

Pengembangan Model Prediksi Risiko Kredit

Auditasari Damayanti putri





Outline

01

Business Understanding

02

Exploratory Data Analyst

03

Data Preprocessing

04

Machine Learning Modeling

05

Business Insight and Recommendation

Business Understanding



Problem

Bank meminjamkan dana kepada peminjam, namun 49,8 % peminjam mengalami kredit macet. Hal ini menyebabkan kerugian. Sehingga bank menginginkan model untuk memprediksi resiko kredit, untuk meminimalisir kerugian dimasa depan.

Goal

Mengurangi peminjam gagal bayar atau kredit macet.

Objective

- Mengembangkan model machine learning untuk memprediksi apakah peminjam memiliki resiko tinggi atau rendah
- Memberikan rekomendasi, berupa segmentasi peminjam yang berpotensi kredit macet atau kredit lancar.



Business Matrics



Mengurangi rasio kredit macet sebesar 10%

Data Overview

Dataset tersusun dari :

Total Baris : 466.285

Total kolom : 75 Kolom

Terdiri dari :

1 kolom target : loan_status

53 kolom numerik

20 kolom kategorik

Terdapat Kolom kosong sebanyak 17 kolom dan 20 kolom yang memiliki nilai null

Kolom yang digunakan :

'term','grade','sub_grade','home_ownership','annual_inc','dti',

'open_acc','revol_util','funded_amnt','emp_length',
'inq_last_6mths','loan_status','verification_status'

Fitur Target

Memiliki nilai sebagai berikut:

- Current
- Fully Paid
- Charged Off
- Late(31-120 days)
- In Grace Period
- Does not meet credit policy: Full Paid
- Does not meet credit policy: Charged off
- Late (16-30days)
- Defalut

Untuk meminimalkan, fitur target dikelompokkan menjadi

- High Risk (Charged Off, Late(31-120days),Late(16-30 days), In Grace Period,Default, Does meet credit policy:Charged Off)
- Low Risk (Current, Fully Paid, Does not meet credit policy: Fully Paid)

Exploratory Data Analyst



Univariet

Distribusi data

- * variabel dti dan funded_mnt berdistribusi normal

- * variabel annual_inc, open_acc, revol_util dan inq_last_6mths tidak berdistribusi normal dan memiliki skewed positif

- * Outlier

Variabel

annual_inc, revol_util, open_acc dan inq_last_6mths terindikasi terdapat outlier

Bivariet

1. Hubungan variabel numerik dengan numerik
Variabel - variabel numerik tidak memiliki hubungan yang signifikan

2. Hubungan variabel kategorik dengan kategorik

Variabel - variabel kategorik memiliki hubungan signifikan terhadap status pinjaman (loan_status)

3. Hubungan variabel numerik dengan kategorik (loan_status)

Variabel total pendapatan tahunan (annual_inc) dan rasio kredit bergulir (revol_util) tidak memiliki perbedaan yang signifikan terhadap status pinjaman (loan_status). Variabel jumlah pinjaman (funded_amnt) memiliki perbedaan yang signifikan terhadap status pinjaman (loan_status)

Bivariet

Korelasi Variabel – variabel numerik tidak ada di atas 0.7. Hal ini menandakan tidak adanya fitur yang redundan (berulang)



