

Final Project

Home Credit Indonesia Data Scientist Project Based Internship Program

Presented by
Taufik Yasir Sukarda

Final Project

Home Credit Data Scientist Based
Internship Program

Supported by:
Rakamin Academy



Created by:

Taufik Yasir Sukarda

taufikyasir4@gmail.com

[My LinkedIn Profile](#)

“My name is Taufik Yasir Sukarda, and I am an enthusiastic and dedicated learner who is eager to explore the world of data. With a background in geology, I am excited to navigate this dynamic field and discover new insights. Currently, I am engaging in self-directed study and online courses to improve my skills and gain proficiency in the data field. I am passionate about using data to make informed decisions, and I am looking forward to contributing my growing expertise, dedication, and enthusiasm to any opportunity that comes my way. ”

Case Study

PT Home Credit Indonesia merupakan perusahaan pembiayaan berbasis teknologi, menghadirkan produk dan layanan keuangan yang dapat membantu menciptakan berbagai kesempatan bagi masyarakat dalam meningkatkan kehidupan mereka dengan menyediakan layanan keuangan yang bertanggung jawab dan dapat dipercaya.

Kesempatan yang diberikan tentunya dibarengi dengan kewajiban klien untuk melakukan pembayaran yang lancar juga maka dari itu PT Home credit sebisa mungkin ingin memberikan layanan kepada klien dengan potensi.

PT home credit telah mengumpulkan data-data dengan berbagai fitur dari klien klien yang ada yang Dimana diharapkan dengan memanfaatkan data ini dapat dirancang suatu model machine learning yang mampu memprediksi klien dengan potensi pembayaran yang lancar sehingga layanan tepat sasaran mendatangkan keuntungan.

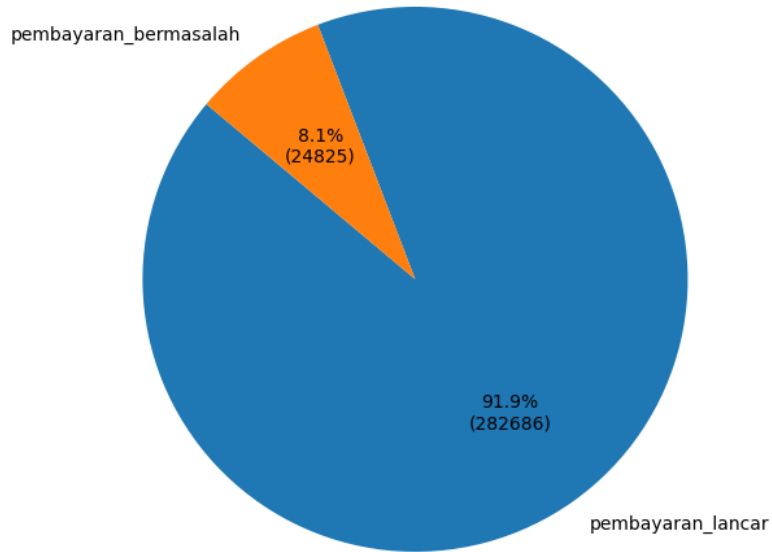
Dataset

Dataset yang tersedia adalah sampel data yang terdiri dari 122 kolom dan 307511 baris data, berisikan fitur fitur klien disertai satu kolom target yang berisikan kemampuan pembayaran dari klien apakah ada masalah pembayaran atau tidak

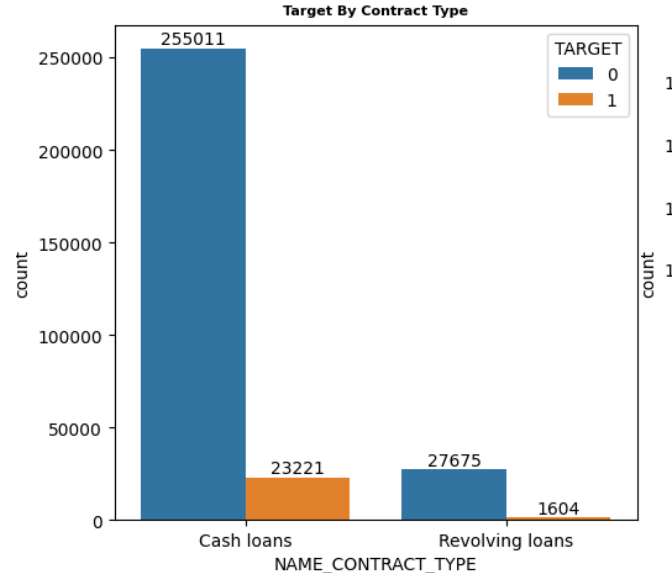
Business Insight

Dari keseluruhan data klien yang ada hanya sebesar 8.1 % klien yang mengalami masalah dalam pembayaran

