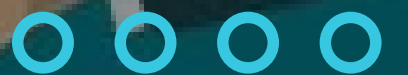




PREDICTING THE CREDIT RISK OF CUSTOMER LOANS

Alvito Dwinovan W
Final Task Data Scientist
IDX Partners X Rakamin Academy



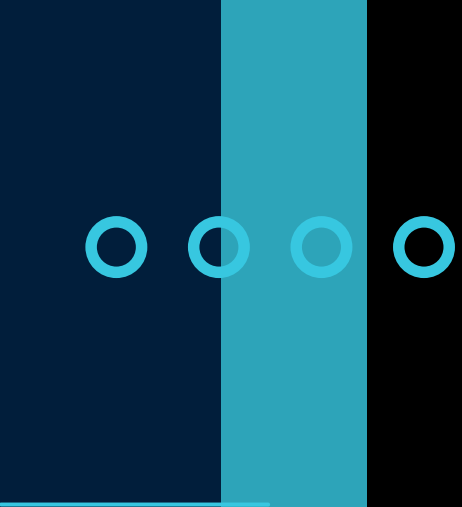


BACKGROUND

Keakuratan dalam memprediksi risiko kredit pinjaman pelanggan merupakan hal yang kritis bagi pemberi pinjaman dan lembaga keuangan. Tanpa metode yang andal dan akurat dalam menilai risiko kredit, keputusan terkait besaran pinjaman dan persetujuan aplikasi pinjaman menjadi kurang tepat.

Prediksi risiko kredit yang tidak tepat mengakibatkan

- Potensi kerugian keuangan
- Pengaruh pada profitabilitas
- Reputasi Perusahaan
- Kepatuhan Regulasi
- Pengaruh pada laju pertumbuhan



OBJECTIVE

Membuat model prediksi yang akurat untuk memprediksi risiko kredit nasabah perusahaan. Model ini akan memberikan informasi apakah seorang nasabah berpotensi default dalam pembayaran pinjaman atau tidak. Diharapkan perusahaan dapat mengurangi jumlah nasabah yang berpotensi gagal bayar, sehingga resiko kerugian perusahaan berkurang.



DATASET

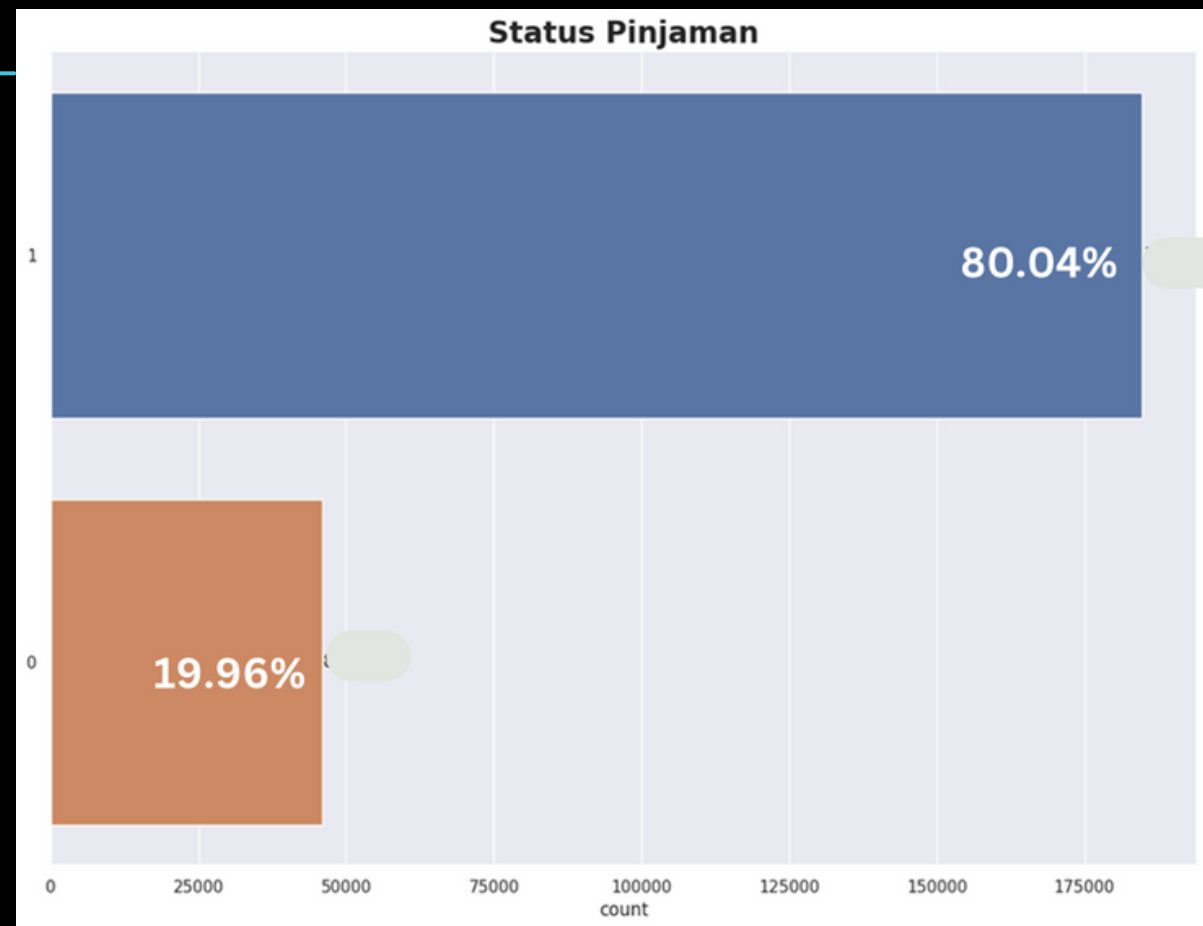
Data perusahaan terdiri dari 466.285 baris dan 75 kolom yang diduga dibutuhkan dalam memprediksi resiko pinjaman.

