



# Classification Loan With Home Credit Default Risk

Kelompok 12

## Anggota Kelompok 12 :

1

Adinda Permata Sari

2

Ardia Fatma Sari

3

Himma Faicha Hubbiya

4

Raihanah Assa`Adah

5

Raihan Tsabita Sabil



# LINK GCOLAB DAN DASHBOARD

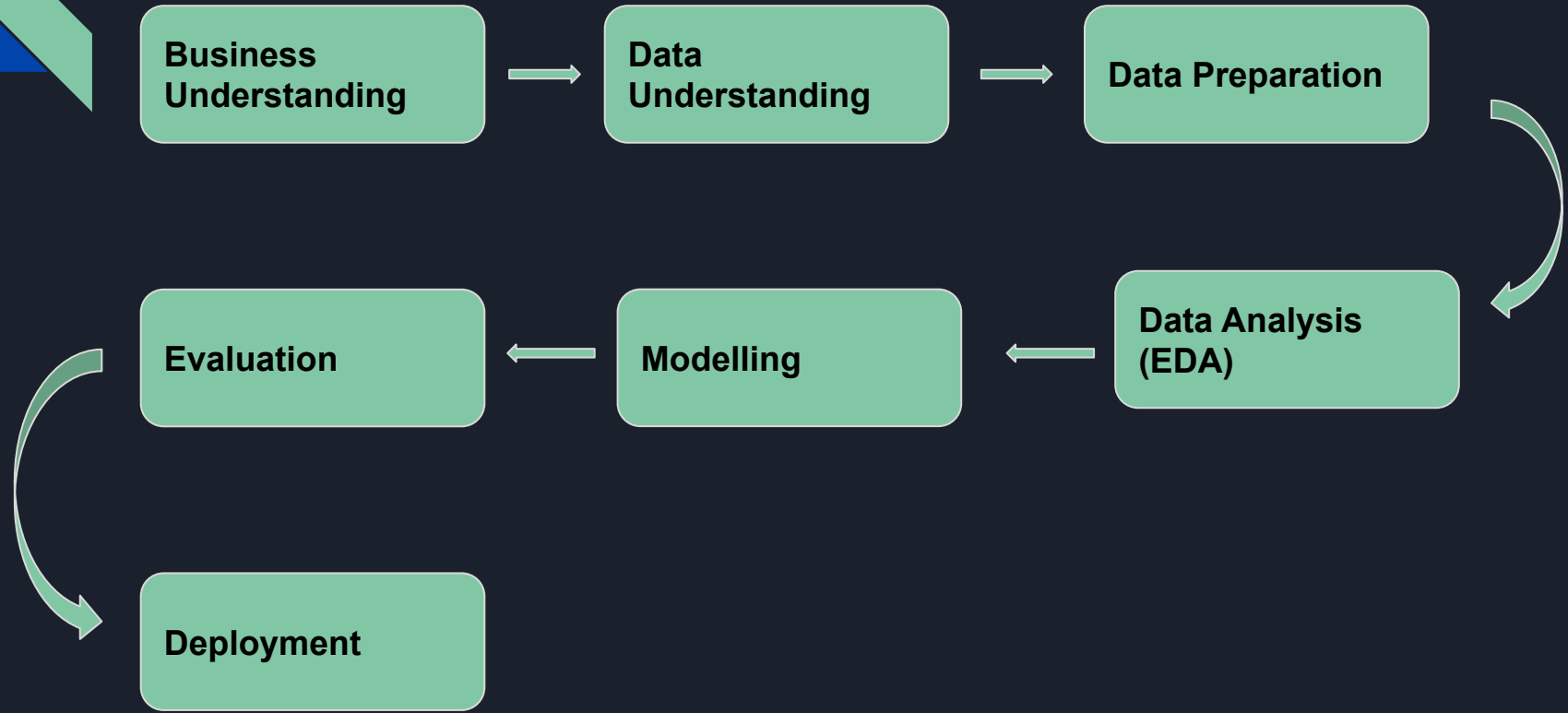
## Google Colab:

<https://colab.research.google.com/drive/1tSGqgEmhr273qubDM-782V7999mEspPv#scrollTo=QrohCdJo47Jb>

## Dashboard:

<https://lookerstudio.google.com/reporting/0b83642c-2d7b-4109-bde4-31c6b71dbf02>

# Alur Pembahasan



A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is a light greenish-blue. Both are tilted at an angle.

# **Business Understanding**



# Bussiness Understanding

## Latar Belakang:

Dataset yang disediakan di Kaggle terkait dengan Home Credit, sebuah lembaga keuangan yang mengkhususkan diri dalam layanan keuangan non-bank. Mereka fokus pada memberikan pinjaman kepada individu yang mungkin memiliki riwayat kredit terbatas atau tidak ada. Tujuannya adalah memprediksi kemungkinan seorang klien gagal membayar pinjamannya, berdasarkan berbagai fitur dan atribut yang tersedia.

## Tujuan:

- Dapat memprediksi risiko kredit sehingga bank/perusahaan/institusi keuangan dapat memberikan pinjaman kepada pelanggan dengan tepat sasaran sehingga dapat mengantisipasi adanya gagal bayar.
- Meningkatkan akurasi dan efisiensi dalam menilai risiko kredit
- Perusahaan dapat meminimalkan kerugian dan memaksimalkan keuntungan



## Permasalahan

Perusahaan selektif dalam memberikan pinjaman ke calon nasabah, namun sebagian calon tersebut pengajuannya tidak diterima. Bagaimana cara memprediksi kelayakan nasabah dalam pemberian pinjaman berdasarkan data-data pelanggan terdahulu?

## Solusi

Digunakan pemodelan dalam machine learning dengan memanfaatkan data-data pelanggan terdahulu untuk memprediksi kelayakan nasabah baru dalam pemberian pinjaman.

