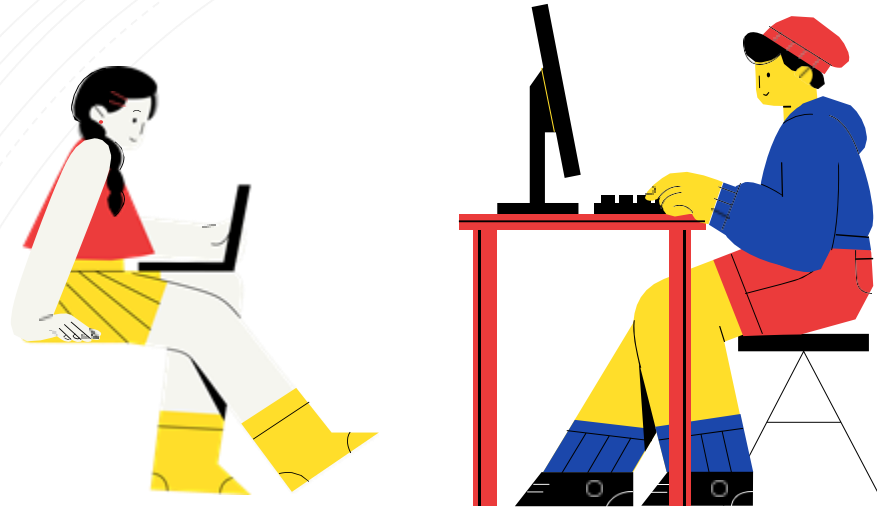




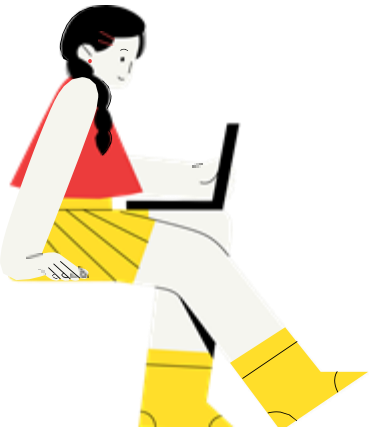
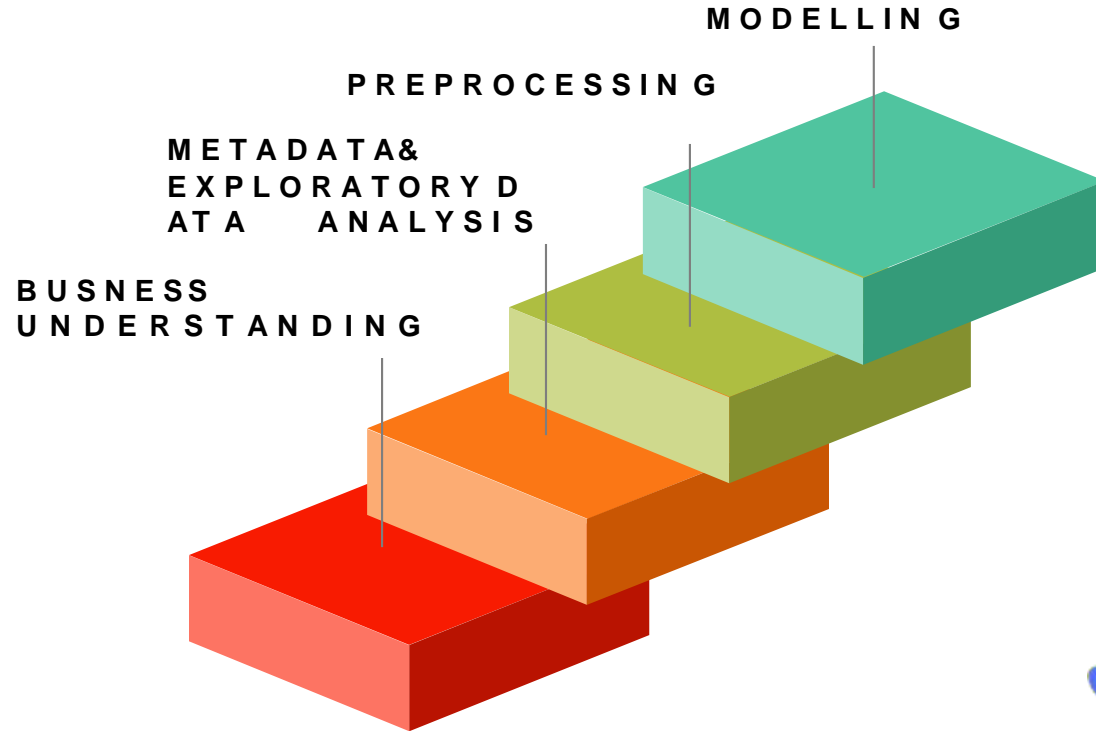
# HOME CREDIT DEFAULT RISK

