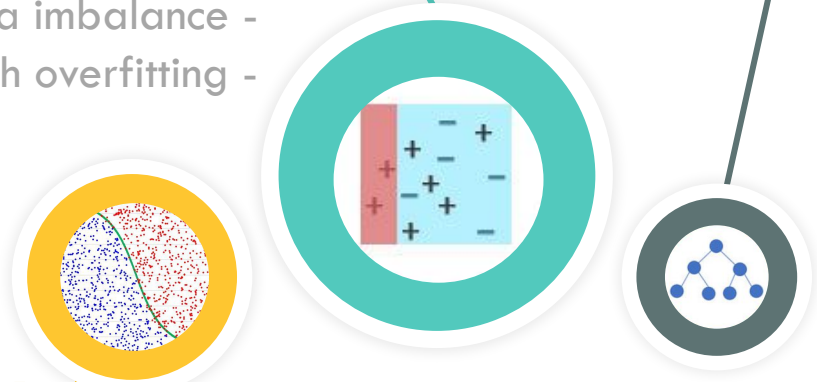
a. Algoritma

01
Logistic Regression

prediksi bertipe binary -
data terbatas, tidak ada time series -
memanfaatkan teknik regresi -

02
ADABoost

class biner -
data imbalance -
tidak mudah overfitting -



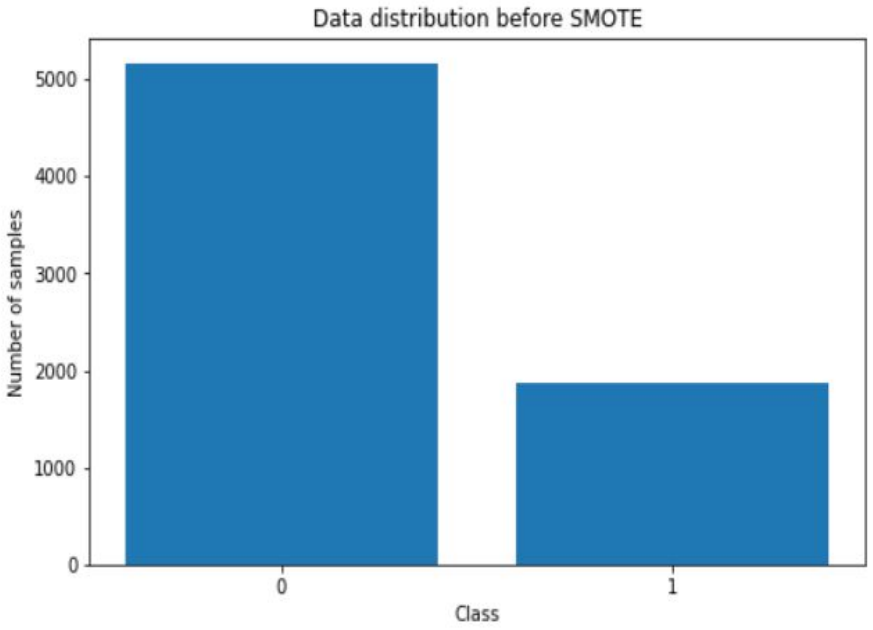
03
XGBoost

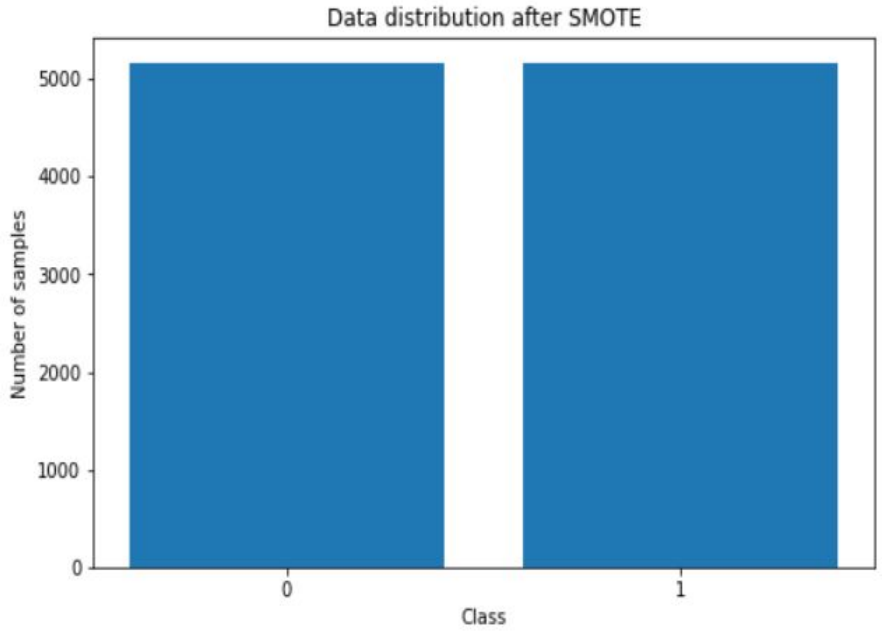
- teknik reguralisasi
- data imbalance

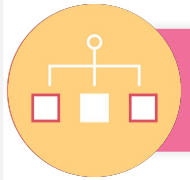
Modelling

b. Imbalance

handling imbalance data menggunkaan SMOTE.





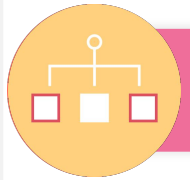


Modelling

b. Split Data (Train & Test)

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.20, random_state = 1)
print("Shape of x_train: ", X_train.shape)
print("Shape of x_test: ", X_test.shape)
```

```
Shape of x_train:  (8260, 18)
Shape of x_test:  (2066, 18)
```



Modelling

b. Modelling Process

Menggunakan 2 proses modelling, dimana pada **modeling #1**

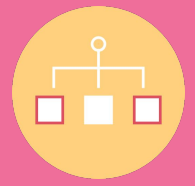
1. Model selection
2. Hyperparameter Tunning : Random Search

Objectives

Business Understanding

Data Understanding

Data Preprocessing



Modelling

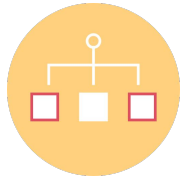
Evaluation

Objectives

Business Understanding

Data Understanding

Data Preprocessing



Modelling

Evaluation

Evaluation

Purpose

memprediksi apakah pelanggan akan churn apa tidak, maka:

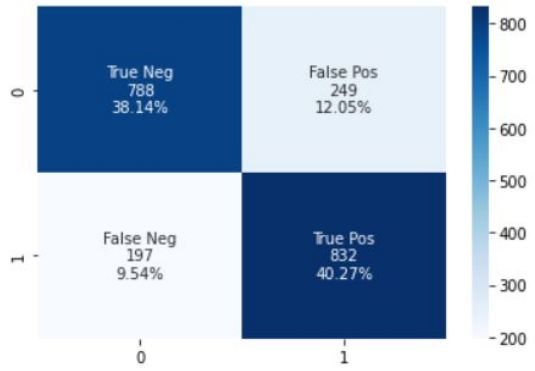
- Metric evaluasi yang akan kita fokuskan adalah **F1 Score** dimana kita ingin mempertimbangkan Precision dan Recall.