A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is light green, both with a slight 3D effect.

# **Data Understanding**





# Data Understanding

Dataset utama yang digunakan:

**application\_train:** File ini berisi data yang akan digunakan untuk melatih model prediktif. Setiap baris dalam file ini mewakili satu peminjam dan berisi berbagai fitur (variabel) yang digunakan untuk memprediksi risiko gagal bayar. Beberapa fitur yang dapat ditemukan dalam file ini antara lain adalah informasi pribadi peminjam (misalnya jenis kelamin, usia), informasi kredit (misalnya jumlah pinjaman, jangka waktu), dan informasi historis kredit (misalnya jumlah tunggakan, jumlah kredit yang telah dilunasi).

**Tujuan utama** dari analisis data ini adalah untuk memahami karakteristik dataset, mengeksplorasi hubungan antara variabel-variabel yang ada, dan mempersiapkan data untuk proses pemodelan. Beberapa tugas yang perlu dilakukan dalam tahap pemahaman data (data understanding) antara lain:

- **Menjelajahi Data:** Membaca dan memahami struktur data dalam file train. Melihat jumlah baris dan kolom dalam dataset, serta melihat beberapa contoh data untuk mendapatkan gambaran awal tentang informasi yang ada.
- **Variabel Target:** Identifikasi variabel target yang akan diprediksi. Dalam kasus ini, variabel target adalah "TARGET" yang menunjukkan apakah peminjam mengalami gagal bayar (1) atau tidak (0).

A blue parallelogram and a light green parallelogram are positioned in the top-left corner of the slide. The background is dark blue with diagonal stripes in a slightly lighter shade of blue.

# **Data Preparation**

# Cek Missing value

```
COMMONAREA_MEDI      214865
COMMONAREA_AVG       214865
COMMONAREA_MODE       214865
NONLIVINGAPARTMENTS_MODE 213514
NONLIVINGAPARTMENTS_AVG 213514
...
NAME_HOUSING_TYPE      0
NAME_FAMILY_STATUS     0
NAME_EDUCATION_TYPE    0
NAME_INCOME_TYPE       0
SK_ID_CURR             0
Length: 122, dtype: int64
```

Mengecek apakah ada missing value, kemudian melakukan drop data yang missing valuenya >50%.

	Percent Miss	Total Miss
COMMONAREA_MEDI	69.872297	214865
COMMONAREA_AVG	69.872297	214865
COMMONAREA_MODE	69.872297	214865
NONLIVINGAPARTMENTS_MODE	69.432963	213514
NONLIVINGAPARTMENTS_AVG	69.432963	213514
NONLIVINGAPARTMENTS_MEDI	69.432963	213514
FONDKAPREMONT_MODE	68.386172	210295
LIVINGAPARTMENTS_MODE	68.354953	210199
LIVINGAPARTMENTS_AVG	68.354953	210199
LIVINGAPARTMENTS_MEDI	68.354953	210199
FLOORSMIN_AVG	67.848630	208642
FLOORSMIN_MODE	67.848630	208642
FLOORSMIN_MEDI	67.848630	208642

# Handling Missing Value

	Percent Miss	Total Miss
COMMONAREA_MEDI	69.872297	214865
COMMONAREA_AVG	69.872297	214865
COMMONAREA_MODE	69.872297	214865
NONLIVINGAPARTMENTS_MODE	69.432963	213514
NONLIVINGAPARTMENTS_AVG	69.432963	213514
NONLIVINGAPARTMENTS_MEDI	69.432963	213514
FONDKAPREMONT_MODE	68.386172	210295
LIVINGAPARTMENTS_MODE	68.354953	210199
LIVINGAPARTMENTS_AVG	68.354953	210199

	Percent Miss	Total Miss
FLOORSMAX_AVG	49.761142	153018
FLOORSMAX_MODE	49.761142	153018
FLOORSMAX_MEDI	49.761142	153018
YEARS_BEGINEXPLUATATION_AVG	48.781321	150005
YEARS_BEGINEXPLUATATION_MODE	48.781321	150005
YEARS_BEGINEXPLUATATION_MEDI	48.781321	150005
TOTALAREA_MODE	48.268809	148429
EMERGENCYSTATE_MODE	47.398579	145753
OCCUPATION_TYPE	31.345507	96389

Pada data yang hilang lebih dari 50% akan didrop dan untuk data kolom yang bertipe numerik akan diisi oleh median serta data kolom yang bertipe kategorikal akan diisi oleh modus

# Missing Value Identification

```
Value count kolom CODE_GENDER:  
F      202448  
M      105059  
XNA      4  
Name: CODE_GENDER, dtype: int64
```

```
Value count kolom NAME_FAMILY_STATUS:  
Married      196432  
Single / not married  45444  
Civil marriage  29775  
Separated      19770  
Widow      16088  
Unknown      2  
Name: NAME_FAMILY_STATUS, dtype: int64
```

Terdapat unique value yang kurang relevan dengan variabel data:

1. Terdapat unique XNA pada Variabel "CODE\_GENDER"
2. Terdapat unique Unknown pada Variabel "NAME\_FAMILY\_STATUS"

Adanya unique value yang tidak relevan dengan informasi variabel dapat membuat hasil analisis kurang valid sehingga perlu dilakukan identifikasi penyebabnya.

