

FINAL TASK

ID/X PARTNER – VIX RAKAMIN

BY: ANANTA ALFALLAH TARIGAN





Latar Belakang

Dalam dunia keuangan yang dinamis, evaluasi risiko kredit telah menjadi suatu keharusan untuk menilai kemungkinan pembayaran pinjaman. Kemajuan teknologi yang cepat, terutama di bidang data science dan machine learning, telah membuka peluang baru bagi lembaga keuangan untuk merevolusi prosedur penilaian risiko kredit mereka. Bank dan perusahaan pemberi pinjaman kini memiliki kesempatan untuk memanfaatkan kekuatan algoritma canggih untuk memperkuat upaya pemodelan risiko kredit mereka.

Tujuan

- Membangun model yang dapat memprediksi credit risk menggunakan dataset yang disediakan oleh company yang terdiri dari data pinjaman yang diterima dan yang ditolak.
- Nantinya model tersebut dapat memprediksi apakah calon peminjam layak atau tidak mendapatkan pinjaman, lewat variabel-variabel yang telah disesuaikan.

Metode

01

Pengumpulan dan Pra-pemrosesan Data

Tahap awal pemodelan risiko kredit melibatkan pengumpulan dan persiapan data. Dataset yang luas yang mencakup berbagai informasi, disesuaikan untuk memastikan keandalan dan kesesuaian data untuk pemodelan.

02

Seleksi dan Rekayasa Fitur

Seleksi fitur melibatkan identifikasi variabel yang paling signifikan yang memengaruhi evaluasi risiko kredit. Rekayasa fitur mencakup pembuatan variabel baru atau transformasi yang lebih akurat untuk menangkap pola dan hubungan yang mendasar di dalam data.

03

Pelatihan dan Evaluasi Model

Memilih dari berbagai algoritma machine learning berdasarkan persyaratan yang sesuai tugas pemodelan risiko kredit. Model ini kemudian dievaluasi menggunakan metrik seperti akurasi, presisi, recall.

Gambaran Umum Dataset

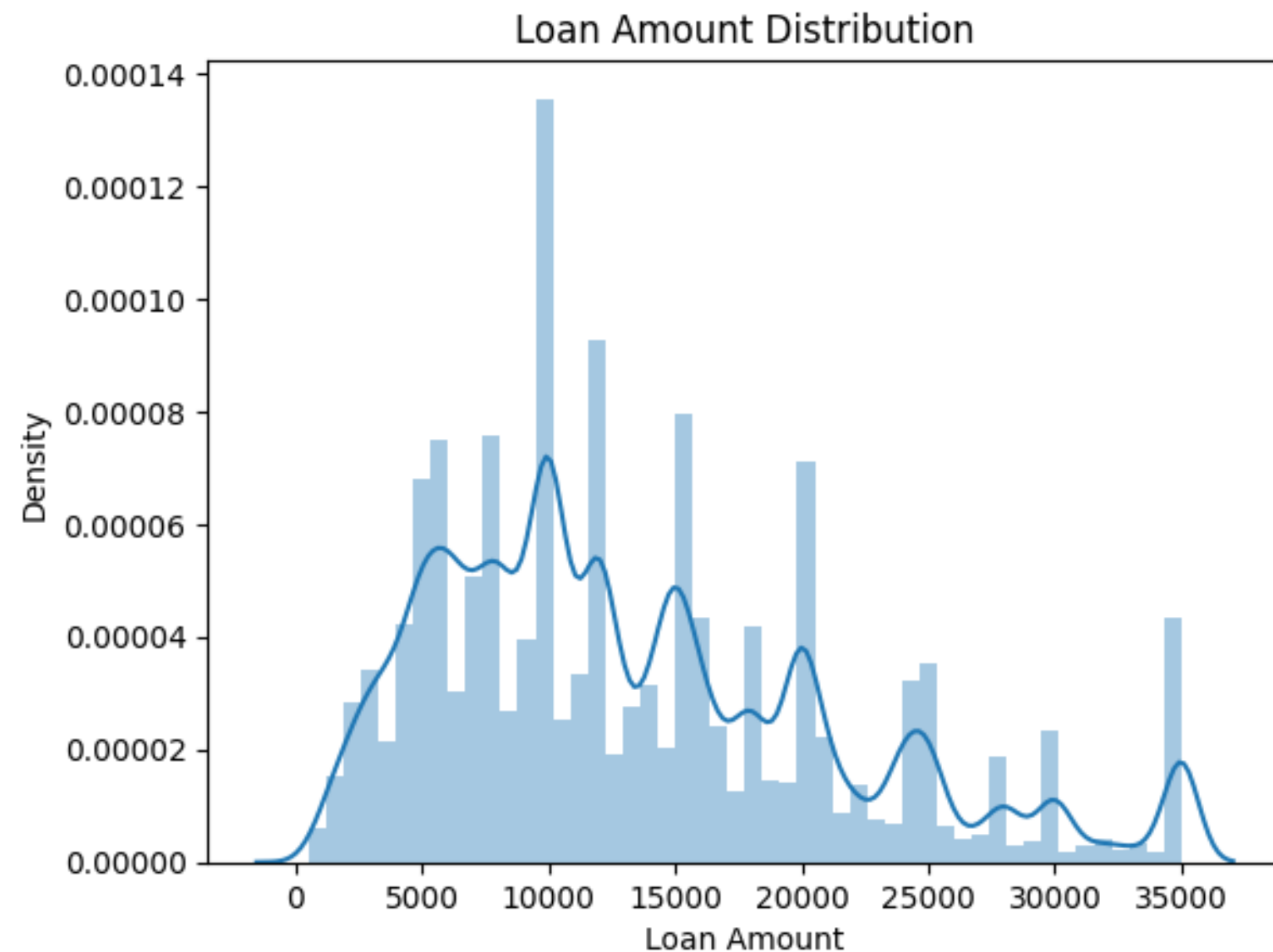
- Ukuran: 466285 baris × 75 kolom
- Jumlah duplikasi: 0
- Jumlah nilai kosong: sangat banyak dan beragam di tiap kolom