1. Iganasius Frans De Sale Tyas Neno
2. Hendy Nurhidayat
3. Ira Rizkillah Koswara



## Latar Belakang

# Business Step



# Profile Perusahaan

# HOME CREDIT

PT Home Credit Indonesia atau yang lebih dikenal dengan Home Credit merupakan perusahaan pembiayaan multiguna multinasional. Perusahaan ini menyediakan layanan pembiayaan berbasis teknologi.

## Pembiayaan

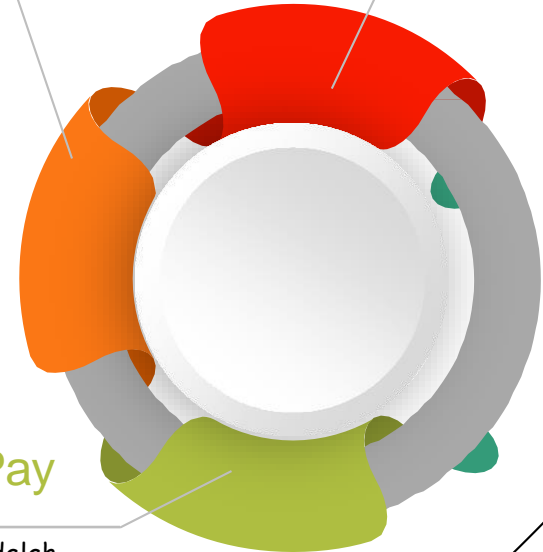
Pembiayaan bagi pelanggan yang berbelanja secara online maupun offline

Menyediakan pembiayaan untuk financial pelanggan

## Peminjaman

## Home Credit Pay

Home Credit Pay adalah layanan e-money atau uang elektronik yang bekerja sama dengan KasPro



# Business Understanding



Banyak orang berjuang untuk mendapatkan pinjaman karena sejarah kredit yang tidak mencukupi atau tidak ada. Dan, sayangnya, populasi ini sering dimanfaatkan oleh pemberi pinjaman yang tidak dapat dipercaya.



Home Credit berupaya untuk memperluas inklusi keuangan bagi populasi yang tidak memiliki rekening bank dengan memberikan pengalaman meminjam yang positif dan aman. Home Credit menggunakan berbagai data alternatif - termasuk telekomunikasi dan informasi transaksional - untuk memprediksi kemampuan pembayaran kembali klien mereka.