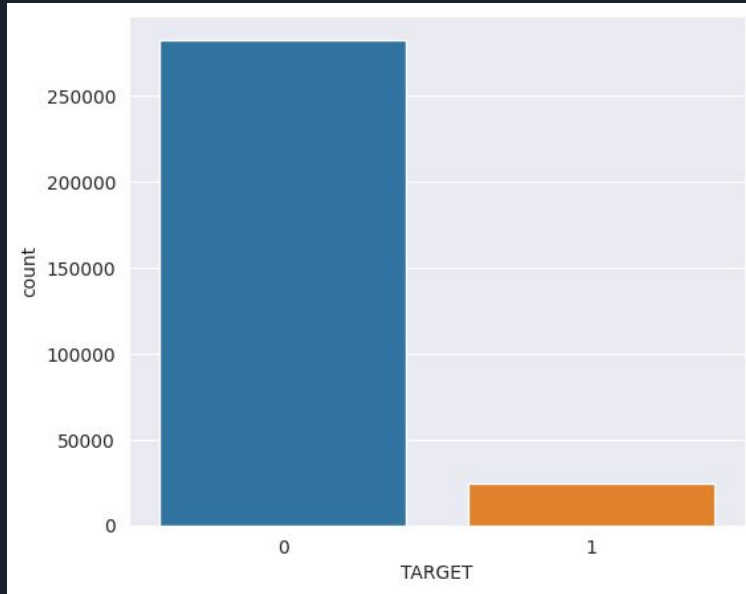
# Handling Missing Value

```
Value count kolom CODE_GENDER:  
F      202448  
M      105059  
XNA      4  
Name: CODE_GENDER, dtype: int64
```

```
Value count kolom NAME_FAMILY_STATUS:  
Married      196432  
Single / not married  45444  
Civil marriage  29775  
Separated     19770  
Widow         16088  
Unknown        2  
Name: NAME_FAMILY_STATUS, dtype: int64
```

Nilai XNA pada kolom 'CODE\_GENDER' dan nilai Unknown pada kolom 'NAME\_FAMILY\_STATUS' tidak dipengaruhi oleh kolom lain karena tidak ada pola keterikatan tertentu antara unique value tersebut dengan unique value dari kolom lain sehingga kemungkinan disebabkan oleh human error. Dengan demikian, dapat dihandling dengan menghapus baris karena jumlah unique value tersebut hanya sedikit maka tidak memiliki pengaruh signifikan terhadap hasil analisis.

# Cek Unbalance Data



Dari hasil value count dan count plot di atas, dapat terlihat bahwa data pada variabel target tidak seimbang dan terjadi ketimpangan.

1. Untuk variabel dengan target==1 (membayar tepat jatuh temponya) sebanyak 282686 orang.
2. Untuk variabel dengan target==0 (tidak bisa membayar tepat jatuh tempo) sebanyak 24825 orang.

```
0    282680
1     24825
Name: TARGET, dtype: int64
```

# Cek Outlier





# Cek Outlier


Jumlah outlier pada kolom numerik:

```
{'SK_ID_CURR': 0,  
'TARGET': 24825,  
'CNT_CHILDREN': 4272,  
'AMT_INCOME_TOTAL': 14034,  
'AMT_CREDIT': 6562,  
'AMT_ANNUITY': 7504,  
'AMT_GOODS_PRICE': 14728,  
'REGION_POPULATION_RELATIVE': 8412,  
'DAYS_BIRTH': 0,  
'DAYS_EMPLOYED': 72216,  
'DAYS_REGISTRATION': 659,  
'DAYS_ID_PUBLISH': 0,  
'FLAG_MOBIL': 1,  
'FLAG_EMP_PHONE': 55386,  
'FLAG_WORK_PHONE': 61303,  
'FLAG_CONT_MOBILE': 574,  
'FLAG_PHONE': 0,  
'FLAG_EMAIL': 17442,  
'CNT_FAM_MEMBERS': 4007,  
'REGION_RATING_CLIENT': 80526,  
'REGION_RATING_CLIENT_W_CITY': 78026,  
'HOUR_APPR_PROCESS_START': 2257,  
'REG_REGION_NOT_LIVE_REGION': 4657,  
'REG_REGION_NOT_WORK_REGION': 15611,  
'LIVE_REGION_NOT_WORK_REGION': 12502,  
'REG_CITY_NOT_LIVE_CITY': 24039,  
'REG_CITY_NOT_WORK_CITY': 70866,  
'LIVE_CITY_NOT_WORK_CITY': 55214,  
'EXT_SOURCE_2': 0,  
'EXT_SOURCE_3': 4312,  
'YEARS_BEGINEXPLUATATION_AVG': 142013,
```

```
'FLOORSMAX_MEDI': 90881,  
'TOTALAREA_MODE': 141913,  
'OBS_30_CNT_SOCIAL_CIRCLE': 19969,  
'DEF_30_CNT_SOCIAL_CIRCLE': 35164,  
'OBS_60_CNT_SOCIAL_CIRCLE': 19562,  
'DEF_60_CNT_SOCIAL_CIRCLE': 25767,  
'DAYS_LAST_PHONE_CHANGE': 435,  
'FLAG_DOCUMENT_2': 13,  
'FLAG_DOCUMENT_3': 0,  
'FLAG_DOCUMENT_4': 25,  
'FLAG_DOCUMENT_5': 4647,  
'FLAG_DOCUMENT_6': 27078,  
'FLAG_DOCUMENT_7': 58,  
'FLAG_DOCUMENT_8': 25024,  
'FLAG_DOCUMENT_9': 1198,  
'FLAG_DOCUMENT_10': 7,  
'FLAG_DOCUMENT_11': 1203,  
'FLAG_DOCUMENT_12': 2,  
'FLAG_DOCUMENT_13': 1084,  
'FLAG_DOCUMENT_14': 903,  
'FLAG_DOCUMENT_15': 372,  
'FLAG_DOCUMENT_16': 3053,  
'FLAG_DOCUMENT_17': 82,  
'FLAG_DOCUMENT_18': 2500,  
'FLAG_DOCUMENT_19': 183,  
'FLAG_DOCUMENT_20': 156,  
'FLAG_DOCUMENT_21': 103,  
'AMT_REQ_CREDIT_BUREAU_HOUR': 1626,  
'AMT_REQ_CREDIT_BUREAU_DAY': 1489,  
'AMT_REQ_CREDIT_BUREAU_WEEK': 8536,  
'AMT_REQ_CREDIT_BUREAU_MON': 43758,  
'AMT_REQ_CREDIT_BUREAU_QRT': 50574,  
'AMT_REQ_CREDIT_BUREAU_YEAR': 7233}
```