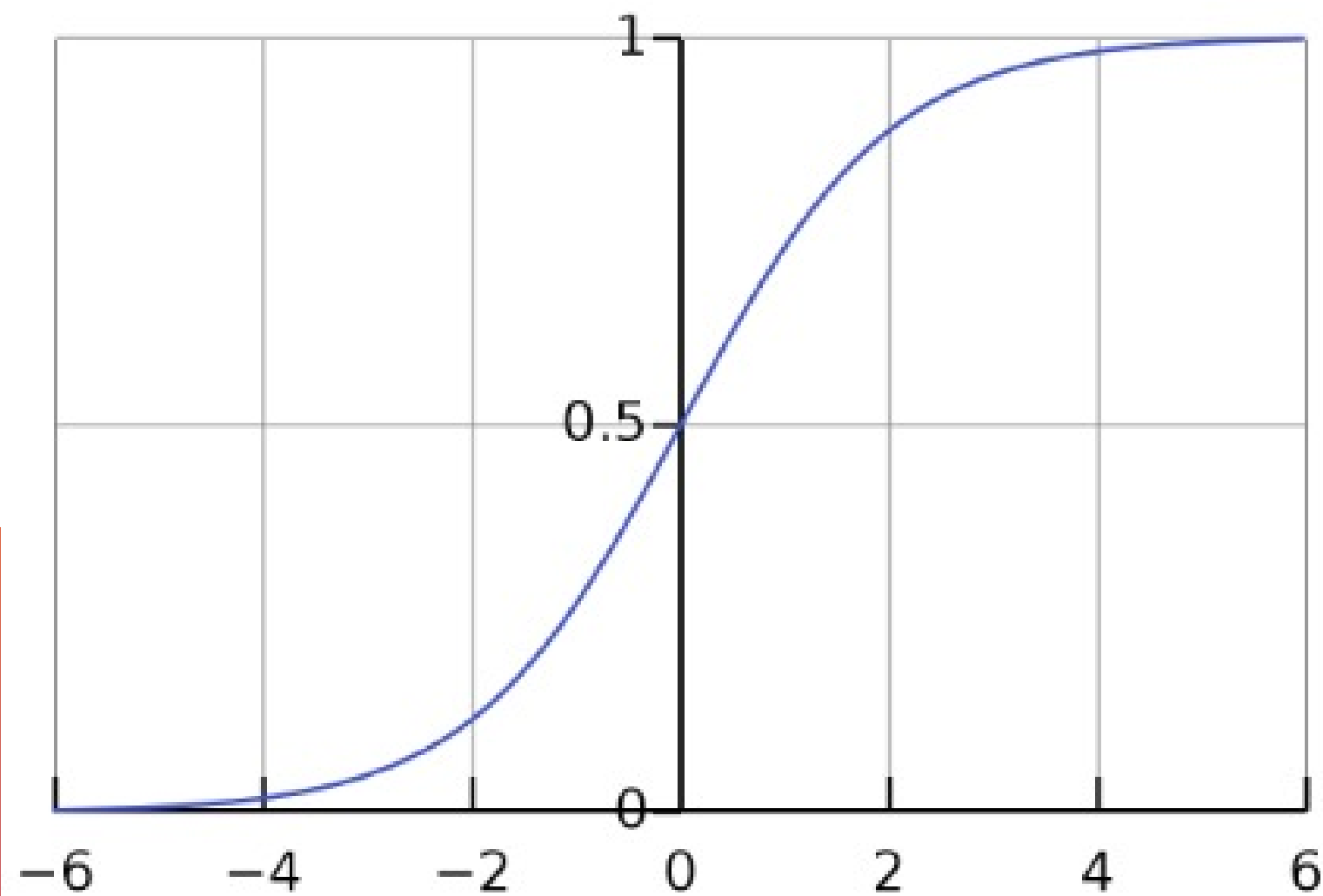
Unnamed: 0	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	grade	...	total_bal_il	il_util	open_rv_12m	open_rv_...
0	0	1077501	1296599	5000	5000	4975.0	36 months	10.65	162.87	B	...	NaN	NaN	NaN
1	1	1077430	1314167	2500	2500	2500.0	60 months	15.27	59.83	C	...	NaN	NaN	NaN
2	2	1077175	1313524	2400	2400	2400.0	36 months	15.96	84.33	C	...	NaN	NaN	NaN
3	3	1076863	1277178	10000	10000	10000.0	36 months	13.49	339.31	C	...	NaN	NaN	NaN
4	4	1075358	1311748	3000	3000	3000.0	60 months	12.69	67.79	B	...	NaN	NaN	NaN
...
466280	466280	8598660	1440975	18400	18400	18400.0	60 months	14.47	432.64	C	...	NaN	NaN	NaN
466281	466281	9684700	11536848	22000	22000	22000.0	60 months	19.97	582.50	D	...	NaN	NaN	NaN
466282	466282	9584776	11436914	20700	20700	20700.0	60 months	16.99	514.34	D	...	NaN	NaN	NaN
466283	466283	9604874	11457002	2000	2000	2000.0	36 months	7.90	62.59	A	...	NaN	NaN	NaN
466284	466284	9199665	11061576	10000	10000	9975.0	36 months	19.20	367.58	D	...	NaN	NaN	NaN