Data Pre- Processing



Data Cleanising

- Handle Missing Value
- Handle Outlier

Feature Enginerring

- 2 Feature Extraction
- 5 Fitur label Encoder
- 2 Fitur One Hot Encoding

Feature Transformation

- Normalization Standarscaler

Split Data

- Train : 296.632
- Test : 148.137



Machine Learning Model



Model Evaluation

F1-Score



Model
Metrics

Logistic Regression Decision Tree Random Forest XGBoost Adaboost

Precision

0.65 0.61 0.66 0.66 0.66

Recall

0.58 0.61 0.74 0.79 0.74

F1-Score

0.61 0.61 0.70 0.71 0.70

Roc_auc

0.70 0.63 0.77 0.78 0.77



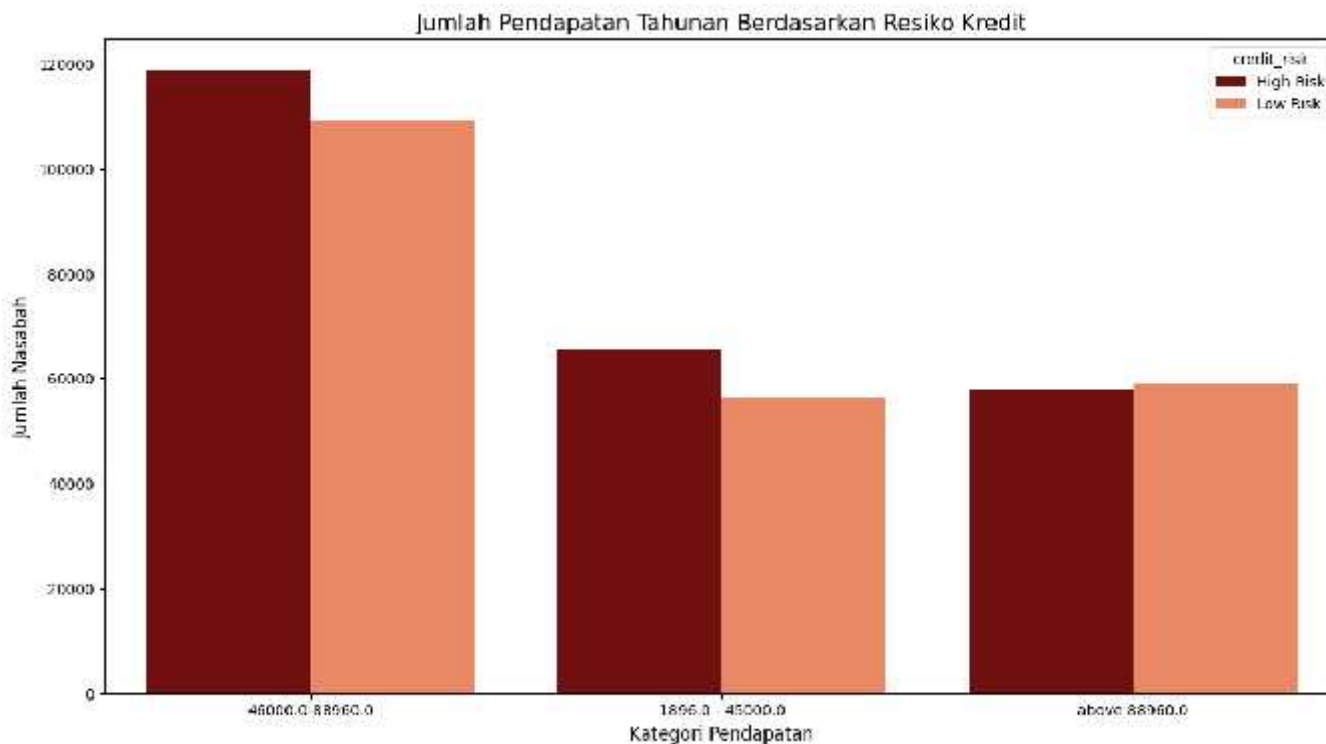
Hyperparameter Tuning

<div>Model Metrics</div>	Logistic Regression	XGBoost	Adaboost
Precision	0.65	0.66	0.66
Recall	0.58	0.77	0.76
F1-Score	0.61	0.71	0.71
Roc_auc	0.70	0.78	0.78

