



HOME CREDIT DEFAULT RISK

BY SUKRI ALFIAN



DEFAULT RISK

Default Risk merupakan resiko karena memberikan kredit kepada nasabah yang memiliki kualifikasi yang berbeda-beda. Biasanya nasabah dengan kualifikasi tertentu cenderung kesulitan untuk membayar kembali kredit mereka. Oleh karena itu HOME CREDIT sebagai penyedia kredit perlu melakukan analisa Default Risk berdasarkan metode Data Mining dan Machine Learning Untuk memprediksi apakah seorang nasabah memenuhi kualifikasi sebagai orang yang dapat membayar kredit dengan teratur.

METODE 1

LOGISTIC REGRESSION

METODE 2

RANDOM FOREST

DATA OVERVIEW

Berdasarkan data yang ada dipilih *application* sebagai dataset untuk pemodelan. Pada *application* juga terdapat variabel sebagai berikut :

- EXT_SOURCE_1, EXT_SOURCE_2, dan EXT_SOURCE_3: Skor kredit eksternal yang disediakan oleh lembaga kredit.
- AMT_INCOME_TOTAL: Total pendapatan pemohon.
- AMT_CREDIT: Jumlah total kredit yang diajukan. Ini adalah variabel numerik yang diharapkan memiliki korelasi positif dengan risiko default.
- NAME_EDUCATION_TYPE: Tingkat pendidikan tertinggi yang dicapai oleh pemohon.
- NAME_INCOME_TYPE: Jenis penghasilan yang dimiliki oleh pemohon.
- DAYS_EMPLOYED: Jumlah hari pemohon bekerja.

DATA EXPLORATION

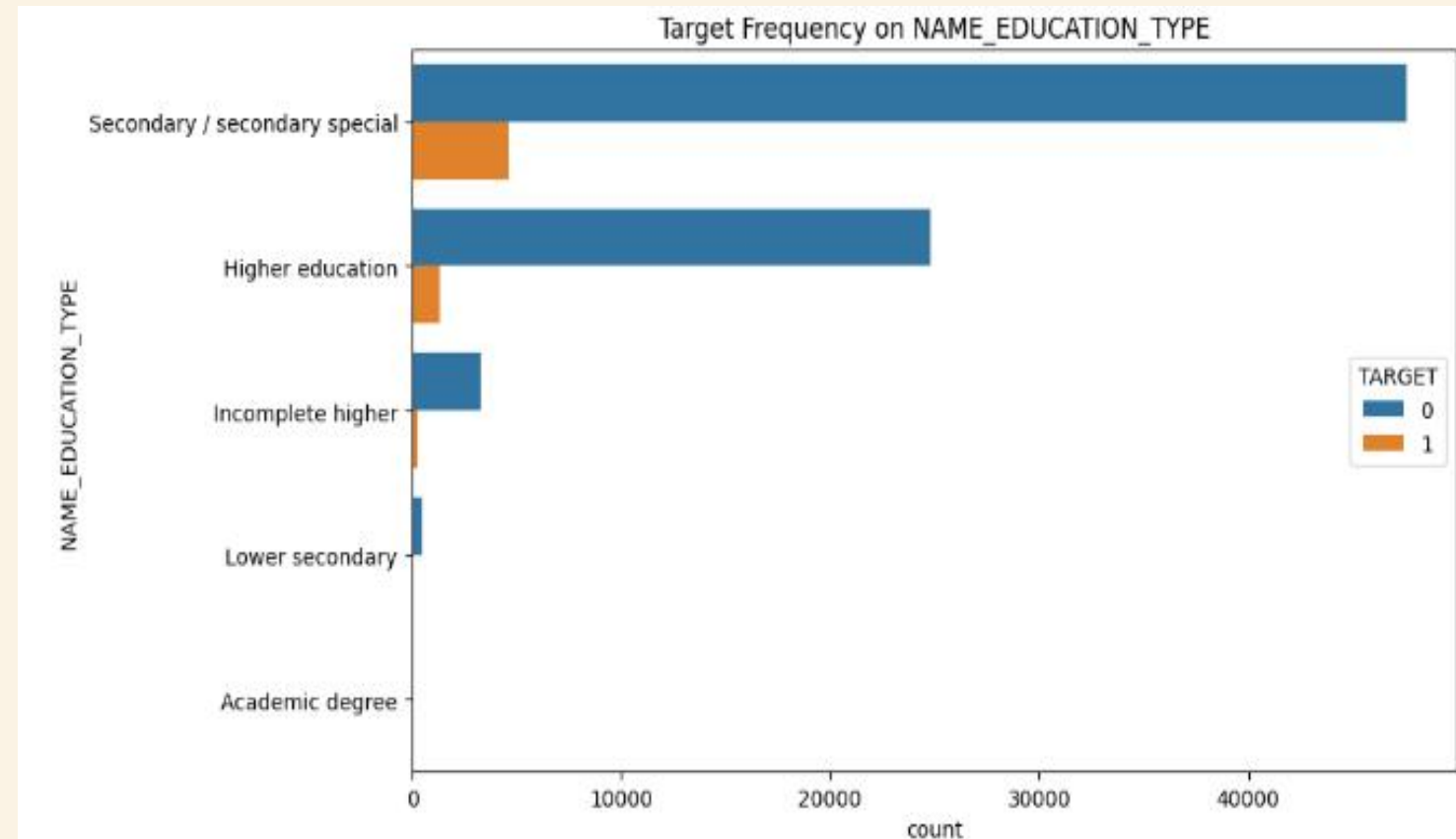
Data columns (total 15 columns):