# Business Understanding



## TARGET

Menemukan kelayakan user yang tidak memiliki rekening bank dalam melakukan pengajuan kredit berdasarkan parameter-parameter tertentu yang telah tersedia dari fitur-fitur aplikasi pengajuan kredit yang telah dilakukan sebelumnya oleh user tersebut

# Business Understanding

## Strategi Bisnis



01 Memaksimalkan Metode Statistika



02 Memaksimalkan Metode Machine Learning



# METADATA





# METADATA



```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 307511 entries, 0 to 307510  
Columns: 122 entries, SK_ID_CURR to AMT_REQ_CREDIT_BUREAU_YEAR  
dtypes: float64(65), int64(41), object(16)  
memory usage: 286.2+ MB
```

Kita mempunyai 122 kolom dengan 307511 baris data #fact1



# Exploratory Data Analysis



# EDA

## Dataset statistics

Number of variables	122
Number of observations	307511
Missing cells	9152465
Missing cells (%)	24.4%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	286.2 MiB
Average record size in memory	976.0 B

## Variable types

Numeric	70
Categorical	49
Boolean	3



# EDA

Dalam data yang telah di eksplor di temukan missing value yang bernilai 24.4% dan dihitung berdasarkan cells yang memiliki nilai missingvalue

## Dataset statistics

Number of variables	122
Number of observations	307511
Missing cells	9152465
Missing cells (%)	24.4%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	286.2 MiB
Average record size in memory	976.0 B



## Variable types

Numeric

Categorical

Boolean

70

49

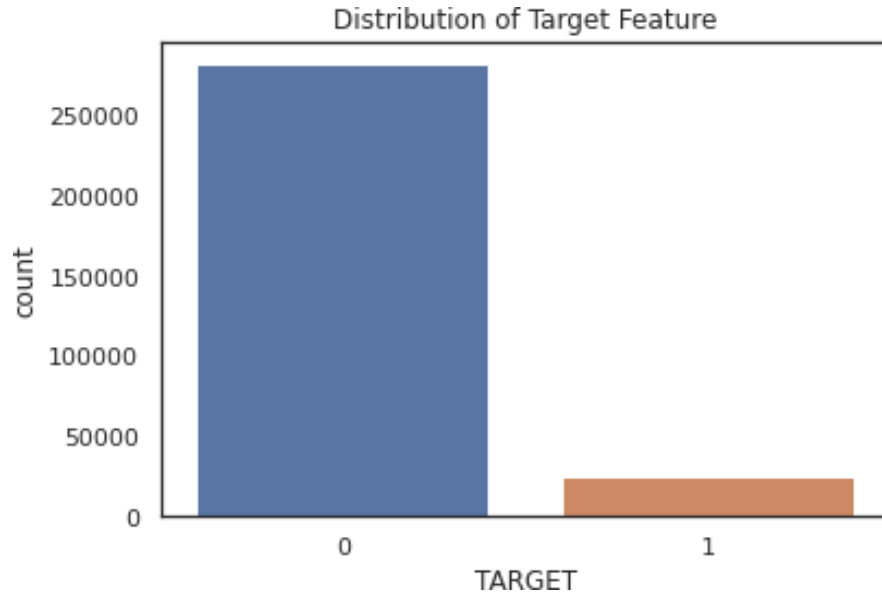
3

Dari 122 Variable yang kita miliki kita mendapat 3 kategori tipe data yaitu Numeric, Categorical, dan Boolean.

Jumlah masing-masing variable berdasarkan kategori tipe data yang ada.



TARGET  
Categorical

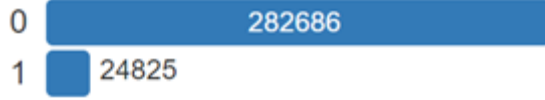


Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	2.3 MiB



## TARGET

### Categorical



Pada kolom target bisa di katakan terjadi imbalance sehingga perlu di lakukan sampling pada kolom ini untuk meningkatkan akurasi pada model yang akan kita buat.