Jumlah outlier pada kolom kategorikal:

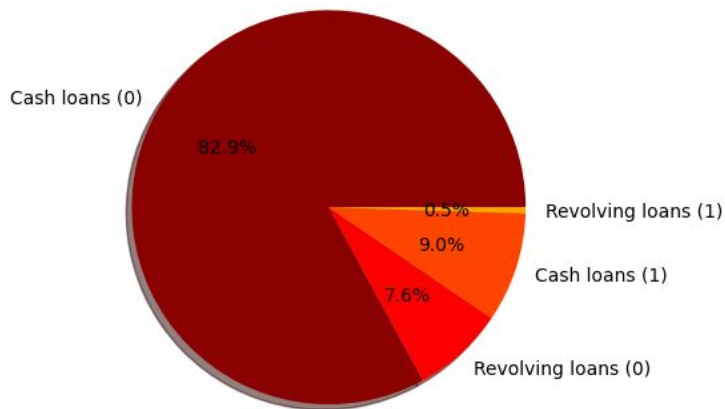
```
{'NAME_CONTRACT_TYPE': 0,  
'CODE_GENDER': 0,  
'FLAG_OWN_CAR': 0,  
'FLAG_OWN_REALTY': 0,  
'NAME_TYPE_SUITE': 0,  
'NAME_INCOME_TYPE': 1,  
'NAME_EDUCATION_TYPE': 0,  
'NAME_FAMILY_STATUS': 0,  
'NAME_HOUSING_TYPE': 0,  
'OCCUPATION_TYPE': 0,  
'WEEKDAY_APPR_PROCESS_START': 0,  
'ORGANIZATION_TYPE': 0,  
'EMERGENCYSTATE_MODE': 0}
```



# **Exploratory Data Analysis (EDA)**

# Tipe Kontrak **vs** Target

Kemampuan Membayar Loans berdasarkan tipe kontrak

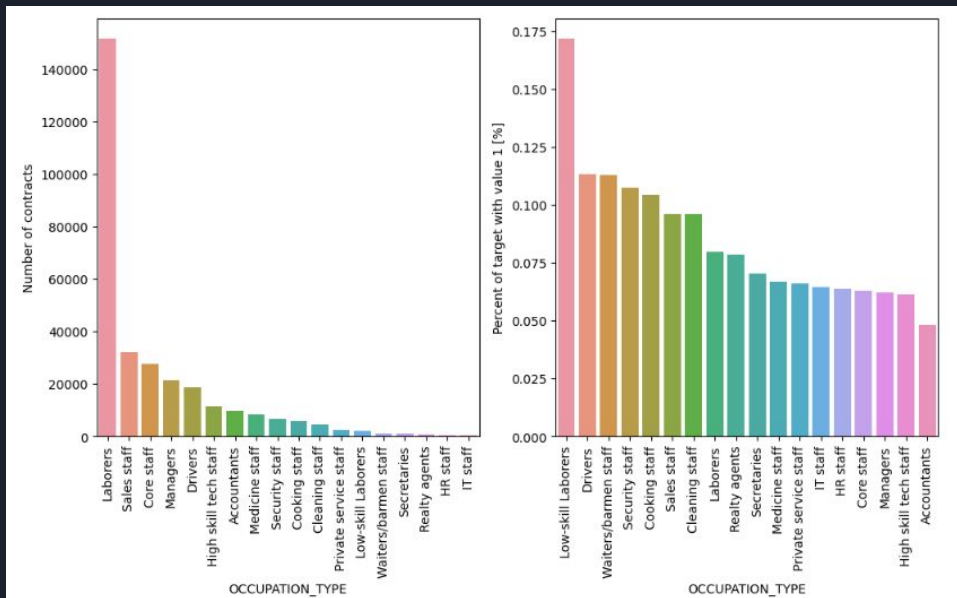


NAME_CONTRACT_TYPE	TARGET	SK_ID_CURR
Cash loans	0	255011
Revolving loans	0	27671
Cash loans	1	23221
Revolving loans	1	1604

Peminjam dengan type pinjaman Cash lebih banyak jumlahnya dibandingkan dengan Revolving. Dari barplot disamping terlihat juga bahwa terdapat ketimpangan data.