- Karena data target pada kasus kita terjadi imbalance class sehingga kita juga akan mempertimbangkan nilai **AUC** sebagai pembeda antar kelas negative dan positive.
- **Cross Validation Score** digunakan untuk mengevaluasi kinerja model machine learning, khususnya mendeteksi saat terjadinya data leakage.

Evaluation

a. Data Modelling #1 - Default

Logistic Regression

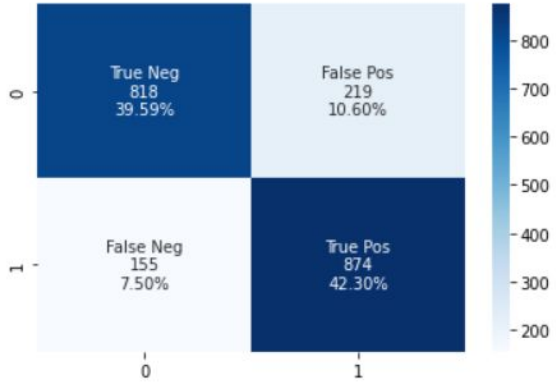
	precision	recall	f1-score	support
0	0.80	0.76	0.78	1037
1	0.77	0.81	0.79	1029
accuracy			0.78	2066
macro avg	0.78	0.78	0.78	2066
weighted avg	0.78	0.78	0.78	2066



Cross Validation Score : 87.37%

AdaBoost

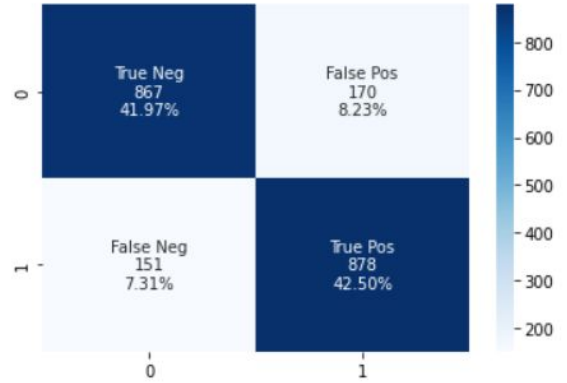
	precision	recall	f1-score	support
0	0.84	0.79	0.81	1037
1	0.80	0.85	0.82	1029
accuracy			0.82	2066
macro avg	0.82	0.82	0.82	2066
weighted avg	0.82	0.82	0.82	2066



Cross Validation Score : 87.37%

XGBoost

	precision	recall	f1-score	support
0	0.85	0.84	0.84	1037
1	0.84	0.85	0.85	1029
accuracy			0.84	2066
macro avg	0.84	0.84	0.84	2066
weighted avg	0.84	0.84	0.84	2066



Cross Validation Score : 92.45%

Evaluation

a. Data Modelling #1 - Default

Scoring base model

	Model	Accuracy	Precision	Recall	F1 Score	AUC (Test)	AUC (Train)	Cross Validation Score
0	Logistic Regression	0.7841	0.7697	0.8086	0.7886	0.8660	0.8776	0.873000
1	AdaBoost	0.8190	0.7996	0.8494	0.8238	0.8986	0.9084	0.903227
2	XGBoost	0.8446	0.8378	0.8533	0.8455	0.9250	0.9893	0.925448

Evaluation

b. Hyperparameter Tunning : Random Search