# Pre-Processing





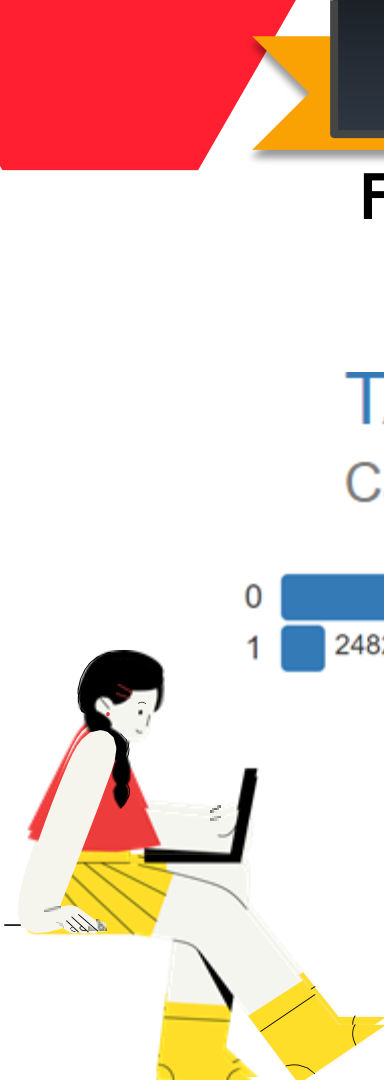
# Pre-processing

## Feature Selection

TARGET  
Categorical



Pada kolom target bisa di katakan terjadi imbalance sehingga perlu di lakukan sampling pada kolom ini untuk meningkatkan akurasi pada model yang akan kita buat.