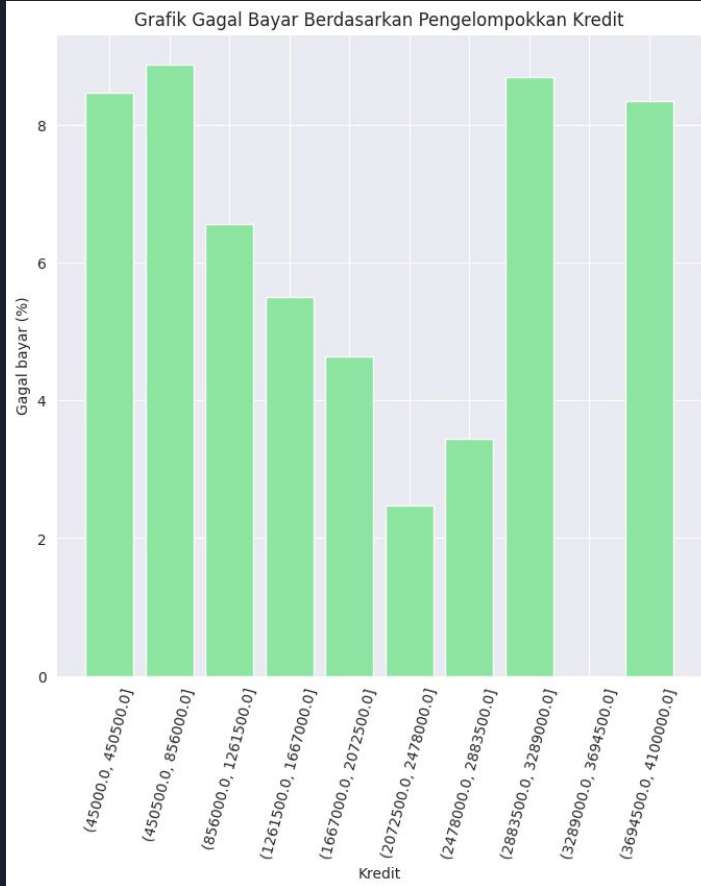
# Visualisasi Pekerjaan Loans

OCCUPATION_TYPE	TARGET	SK_ID_CURR
Laborers	0	139458
Sales staff	0	29010
Core staff	0	25831
Managers	0	20043
Drivers	0	16486
Laborers	1	12116
High skill tech staff	0	10679
Accountants	0	9339
Medicine staff	0	7965
Security staff	0	5999
Cooking staff	0	5325
Cleaning staff	0	4206
Sales staff	1	3052
Private service staff	0	2477
Drivers	1	2107
Core staff	1	1738
Low-skill Laborers	0	1733
Managers	1	1328
Secretaries	0	1213
Waiters/barmen staff	0	1196
Security staff	1	722
High skill tech staff	1	701
Realty agents	0	692
Cooking staff	1	621
Medicine staff	1	572
HR staff	0	527
IT staff	0	492
Accountants	1	474
Cleaning staff	1	447
Low-skill Laborers	1	359



Loans dengan pekerjaan sebagai karyawan adalah peminjam terbanyak. Untuk prosentase target==1, Karyawan low-skill menunjukkan prosentase terbanyak yaitu < 17,5%. Sedangkan prosentase terendah untuk target==1 adalah pekerja accountants, sehingga dimungkinkan lebih aman jika memberikan pinjaman ke pada pekerja akuntan.

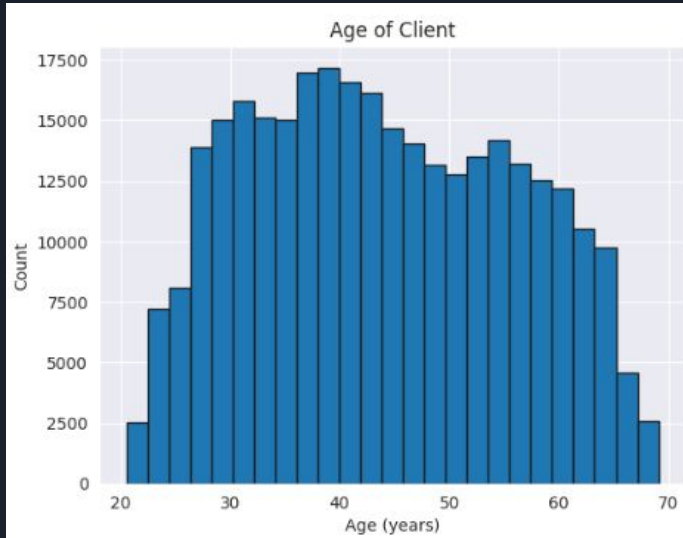
# Distribusi Kredit



Secara umum, persentase gagal bayar cenderung menurun seiring dengan peningkatan jumlah kredit.

Terdapat satu kelompok kredit dengan rentang (3289000.0, 3694500.0) yang memiliki persentase gagal bayar sebesar 0%. Ini berarti tidak ada kejadian gagal bayar dalam kelompok kredit tersebut.

# Distribusi Days from Birth



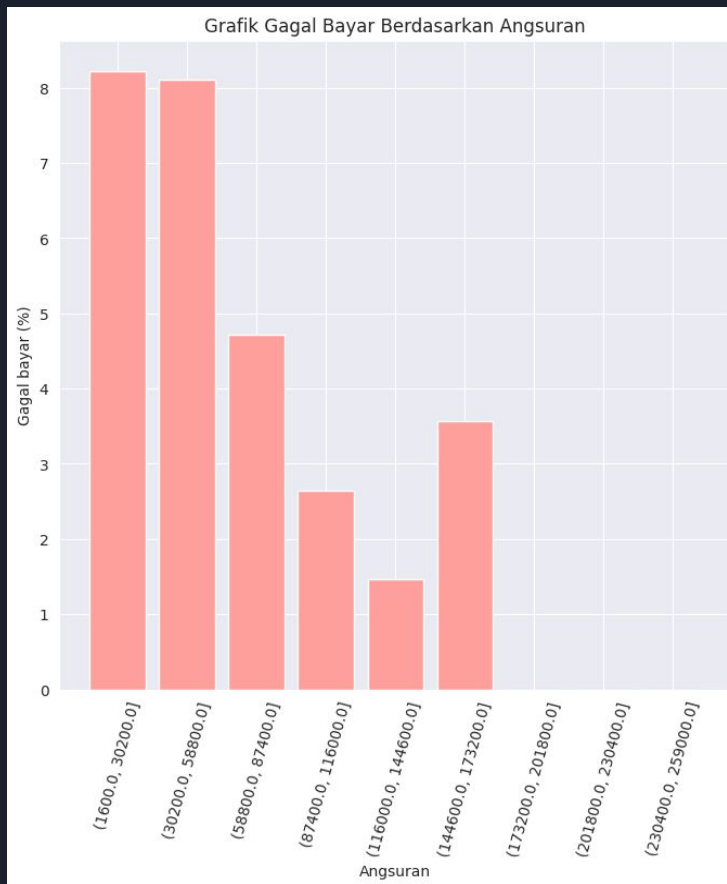
	TARGET	DAYS_BIRTH	YEARS_BIRTH	YEARS_BINNED
0	1	9461	25.920548	(25.0, 30.0]
1	0	16765	45.931507	(45.0, 50.0]
2	0	19046	52.180822	(50.0, 55.0]
3	0	19005	52.068493	(50.0, 55.0]
4	0	19932	54.608219	(50.0, 55.0]
5	0	16941	46.413699	(45.0, 50.0]
6	0	13778	37.747945	(35.0, 40.0]
7	0	18850	51.643836	(50.0, 55.0]
8	0	20099	55.065753	(55.0, 60.0]
9	0	14469	39.641096	(35.0, 40.0]