466285 rows × 75 columns

Poin Penting

- Setelah dilakukan cleaning dan penyesuaian data diketahui:
- Bahwa sebaran jumlah nominal pinjaman sebagian besar berada pada kisaran rendah sampai menengah.
 - Hanya sekitar 18,8%, dari total peminjam, yang terindikasi sebagai peminjam yang buruk/tidak layak (bad loan). Sedangkan sisanya 81,2% merupakan peminjam yang baik/layak.





Pelatihan Model: Logistic regression

- Dalam kasus pengklasifikasian 2 kelas peminjam, yakni good loan dan bad loan, kita dapat menggunakan algoritma logistic regression.
- Logistic regression merupakan salah satu metode yang umum digunakan untuk klasifikasi.
- Pada kasus klasifikasi peminjam ini, logistic regression bekerja dengan menghitung probabilitas kelas dari sebuah sampel good loan dan bad loan. Apabila sebuah calon peminjam memiliki probabilitas $> 50\%$ maka akan dikategorikan sebagai good loan, begitu pun sebaliknya.

Hasil Evaluasi Model

Menggunakan 40 fitur/kolom untuk memprediksi klasifikasi peminjam, dengan data latih sebesar 80% dan data test sebesar 20%, maka diperoleh metrik evaluasi sebagai berikut:

```
Accuracy: 0.9938710610002891
Classification Report:
```

	precision	recall	f1-score	support
0	1.00	0.99	0.99	8624
1	0.99	1.00	0.99	8671
accuracy			0.99	17295
macro avg	0.99	0.99	0.99	17295
weighted avg	0.99	0.99	0.99	17295

- Dalam kasus ini, akurasi model adalah sekitar **99%**, nilai yang tergolong sangat tinggi dan menunjukkan model memiliki kemampuan yang baik untuk memprediksi dengan benar.
- Precision (presisi), sebesar 1 dan 0.99, menandakan bahwa **prediksi positif** model akurat.
- Recall (recall), sebesar 0.99 dan 1, menandakan bahwa model dapat mengidentifikasi dengan benar **nilai yang sebenarnya positif**.



Kesimpulan

Model yang dibangun menunjukkan kinerja yang sangat baik dalam memprediksi klasifikasi biner `loan_status` dengan akurasi sekitar 99%. Hal ini menunjukkan kemampuan model untuk secara konsisten membuat prediksi calon peminjam yang layak atau tidak dengan benar.

Saran

Meski model dirasa sudah sangat baik, perlu adanya validasi eksternal menggunakan dataset yang belum pernah dilihat sebelumnya. Hal ini dapat memberikan gambaran tentang sejauh mana model dapat melakukan generalisasi pada data baru. Selain itu, apabila dirasa performa menurun maka mungkin diperlukan penyesuaian hyperparameter atau menggunakan algoritma lain, seperti jaringan saraf tiruan untuk membuat model yang lebih baik