Sebaran Kemampuan pembayaran Klien



Jika dilihat berdasarkan tipe kontrak nya mayoritas klien menginginkan tipe kontrak cash loan sekitar 90% total klien berada pada tipe kontrak ini sedangkan untuk tipe revolving loan klien nya hanya 10% dari total klien terdata, ada potensi yang belum termaksimalkan pada revolving loan ini mungkin bisa dilakukan studi yang lebih mendalam serta dicari data tamban lainnya yang dapat menjelaskan rendahnya minat klien pada tipe kontrak ini, sehingga dapat diatur berbagai strategi yang ampuh agar potensi revolving loan ini termaksimalkan



Business Insight

	NAME_INCOME_TYPE	TARGET	Jumlah_klien	groupptc	totalpct
0	Businessman	0	10	100.000000	0.003252
1	Commercial associate	0	66257	92.515743	21.546221
2	Commercial associate	1	5360	7.484257	1.743027
3	Maternity leave	0	3	60.000000	0.000976
4	Maternity leave	1	2	40.000000	0.000650
5	Pensioner	0	52380	94.613634	17.033537
6	Pensioner	1	2982	5.386366	0.969721
7	State servant	0	20454	94.245035	6.651469
8	State servant	1	1249	5.754965	0.406164
9	Student	0	18	100.000000	0.005853
10	Unemployed	0	14	63.636364	0.004553
11	Unemployed	1	8	36.363636	0.002602
12	Working	0	143550	90.411528	46.681257
13	Working	1	15224	9.588472	4.950717

Mayoritas klien adalah pekerja, namun dapat dilihat ada potensi juga pada klien dengan jenis income commercial associate, dan sayang nya klien dengan tipe income businessman jumlahnya rendah ada kemungkinan kurang exposure terhadap tipe ini dengan potensi yang belum tergali, karna penyediaan layanan keuangan untuk bisnis tentunya sangat penting, mungkin diperlukan strategi yang lebih ampuh untuk meningkatkan keterlibatan klien

Machine Learning Model



DATASET

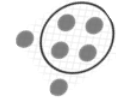
307511 rows Data
&
122 columns

NULL

Banyak kolom dengan nilai null melebihi 40% sehingga dilakukan drop kolom, untuk sisanya dilakukan imputasi median untuk data numerik dan imputasi modus untuk data kategorik



Duplicate:
0 duplicated



Outlier:
menggunakan model yang robust outlier sehingga tidak dilakukan handling



Feature Engineering:

Melakukan Drop fitur-fitur yang tidak relevan serta melakukan encoding pada fitur kategorik beberapa fitur dengan nilai negative juga diubah ke nilai absolutnya, kemudian fitur diseleksi dengan menggunakan selectKbest melihat fitur- yang memiliki pengaruh yang signifikan terhadap target



Split train_test data:
Menjadi 70% train dan
30% test data

Scaling:
Menstandarisasi data
sehingga siap untuk
Modeling

Machine Learning Model Evaluation

Model akhir yang dipilih adalah menggunakan algoritma model Logistic Regresion dengan Teknik pemilihan fitur melalui SelectKbest setelah dicoba ditetapkan pada 12 fitur yang memiliki skor tertinggi dengan matrik evaluasi berfokus pada akurasi dan juga recall algoritma Logistic Regresion memiliki skor evaluasi yang lumayan bagus dan seimbang untuk ke-dua matrik tersebut yaitu sebagai berikut ini:

```
accuracy(test): 0.6939536497062458  
accuracy(train): 0.7080113219944907  
recall(test): 0.5034399028733306  
recall(train): 0.7061487528115444
```

Walau nilai recall yang lumayan jauh antara train dan test namum algoritma ini adalah yang paling stabil jika disbanding algorithma lain yang terlalu overfitting skor diatas lebih stabil.

Insights and Recommendations

1. Ada potensi dari tipe kontrak revolving loans yang masih belum tergali maksimal hal ini dicerminkan dari sedikitnya klien pada tipe kontrak ini, perlu disusun strategi dan disertai exposure yang lebih untuk meningkatkan keterlibatan klien
2. Untuk klien dengan tipe income commercial associate memiliki potensi dilihat dari 20% klien adalah dari tipe ini sedangkan untuk klien dengan tipe income businessman keterlibatannya sangat rendah mungkin bisa digali lebih dalam potensi klien tipe ini

Tambahan terkait repositori Github pengerjaan project ini ada pada:

This Project Github

Thank You



Rakamin
Academy



Bank
Muamalat