- Pada baris pertama, peminjam memiliki nilai 'TARGET' = 1, artinya mereka memiliki risiko kredit yang tinggi.
- Usia peminjam dihitung dalam tahun ('YEARS\_BIRTH'), dan dalam contoh ini peminjam pertama memiliki usia sekitar 25 tahun.
- Usia peminjam kemudian dikategorikan dalam rentang usia ('YEARS\_BINNED'). Peminjam pertama termasuk dalam kategori usia antara 25 hingga 30 tahun.

# Distribusi AMT-Annuity

TARGET	AMT_ANNUITY	AMT_ANNUITY_BINNED
1	24700.5	(1600.0, 30200.0]
0	35698.5	(30200.0, 58800.0]
0	6750.0	(1600.0, 30200.0]
0	29686.5	(1600.0, 30200.0]
0	21865.5	(1600.0, 30200.0]

	TARGET	AMT_ANNUITY
AMT_ANNUITY_BINNED		
(1600.0, 30200.0]	0.082187	18878.878432
(30200.0, 58800.0]	0.081122	39831.979720
(58800.0, 87400.0]	0.047131	66972.765846
(87400.0, 116000.0]	0.026437	97831.779310
(116000.0, 144600.0]	0.014599	126411.043796
(144600.0, 173200.0]	0.035714	153360.642857
(173200.0, 201800.0]	0.000000	178010.470588
(201800.0, 230400.0]	0.000000	221277.728571
(230400.0, 259000.0]	0.000000	258025.500000





# Korelasi

## Most Positive Correlations:

FLAG_DOCUMENT_3	0.044338
REG_CITY_NOT_LIVE_CITY	0.044394
FLAG_EMP_PHONE	0.045985
REG_CITY_NOT_WORK_CITY	0.050994
DAYS_ID_PUBLISH	0.051457
DAYS_LAST_PHONE_CHANGE	0.055220
REGION_RATING_CLIENT	0.058901
REGION_RATING_CLIENT_W_CITY	0.060895
DAYS_BIRTH	0.078244
TARGET	1.000000

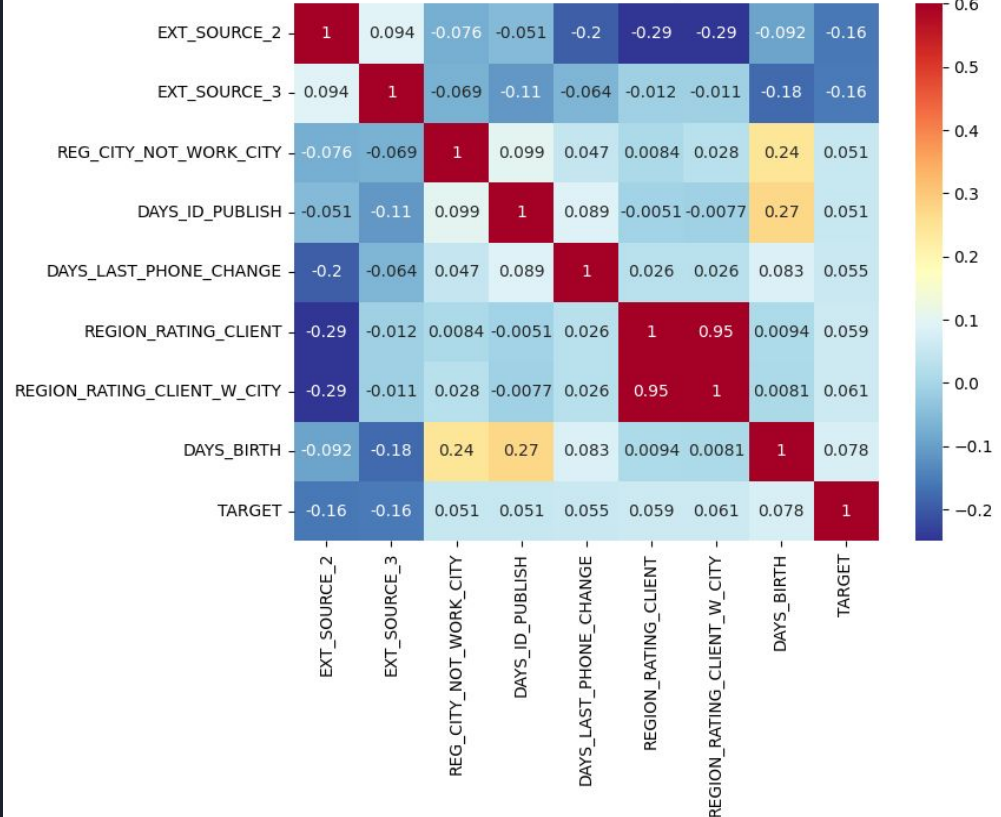
Name: TARGET, dtype: float64

## Most Negative Correlations:

EXT_SOURCE_2	-0.160293
EXT_SOURCE_3	-0.155898
DAYS_EMPLOYED	-0.044935
AMT_GOODS_PRICE	-0.039625
FLOORSMAX_AVG	-0.039388
FLOORSMAX_MEDI	-0.039159
FLOORSMAX_MODE	-0.038379
REGION_POPULATION_RELATIVE	-0.037225
AMT_CREDIT	-0.030371
TOTALAREA_MODE	-0.030216