Data Insight & Recommendation



Jumlah Pendapatan Tahunan Berdasarkan Resiko Kredit

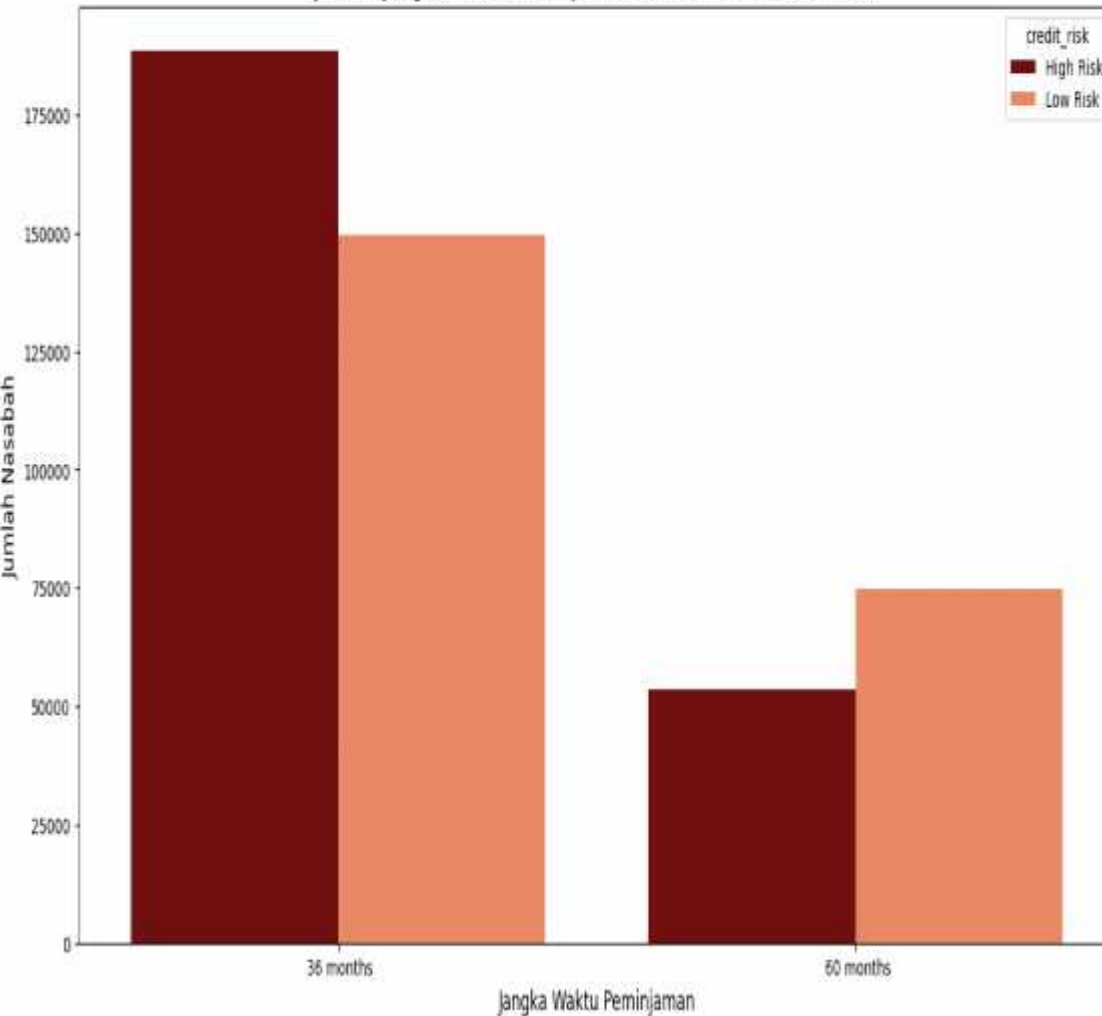


Peminjam dana dengan pendapatan tahunan dari 1896 sampai 88.960 memiliki resiko tertinggi atau kredit macet. Sedangkan Peminjam dengan pendapatan diatas 88.960 memiliki selisih yang rendah antara resiko tinggi dan resiko rendah.

Rekomendasi :
Bank dapat menerapkan penilaian risiko yang lebih ketat terhadap peminjam dengan pendapatan tahunan di bawah 88.960.