```
list_hyperparameters = [
    { # Logistic Regression
        'penalty': ['l2','l1','elasticnet'],
        'C': [float(x) for x in np.logspace(-3, 3, 20)]
    },
    { # Adaboost
       'n_estimators' : [int(x) for x in np.linspace(100, 2000, 1000)],
       'learning_rate' : [float(x) for x in np.linspace(0.001, 0.1, 100)],
       'algorithm' : ['SAMME', 'SAMME.R']
    },
    { # XGBoost
        'max_depth' : [int(x) for x in np.linspace(10, 100, 10)],
        'min_child_weight' : [int(x) for x in np.linspace(1, 10, 11)],
        'gamma' : [float(x) for x in np.linspace(0, 1, 11)],
        'tree_method' : ['auto', 'exact', 'approx', 'hist'],
        'colsample_bytree' : [float(x) for x in np.linspace(0, 1, 11)],
        'learning_rate' : [float(x) for x in np.linspace(0, 1, 100)],
        'reg_lambda' : [float(x) for x in np.linspace(0, 1, 11)],
        'reg_alpha' : [float(x) for x in np.linspace(0, 1, 11)]
    }
]

def show_best_hyperparameter(model, hyperparameters):
    for key, value in hyperparameters.items() :
        print('Best '+key+':', model.best_estimator_.get_params()[key])
```

Evaluation

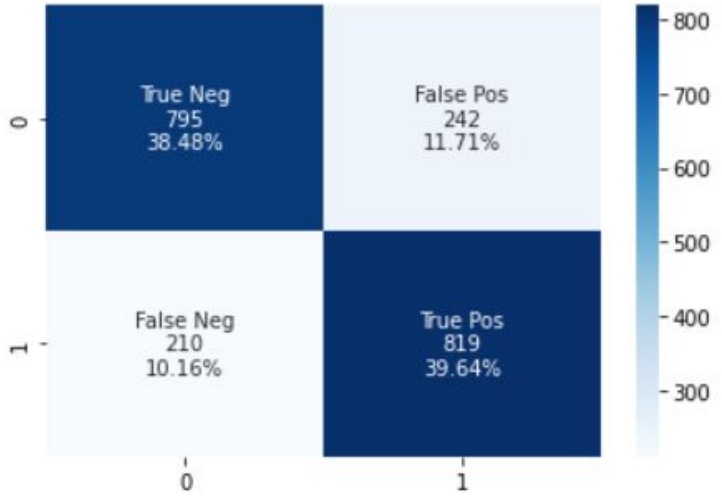
b. Hyperparameter Tuning#1 (Random Search)

#Logistic Regression