Name: TARGET, dtype: float64

Correlation Heatmap





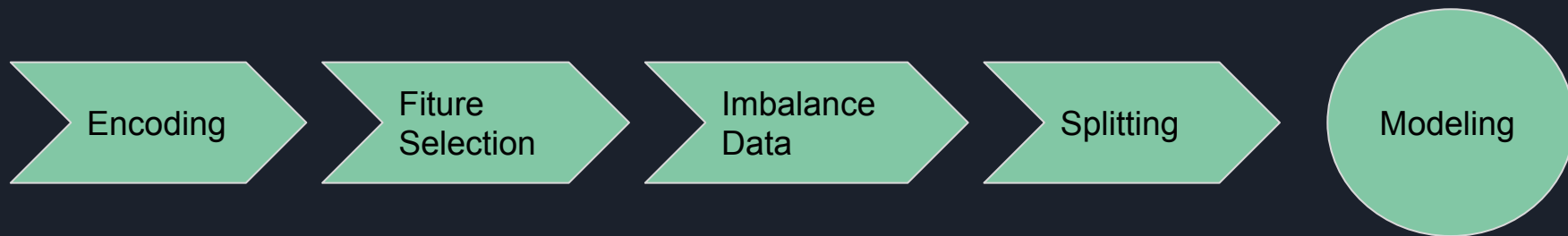
A blue parallelogram and a light green parallelogram are positioned in the top-left corner of the slide. The background is a dark navy blue with several diagonal stripes in varying shades of blue and grey.

# **Modelling**



# Modelling

Beberapa hal yang dilakukan sebelum modelling adalah sebagai berikut:



# Encoding

	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	...	FLAG_DOCUMENT_18	P
0	100002	1	0	1	0	1	0	202500.0	406597.5	24700.5	...	0	
1	100003	0	0	0	0	0	0	270000.0	1293502.5	35698.5	...	0	
2	100004	0	1	1	1	1	0	67500.0	135000.0	6750.0	...	0	
3	100006	0	0	0	0	1	0	135000.0	312682.5	29686.5	...	0	
4	100007	0	0	1	0	1	0	121500.0	513000.0	21865.5	...	0	

Dilakukan untuk mengubah data kategorik menjadi data numerik.

# Feature Selection

Feature selection digunakan untuk memilih fitur yang memiliki korelasi yang tinggi terhadap fitur yang menjadi target. Dengan fitur selection pada saat modelling bisa meningkatkan keakurasian data. Dalam menangani data yang kita gunakan, metode fitur selection yang kita gunakan adalah Pearson Correlation.

Fitur dengan Korelasi Tinggi:

	index	TARGET
0	EXT_SOURCE_2	-0.160294
1	EXT_SOURCE_3	-0.155899
2	REG_CITY_NOT_WORK_CITY	0.050992
3	DAYS_ID_PUBLISH	0.051457
4	NAME_EDUCATION_TYPE	0.054698
5	CODE_GENDER	0.054710
6	DAYS_LAST_PHONE_CHANGE	0.055219
7	REGION_RATING_CLIENT	0.058901
8	REGION_RATING_CLIENT_W_CITY	0.060895
9	DAYS_BIRTH	0.078242
10	TARGET	1.000000



# Imbalance Data

Ketimpangan data dapat membuat model menjadi lebih condong ke data mayoritas sehingga perlu adanya handling terhadap data yang unbalance.

```
train_data_upsampled['TARGET'].value_counts()  
  
1      282686  
0      282682  
Name: TARGET, dtype: int64
```



# Splitting

```
# splitting the data
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.3, random_state=0)
print(x_train.shape, y_train.shape)
print(x_test.shape, y_test.shape)

(395757, 79) (395757,)
(169611, 79) (169611,)
```

## Regresi Logistik

```
Classification Report Training Model (Logistic Regression):
              precision    recall  f1-score   support

      0         0.57         0.60         0.59         85017
      1         0.58         0.55         0.57         84594

 accuracy         0.58         0.58         0.58         169611
 macro avg         0.58         0.58         0.58         169611
weighted avg         0.58         0.58         0.58         169611
```

Accuracy Score :  
0.5757232726650984  
Confusion matrix :  
[[50731 34286]  
 [37676 46918]]  
ROC AUC score is: 0.5756707872470724

## Decision Tree

```
Classification Report Training Model (Decision Tree Classifier):
              precision    recall  f1-score   support

      0         1.00         0.89         0.94         85017
      1         0.90         1.00         0.95         84594

 accuracy         0.95         0.95         0.95         169611
 macro avg         0.95         0.95         0.95         169611
weighted avg         0.95         0.95         0.95         169611
```

Accuracy Score :  
0.9463419235780698  
Confusion matrix :  
[[75933 9084]  
 [ 17 84577]]  
ROC AUC score is: 0.9464749108647165

## Random Forest Classifier

```
Classification Report Training Model (Random Forest Classifier):
              precision    recall  f1-score   support

      0         1.00         0.99         1.00         85017
      1         0.99         1.00         1.00         84594

 accuracy         1.00         1.00         1.00         169611
 macro avg         1.00         1.00         1.00         169611
weighted avg         1.00         1.00         1.00         169611
```

Accuracy Score :  
0.9970992447423811  
Confusion matrix :  
[[84542 475]  
 [ 17 84577]]  
ROC AUC score is: 0.9971059611252526

## KNN

```
Classification Report Training Model (KNeighborsClassifier):
              precision    recall  f1-score   support

      0         0.99         0.72         0.83         85017
      1         0.78         0.99         0.87         84594

 accuracy         0.86         0.86         0.86         169611
 macro avg         0.88         0.86         0.85         169611
weighted avg         0.88         0.86         0.85         169611
```

Accuracy Score :  
0.8558996763181633  
Confusion matrix :  
[[61255 23762]  
 [ 679 83915]]  
ROC AUC score is: 0.856238192121592

Menyeleksi  
beberapa  
Model

A blue parallelogram and a light green parallelogram are positioned in the top-left corner of the slide. The blue shape is on the left, and the green shape is to its right, partially overlapping it.