Pada kolom 'Loan Status' dikelompokkan menjadi pinjaman dengan kemungkinan untuk disetujui ataupun ditolak.

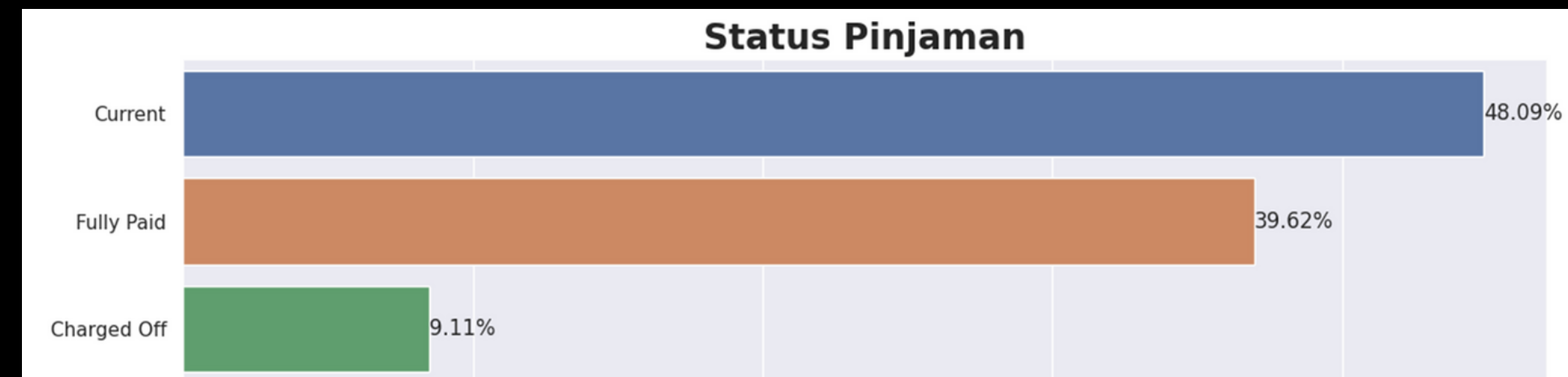
- Disetujui : Fully Paid
- Ditolak : Charged Off, Default, dan Does not meet the credit policy



EXPLORATORY DATA ANALYSIS



Penerimaan peminjaman yang sangat tinggi yaitu 80.04% perlu dipertimbangkan oleh perusahaan karena dapat berdampak besar pada kesehatan keuangan dan dapat meningkatkan risiko bagi perusahaan.



Sebagian besar pinjaman sedang berjalan (tidak mengalami keterlambatan) dan sudah sepenuhnya dilunasi.