#	Column	Non-Null	Count	Dtype
0	SK_ID_CURR	307511	non-null	int64
1	EXT_SOURCE_1	134133	non-null	float64
2	EXT_SOURCE_2	306851	non-null	float64
3	EXT_SOURCE_3	246546	non-null	float64
4	AMT_INCOME_TOTAL	307511	non-null	float64
5	AMT_CREDIT	307511	non-null	float64
6	NAME_EDUCATION_TYPE	307511	non-null	object
7	NAME_INCOME_TYPE	307511	non-null	object
8	DAYS_EMPLOYED	307511	non-null	int64
9	REGION_POPULATION_RELATIVE	307511	non-null	float64
10	OCCUPATION_TYPE	211120	non-null	object
11	AMT_GOODS_PRICE	307233	non-null	float64
12	NAME_FAMILY_STATUS	307511	non-null	object
13	FLAG_OWN_REALTY	307511	non-null	object
14	TARGET	307511	non-null	int64

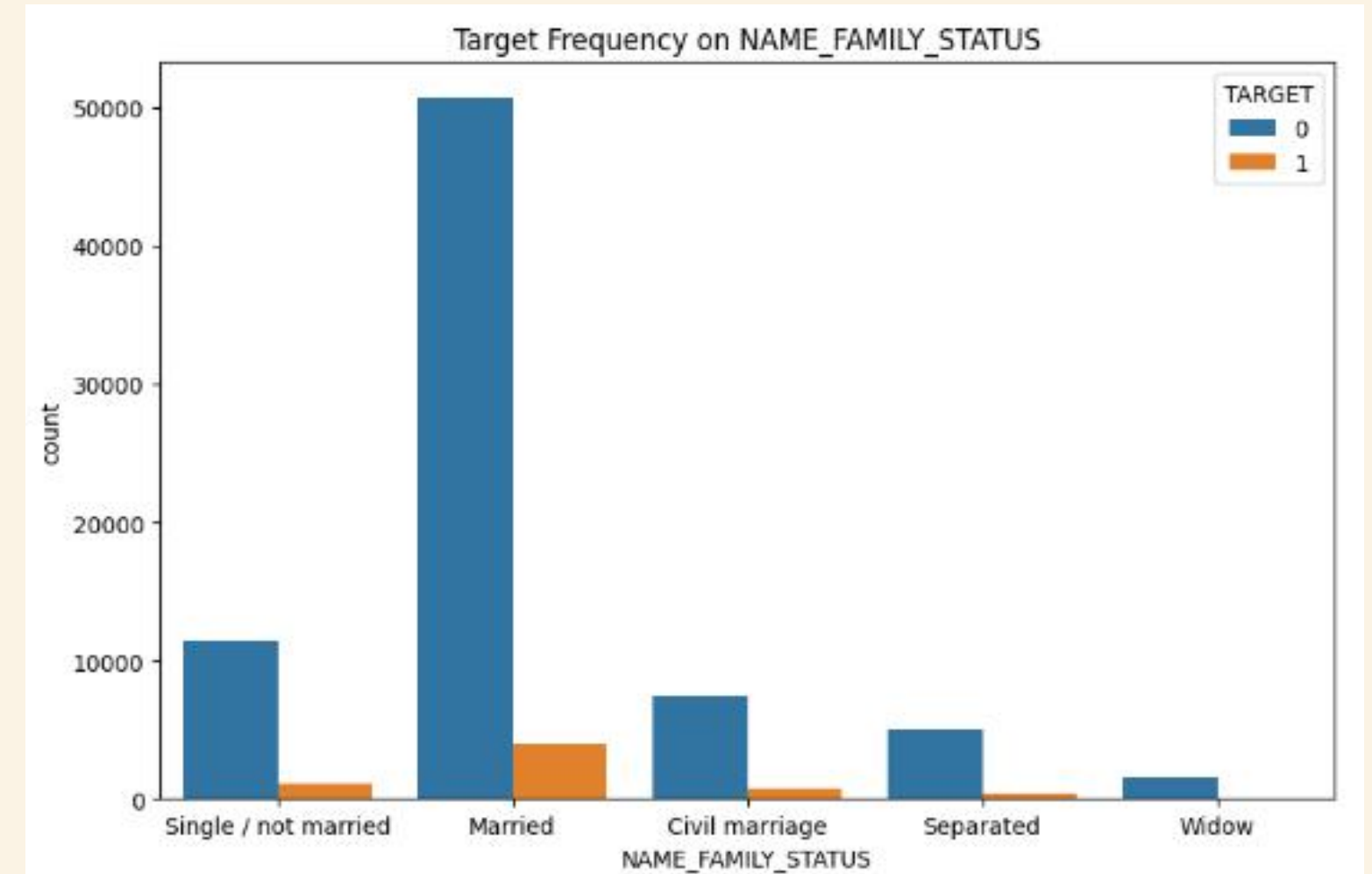
dtypes: float64(7), int64(3), object(5)