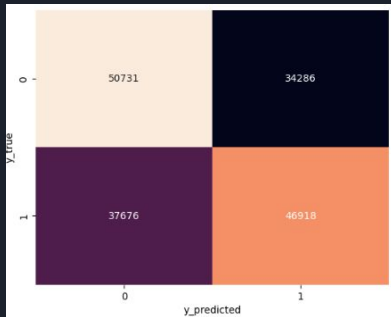
**EVALUATION**



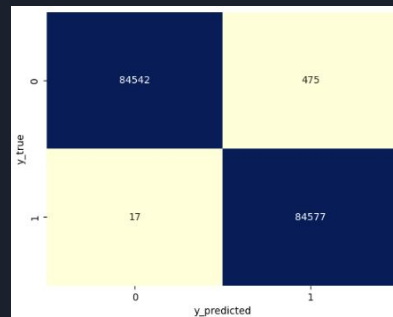
# Evaluasi

	Logistic Regression	Random Forest Classifier	Decision Tree Classifier	KNN
Accuracy	57.572327	99.709924	94.634192	85.589968
Precision	57.580704	99.710707	95.139405	88.417910
Recall	57.567079	99.710596	94.647491	85.623819
ROU_AUC	0.575671	0.997106	0.946475	0.856238

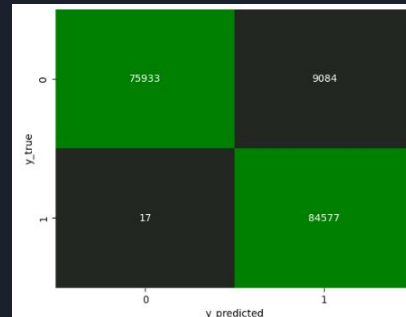
Regresi Logistik



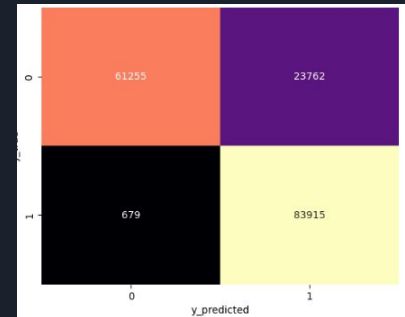
Random Forest



Decision Tree



KNN



Didapatkan model terbaik adalah **Random Forest Classifier**.



**DEPLOYMENT**



# Deployment

Pada tahap deployment, pengetahuan atau informasi yang telah diperoleh akan diatur dan dipresentasikan dalam bentuk khusus sehingga dapat digunakan oleh pengguna. Tahap deployment bisa berbentuk pembuatan laporan yang sederhana atau menerapkan proses data mining yang berulang ke dalam perusahaan. Untuk penyebaran yang lebih kompleks, dapat menerapkan proses tersebut secara paralel di departemen lain. Secara umum ada 2 aktivitas yang dilakukan pada tahap ini, yaitu: perencanaan dan monitoring hasil dari proses deployment serta melengkapi keseluruhan aktivitas sehingga menghasilkan output dari analisis data yang dilakukan.

# Dashboard

## Classification Loan with Home Credit Default Risk



Total Clients  
307.505

Accuracy  
99,7

Recall  
99,7

Precision  
99,7

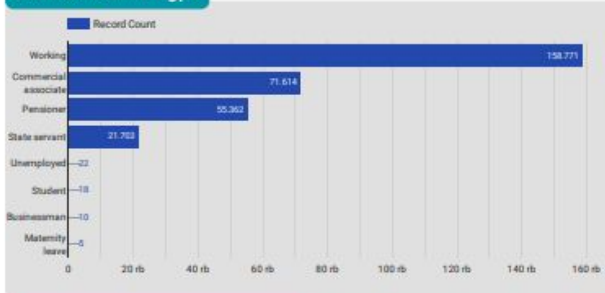
ROC-AUC  
99,70%

LOAN TYPE

TARGET

GENDER

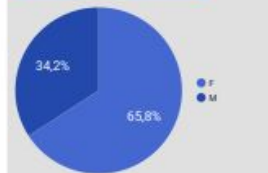
### Clients Income Type



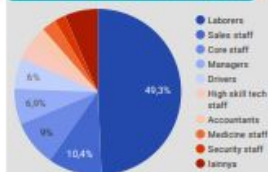
### Features Correlation to Target

FEATURE	TARGET	PERCENTASE
1. EXT_SOURCE_2	0.160294	
2. EXT_SOURCE_3	0.155899	
3. DAYS_BIRTH	0.078242	
4. REGION_RATING_CLIENT_W_CITY	0.060895	
5. REGION_RATING_CLIENT	0.058901	
6. DAYS_LAST_PHONE_CHANGE	0.055219	
7. DAYS_ID_PUBLISH	0.051457	
8. REG_CITY_NOT_WORK_CITY	0.050962	

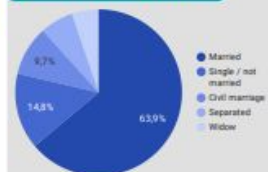
### Gender Distribution



### Clients Occupation Type



### Clients Family Status





THANK YOU

Kelompok 12