DATA PRE-PROCESSING

- Remove kolom dengan Missing value > 50%, kolom dengan data unique 100%, dan kolom yang terdiri dari satu nilai, serta data duplikat
- Remove kolom kategorik yang dianggap tidak berpengaruh terhadap prediksi 'Loan status'.
- Mengisi missing value dengan nilai median dan mean
- Menggabungkan kolom yang berkorelasi
- Melakukan standardisasi pada data numerik
- Pembagian data training dan data test dengan 80:20

Setelah diproses, data yang digunakan untuk lanjut ke tahap permodelan terdiri dari 466285 baris dan 11 kolom. Dimana kolom-kolom tersebut memiliki pengaruh terhadap pengambilan keputusan pada 'Loan status'.

DATA PROCESSING

Data columns (total 12 columns):

#	Column	Non-Null Count	Dtype
0	loan_amnt	466285 non-null	float64
1	int_rate	466285 non-null	float64
2	revol_bal	466285 non-null	float64
3	out_prncp	466285 non-null	float64
4	total_pymnt	466285 non-null	float64
5	total_rec_int	466285 non-null	float64
6	total_rec_late_fee	466285 non-null	float64
7	recoveries	466285 non-null	float64
8	collection_recovery_fee	466285 non-null	float64
9	last_pymnt_amnt	466285 non-null	float64
10	verification_status	466285 non-null	int32
11	loan_status	466285 non-null	int32

dtypes: float64(10), int32(2)

memory usage: 39.1 MB

METHOD

Logistic regression adalah salah satu metode dalam statistik dan machine learning yang digunakan untuk melakukan analisis regresi pada data dengan dua kategori atau lebih. Logistic regression memiliki beberapa kelebihan yang membuatnya menjadi pilihan yang baik dalam prediksi risiko kredit:

- 1. Interpretabilitas:**
- 2. Penanganan Variabel Kategori:**
- 3. Probabilitas:**
- 4. Kemampuan Identifikasi Outlier:**
- 5. Skalabilitas:**
- 6. Pembaruan Model:**
- 7. Pemahaman Penyebab Risiko:**
- 8. Ketahanan Terhadap Overfitting:**

MODEL EVALUATION

confusion matrix

```
[[ 8312      0      0      4      0  275      0      0      0]
 [      2 44447      0      0      0  342      3     16     46]
 [      0   178      0      0      0      1      0      0      1]
 [   150      0      0      1      1     15      0      0      0]
 [      6      0      0      0      5   409      0      0      0]
 [     99      0      0      0      9 36690      0      0      0]
 [      0   623      0      0      0      8      2      0      5]
 [      0   223      0      0      0      1      0      0      4]
 [      0  1330      0      0      0     11      1      0     37]]
```

accuracy

0.9596491416193959

MODEL EVALUATION

```
classification report
              precision    recall  f1-score   support

     0           0.97       0.97       0.97        8591
     1           0.95       0.99       0.97       44856
     2           0.00       0.00       0.00         180
     3           0.20       0.01       0.01         167
     4           0.33       0.01       0.02         420
     5           0.97       1.00       0.98      36798
     6           0.33       0.00       0.01          638

...
 accuracy                   0.96      93257
 macro avg                  0.46      93257
 weighted avg              0.94      93257
```

KESIMPULAN

Berdasarkan hasil evaluasi model mengenai prediksi kredit risk analisis menggunakan metode logistic regression, didapatkan hasil akurasi sebesar 96%. Dengan tingkat akurasi yang cukup tinggi maka model ini mempunyai potensi tinggi dalam membuat prediksi tersebut.



SARAN

- Menerapkan kebijakan kredit yang lebih ketat
- Mengawasi portofolio pinjaman dengan cermat dan melakukan analisis lebih mendalam pada peminjam dengan status buruk untuk mengurangi risiko non-performing loan (NPL).
- Menganalisis profil dan kondisi ekonomi peminjam
- Menilai dengan cermat nilai jaminan dan menghitung Loan-to-Value (LTV) untuk mengurangi risiko kerugian dalam kasus default.

Dengan pendekatan yang berfokus pada analisis data dan risiko, perusahaan dapat mengoptimalkan kinerja portofolio pinjaman dan mengurangi potensi risiko kredit.