```
MODEL LOGISTIC REGRESSION AFTER HYPERPARAMETER TUNING
======================================================
Best penalty: l2
Best C: 6.158482110660261
======================================================

Cross Validation Score :  87.40%
```

	precision	recall	f1-score	support
0	0.79	0.77	0.78	1037
1	0.77	0.80	0.78	1029
accuracy			0.78	2066
macro avg	0.78	0.78	0.78	2066
weighted avg	0.78	0.78	0.78	2066



Evaluation

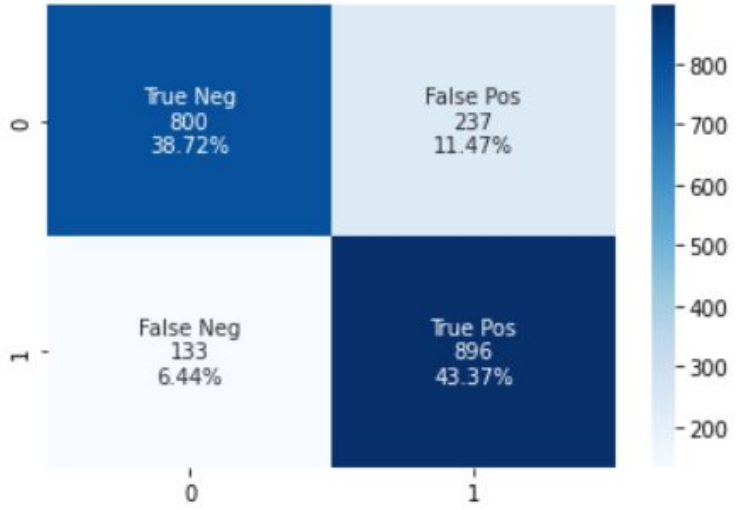
b. Hyperparameter Tuning#1 (Random Search)

#AdaBoost

```
MODEL AdaBoost AFTER HYPERPARAMETER TUNING
==================================================
Best n_estimators: 1748
Best learning_rate: 0.047
Best algorithm: SAMME.R
==================================================

Cross Validation Score :  90.58%
```

	precision	recall	f1-score	support
0	0.86	0.77	0.81	1037
1	0.79	0.87	0.83	1029
accuracy			0.82	2066
macro avg	0.82	0.82	0.82	2066
weighted avg	0.82	0.82	0.82	2066



Evaluation

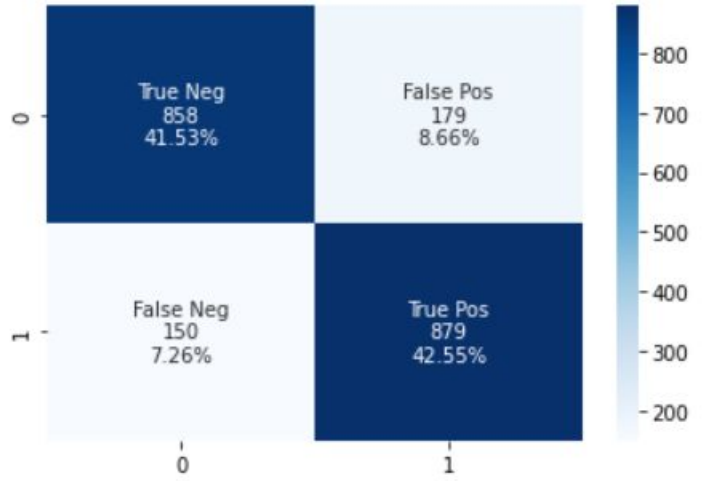
b. Hyperparameter Tuning#1 (Random Search)

#XGBoost

```
MODEL XGBOOST AFTER HYPERPARAMETER TUNING
=========================================
Best max_depth: 10
Best min_child_weight: 8
Best gamma: 0.1
Best tree_method: auto
Best colsample_bytree: 0.7000000000000001
Best learning_rate: 0.13131313131313133
Best reg_lambda: 0.5
Best reg_alpha: 0.6000000000000001
=========================================
```