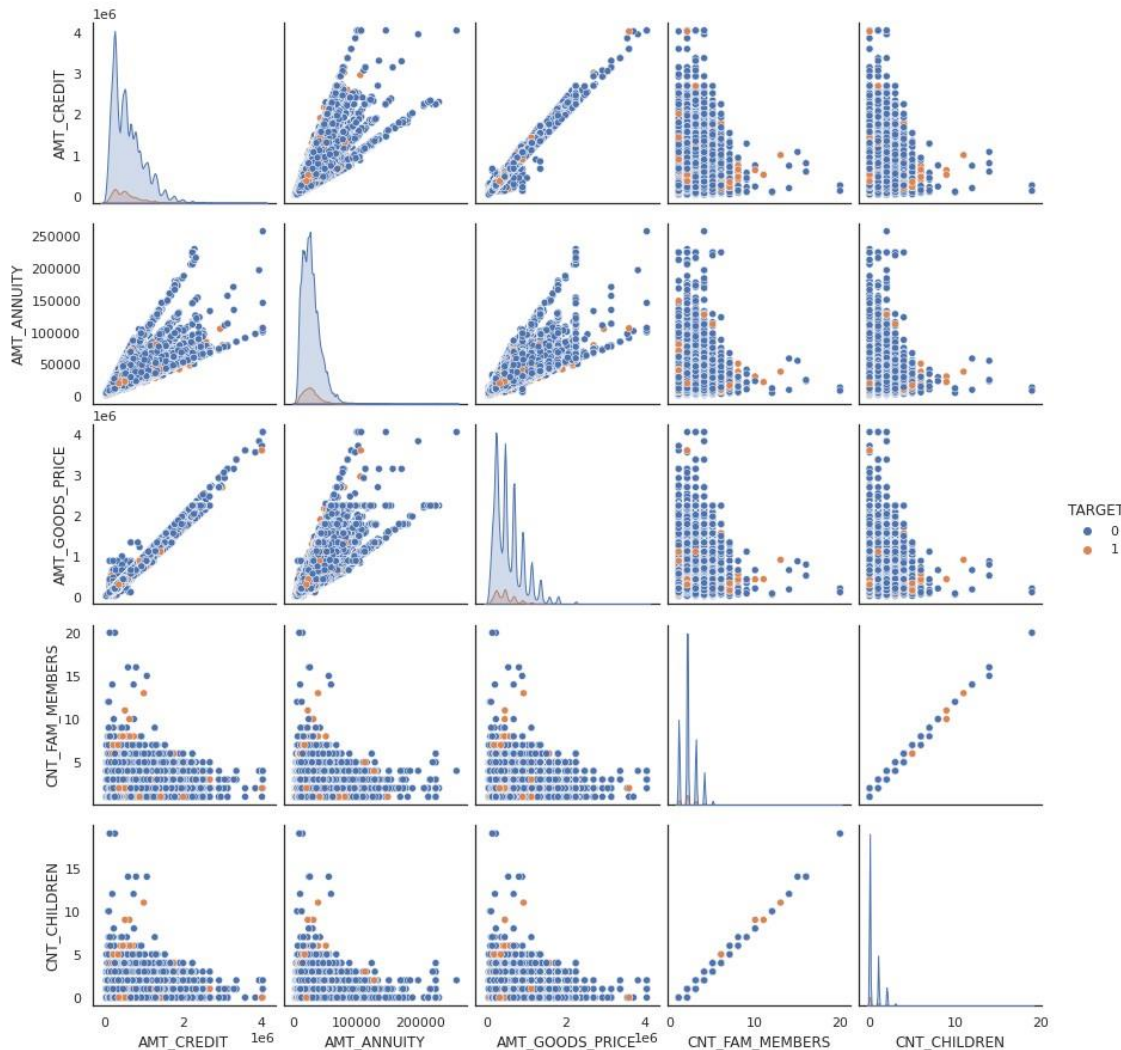


# Pre-processing

## Feature Correlation

Fitur yang digunakan karena memiliki korelasi positif yaitu:

- AMOUNT CREDIT
- AMOUNT ANNUITY
- AMOUNT GOODS PRICE
- CNT\_FAM\_MEMBERS
- CNT\_CHILDREN



# Pre-processing

## Feature Selection

Dalam korelasi fitur yang terlihat kita akan menggunakan beberapa fitur saja yang memungkinkan untuk pemodelan.

01

Target

02

Amount Credit

03

Amount Annuity

04

Amount Goods  
Price

04

Count Family  
Member

04

Count Children

# Pre-processing

## Skewness & Missing Value

Jumlah missing value pada setiap variabel

AMT_GOODS_PRICE	278
AMT_ANNUITY	12
CNT_FAM_MEMBERS	2
CNT_CHILDREN	0
AMT_CREDIT	0
TARGET	0

dtype: int64

Nilai Skewnes dari setiap kategori

AMT_CREDIT	1.234778
AMT_ANNUITY	1.579777
AMT_GOODS_PRICE	1.349000
CNT_FAM_MEMBERS	0.987543
CNT_CHILDREN	1.974604

dtype: float64

Berdasarkan literasi statistik maka pada feature yang telah di pilih, bisa disimpulkan semua missing value akan diisi dengan nilai mean(rata- rata) hal ini di karenakan nilai skewness yang di dapat masuk dalam kategori normal.



Sudah tidak ada nilai missing value

SK_ID_CURR	0
TARGET	0
CNT_CHILDREN	0
AMT_CREDIT	0
AMT_ANNUITY	0
AMT_GOODS_PRICE	0
CNT_FAM_MEMBERS	0

dtype: int64

# Modelling



# Modelling

## Dummy Data

```
dummies_df.head()
```

	SK_ID_CURR	TARGET	CNT_CHILDREN	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	CNT_FAM_MEMBERS
0	100002	1	0	408597.5	24700.5	351000.0	1.0
1	100003	0	0	1293502.5	35698.5	1129500.0	2.0
2	100004	0	0	135000.0	6750.0	135000.0	1.0
3	100006	0	0	312682.5	29686.5	297000.0	2.0
4	100007	0	0	513000.0	21865.5	513000.0	1.0