Jumlah Jangka Waktu Peminjaman Berdasarkan Resiko Kredit



Jangka Waktu Pinjaman

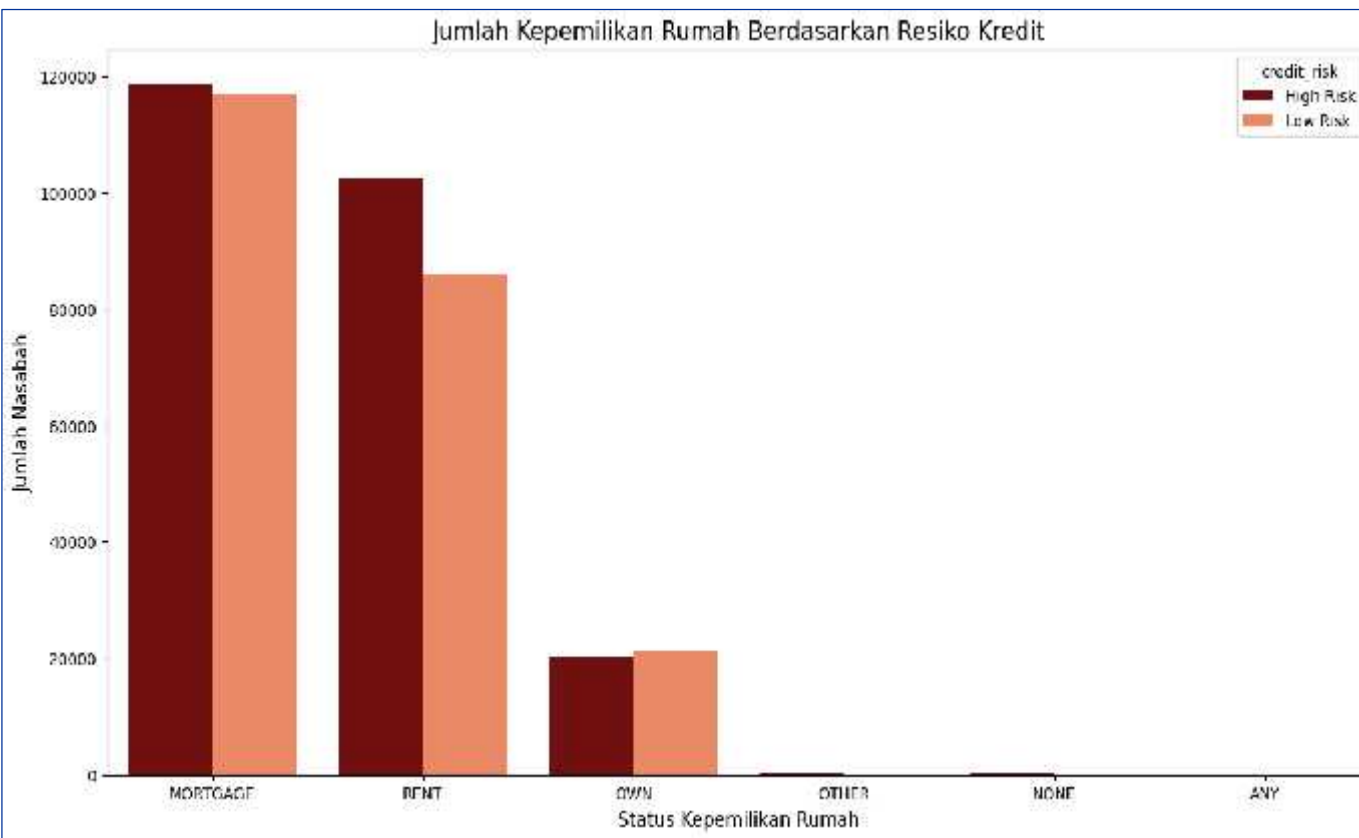
Berdasarkan Resiko Kredit

Peminjam dengan tenor waktu 60 bulan atau 5 tahun memiliki nilai tertinggi resiko kredit renda(Low Risk). Hal ini dapat di sebabkan kecilnya cicilan yang di bebaskan oleh peminjam.

Rekomendasi :

Perusahaan dapat merekomendasikan peminjam untuk mengambil pinjaman dengan waktu tenor 60 bulan atau 5 tahun.

Jumlah Kepemilikan Rumah Berdasarkan Resiko Kredit



Peminjam dana dengan kepemilikan yang tinggi pada resiko tinggi atau kredit macet. Namun secara selisih kepemilikan rumah sewa memiliki selisih antara resiko tinggi dan rendah yang tinggi

Rekomendasi :
Perusahaan dapat memberikan penilaian lebih lanjut apakah peminjam dengan kepemilikan rumah dan sewa dapat membayar pinjaman secara utuh.



Simulasi Business

Before Machine
Learning Model



After Machine
Learning Model

Jumlah Nasabah Yang Terindikasi
High Risk Sebelum Model:
77.682

Jumlah Nasabah Yang
Terindikasi High Risk Setelah
Model: 68.915

 10 %



Terima Kasih