```
Cross Validation Score :  92.32%
```

	precision	recall	f1-score	support
0	0.85	0.83	0.84	1037
1	0.83	0.85	0.84	1029
accuracy			0.84	2066
macro avg	0.84	0.84	0.84	2066
weighted avg	0.84	0.84	0.84	2066



Evaluation

Comparation

#base model

	Model	Accuracy	Precision	Recall	F1 Score	AUC (Test)	AUC (Train)	Cross Validation Score
0	Logistic Regression	0.7841	0.7697	0.8086	0.7886	0.8660	0.8776	0.873000
1	AdaBoost	0.8190	0.7996	0.8494	0.8238	0.8986	0.9084	0.903227
2	XGBoost	0.8446	0.8378	0.8533	0.8455	0.9250	0.9893	0.925448

#tunned model

	Model	Accuracy	Precision	Recall	F1 Score	AUC (Test)	AUC (Train)	Cross Validation Score
0	Logistic Regression	0.7812	0.7719	0.7959	0.7837	0.8629	0.8756	0.874256
1	AdaBoost	0.8209	0.7908	0.8707	0.8289	0.9003	0.9102	0.905594
2	XGBoost	0.8408	0.8308	0.8542	0.8424	0.9252	0.9769	0.921154

Modelling Conclution

Berdasarkan hasil prediksi dan analisis yang telah kami lakukan maka dapat disimpulkan bahwa:

Model yang direkomendasikan adalah XGBoost memiliki **F1 Score sebesar 84%** artinya dapat memprediksi pelanggan yang akan melakukan churn **(Precision) 83%** dengan ketepatan memprediksi pelanggan benar akan churn **(Recall) 85%.**

Recommendation

Terdapat beberapa rekomendasi yang dapat diberikan untuk meningkatkan penghasilan dan mengurangi customer churn.

- Perusahaan perlu memfokuskan layanan pada tiga jenis pelanggan, yaitu senior citizen, pelanggan yang tinggal dengan pasangan(partner), dan pelanggan yang tinggal sendirian. Senior citizen rela membayar harga yang lebih tinggi untuk layanan yang berkualitas, sedangkan pelanggan yang tinggal dengan pasangan dan sendirian lebih memilih layanan dengan biaya bulanan di bawah 65.
- Perusahaan perlu fokus pada layanan seperti OnlineSecurity, OnlineBackup, DeviceProtection, dan TechSupport selama enam bulan pertama karena ini adalah periode yang paling kritis dan tidak pasti bagi pelanggan.
- Perusahaan perlu membuat layanan StreamingTV dan StreamingMovies terjangkau dan memperhatikan konten yang menarik untuk semua jenis pelanggan.
- Perusahaan juga perlu memperhatikan metode pembayaran yang mudah dan lancar, seperti Bank Transfer (automatic) dan Credit Card (automatic). Perusahaan harus menghindari Electronic check karena memiliki tingkat churn yang tinggi.

Conclusion

Berdasarkan hasil prediksi dan analisis yang telah kami lakukan maka dapat disimpulkan bahwa:

1. Model yang direkomendasikan adalah XGBoost memiliki F1 Score sebesar 82.39% artinya dapat memprediksi pelanggan yang akan melakukan churn (Precision) 88.51% dengan ketepatan memprediksi pelanggan benar akan churn (Recall) sebanyak 77.06%.
2. Peningkatan layanan yang sudah dijelaskan pada rekomendasi. Dalam hal ini, CRM dapat membantu perusahaan mengidentifikasi pelanggan yang membutuhkan layanan tertentu dan memberikan layanan yang tepat pada waktu yang tepat. Dengan demikian, CRM dapat membantu perusahaan meningkatkan loyalitas pelanggan dan mengurangi churn.


BCC
Basic Computing Community
Mission
Complete
yeayy, yuhuu 🥳


thank you
danke
謝謝
ngiyabonga
teşekkür ederim
gracias
merci
obrigado
спасибо
bedankt
dziękuję
sukriya
kop khun krap
go raibh maith agat
mochchakkeram
tapadh leat
arigatō
dakujem
grazie
terima kasih
takk
taiku
sagolun
hvala
mauruuru
dank je
мерси
ευχαριστώ
감사합니다
ありがとう
Баярлалаа
рахмат
хвала
tack
mersi
barka
blagodaram
misaotra
matondo
paldies
asante
manana
enkosi
sobodi
dekuji
mèsi
najis tuke
rahmet
tanemirt
diolch
dhanyavadagalu
shukriya
trugarez
xiexie
sulpay
grácies
chnorakaloutioun
djiere dieuf
mamnun

DOKUMEN BRIDGING

https://docs.google.com/document/d/13mMutfONK5XNhuLgo9esA2BOo6ziKyhXyD_zTiIAxKM/edit#heading=h.e10kljp6y29b