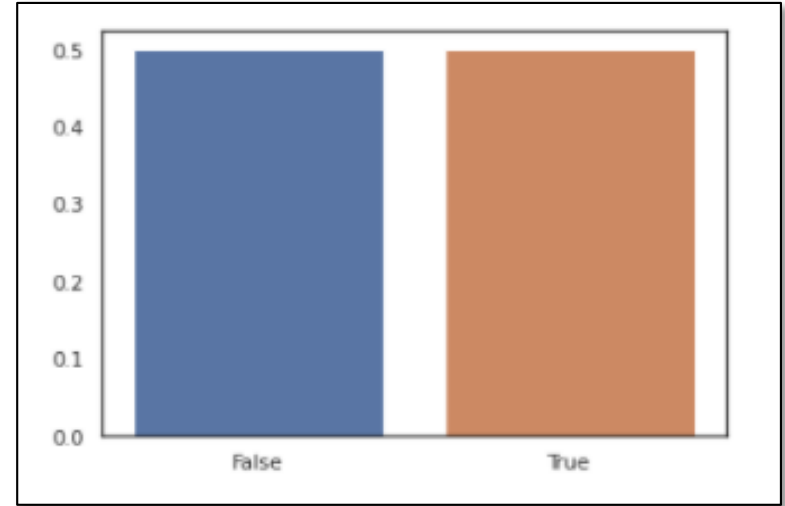
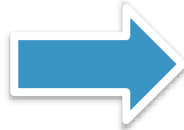
## Split Data

```
X = dummies_df.drop(['TARGET', 'SK_ID_CURR'], axis = 1)  
y = dummies_df['TARGET']
```

```
from sklearn.model_selection import train_test_split  
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.3, stratify = y, random_state = 123)
```

# Modelling

## Imbalance Handling



Undersampling

# Modelling

## MinMax Scaler

```
from sklearn.preprocessing import MinMaxScaler  
scaler = MinMaxScaler()  
X_transform = scaler.fit_transform(X_under)  
X_transform = pd.DataFrame(X_transform, columns = X.columns)
```

X\_transform

	CNT_CHILDREN	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	CNT_FAM_MEMBERS
0	0.090909	0.191743	0.108091	0.177215	0.166667
1	0.000000	0.061864	0.128309	0.063291	0.083333
2	0.090909	0.158185	0.164871	0.177215	0.166667
3	0.090909	0.045196	0.069774	0.050633	0.083333
4	0.090909	0.205279	0.104782	0.189873	0.166667
...	...	...	...	...	...
34749	0.000000	0.146886	0.214891	0.164557	0.083333
34750	0.000000	0.047600	0.099475	0.044304	0.000000
34751	0.090909	0.102143	0.151755	0.091139	0.166667
34752	0.090909	0.045485	0.044794	0.032911	0.083333
34753	0.090909	0.056382	0.073467	0.050633	0.166667

34754 rows × 5 columns





# Modelling

## Accuracy

```
def evaluasi_model(model,X_test,y_test):  
    from sklearn.metrics import accuracy_score  
    y_pred = model.predict(X_test)  
    return accuracy_score(y_test,y_pred)*100
```

## KNN

```
from sklearn.neighbors import KNeighborsClassifier  
knn = KNeighborsClassifier()  
knn.fit(X_transform,y_under)  
accuracy = evaluasi_model(knn,X_test,y_test)  
print('accuracy : {0:.5f}%'.format(accuracy))  
  
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:3: DataConversionWarning:  
A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using  
ravel().  
  
accuracy : 91.92230%
```

## Decission Tree

```
from sklearn.tree import DecisionTreeClassifier  
model_tree = DecisionTreeClassifier()  
model_tree.fit(X_transform, y_under)  
accuracy = evaluasi_model(model_tree,X_test,y_test)  
print('accuracy : {0:.5f}%'.format(accuracy))  
  
accuracy : 91.92664%
```

KNN = 91.92230%  
Decission Tree = 91.92664%



# Conclusion



# Conclusion

Feature yang memiliki nilai korelasi yang tinggi dapat membantu menilai kelayakan customer yang tidak memiliki rekening dalam melakukan pengajuan kredit.

Feature-feature yang dimaksud antara lain:

- ☐ Count Children
- ☐ Amount Credit
- ☐ Amount Annuity
- ☐ Amount Goods Price
- ☐ Count Family Members



TERIMA KASIH