Dataset yang digunakan terdiri dari data *numeric* dan *string*. Pada pemodelan ini akan digunakan data *numeric* untuk memprediksi *default risk*.

DATA EXPLORATION



Nasabah dengan Pendidikan lebih rendah cenderung mengalami *default*. Hal ini dapat disebabkan karena pendapatan mereka tidak stabil karena rendahnya Pendidikan.



Nasabah dengan status menikah memiliki resiko *default* lebih tinggi. Hal ini bisa disebabkan karena mereka memiliki tanggungan hidup lebih dari 1 orang. Sehingga ekonomi mereka cenderung tidak stabil.

DATA PREPROCESSING

**DATA CLEANING AND
MANIPULATION**

NORMALIZATION

FEATURE SELECTION

RESULT AND EVALUATION

LOGISTIC REGRESSION

Train Result:

Accuracy Score: 92.44%

CLASSIFICATION REPORT:

	0	1	accuracy	macro avg	weighted avg
precision	0.924664	0.600000	0.924418	0.762332	0.900076
recall	0.999672	0.006005	0.924418	0.502838	0.924418
f1-score	0.960706	0.011891	0.924418	0.486298	0.888849
support	60972.000000	4996.000000	0.924418	65968.000000	65968.000000

Confusion Matrix:

```
[[60952  20]
 [ 4966  30]]
```

Test Result:

Accuracy Score: 92.42%

CLASSIFICATION REPORT:

	0	1	accuracy	macro avg	weighted avg
precision	0.924355	0.625000	0.92421	0.774678	0.901631
recall	0.999803	0.003994	0.92421	0.501898	0.924210
f1-score	0.960600	0.007937	0.92421	0.484268	0.888282
support	15241.000000	1252.000000	0.92421	16493.000000	16493.000000

Confusion Matrix:

```
[[15238   3]
 [ 1247   5]]
```

RANDOM FOREST

Train Result:

Accuracy Score: 100.00%

CLASSIFICATION REPORT:

	0	1	accuracy	macro avg	weighted avg
precision	1.0	1.0	1.0	1.0	1.0
recall	1.0	1.0	1.0	1.0	1.0
f1-score	1.0	1.0	1.0	1.0	1.0
support	60972.0	4996.0	1.0	65968.0	65968.0

Confusion Matrix:

```
[[60972   0]
 [   0 4996]]
```

Test Result:

Accuracy Score: 92.40%

CLASSIFICATION REPORT:

	0	1	accuracy	macro avg	weighted avg
precision	0.925426	0.490566	0.924028	0.707996	0.892415
recall	0.998228	0.020767	0.924028	0.509498	0.924028
f1-score	0.960449	0.039847	0.924028	0.500148	0.890566
support	15241.000000	1252.000000	0.924028	16493.000000	16493.000000

Confusion Matrix:

```
[[15214  27]
 [ 1226  26]]
```

REKOMENDASI

1. Nasabah Dengan Pendidikan Lebih Rendah Dan Pendapatan Tidak Stabil Cenderung Mengalami *default*. Memberikan pinjaman kepada nasabah dengan kualifikasi ini harus dipertimbangkan dengan baik.
2. Nasabah dengan status menikah atau tanggungan lebih dari satu memiliki resiko *default* yang tinggi. Perlu melakukan pertimbangan terhadap pendapatan dan jenis pekerjaan nasabah untuk menghindari resiko *default*.