

Optimasi Strategi Lending melalui Prediksi Risiko Kredit Berbasis Machine Learning

Data Scientist Project-Based Internship

ID/X Partners



• Latar Belakang & Masalah

The Problem

- **High Credit Risk** : Tingginya angka gagal bayar (Bad Loan) yang mengancam stabilitas finansial perusahaan.
- **Data Imbalance** : Data nasabah macet jauh lebih sedikit dibanding nasabah lancar, sehingga model tradisional sulit mengenali pola risiko.

The Objective

- **Predictive System** : Membangun sistem klasifikasi otomatis yang mampu mendekripsi potensi gagal bayar sejak dini.
- **Actionable Insights** : Memberikan rekomendasi strategis bagi tim manajemen untuk mitigasi risiko kredit secara terukur.





Data preparation Data understanding

Di tahap ini, data mentah dilakukan proses "pembersihan" dari missing values dan outliers, sehingga data mentah diubah menjadi format yang siap diproses oleh algoritma Machine Learning.

Target Labeling (Loan Status)

- 0 : Good Loan
- 1 : Bad Loan

Data Transformation

```
## Rows: 465,945
## Columns: 14
## $ loan_status <fct> 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0...
## $ loan_amnt <int> 5000, 2500, 2400, 10000, 3000, 5000, 7000, 3000, 5600, ...
## $ term <dbl> 36, 60, 36, 36, 60, 36, 60, 60, 36, 36, 36, 36, ...
## $ int_rate <dbl> 10.65, 15.27, 15.96, 13.49, 12.69, 7.90, 15.96, 18.64, ...
## $ sub_grade <fct> B2, C4, C5, C1, B5, A4, C5, E1, F2, B5, C3, B5, C1, B1, ...
## $ annual_inc <dbl> 24000.00, 30000.00, 12252.00, 49200.00, 80000.00, 36000...
## $ dti <dbl> 27.65, 1.00, 8.72, 20.00, 17.94, 11.20, 23.51, 5.35, 5...
## $ emp_length <dbl> 10, 0, 10, 10, 1, 3, 8, 9, 4, 0, 5, 10, 0, 3, 3, 0, 4, ...
## $ home_ownership <fct> RENT, RENT, RENT, RENT, RENT, RENT, RENT, OWN, RE...
## $ total_acc <dbl> 9, 4, 10, 37, 38, 12, 11, 4, 13, 3, 23, 34, 9, 11, 29, ...
## $ delinq_2yrs <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2...
## $ inq_last_6mths <dbl> 1, 5, 2, 1, 0, 3, 1, 2, 2, 0, 2, 0, 1, 2, 2, 1, 2, 0, ...
## $ revol_util <dbl> 83.70, 9.40, 98.50, 21.00, 53.90, 28.30, 85.60, 87.50, ...
## $ purpose <fct> credit_card, car, small_business, other, other, wedding...
```

→ Variabel Y

Variabel X

• Metodologi & Workflow End-to-End Solution

Exploratory Data Analysis (EDA)

Statistika Deskriptif

```
## # loan_status loan_amnt term int_rate sub_grade
## 0:415035 Min. : 500 Min. :36.00 Min. : 5.42 B3 : 31669
## 1: 50910 1st Qu.: 8000 1st Qu.:36.00 1st Qu.:10.99 B4 : 30484
## Median :12000 Median :36.00 Median :13.66 C1 : 26936
## Mean :14320 Mean :42.61 Mean :13.83 C2 : 26720
## 3rd Qu.:20000 3rd Qu.:60.00 3rd Qu.:16.49 B2 : 26600
## Max. :35000 Max. :60.00 Max. :26.06 C3 : 25301
```

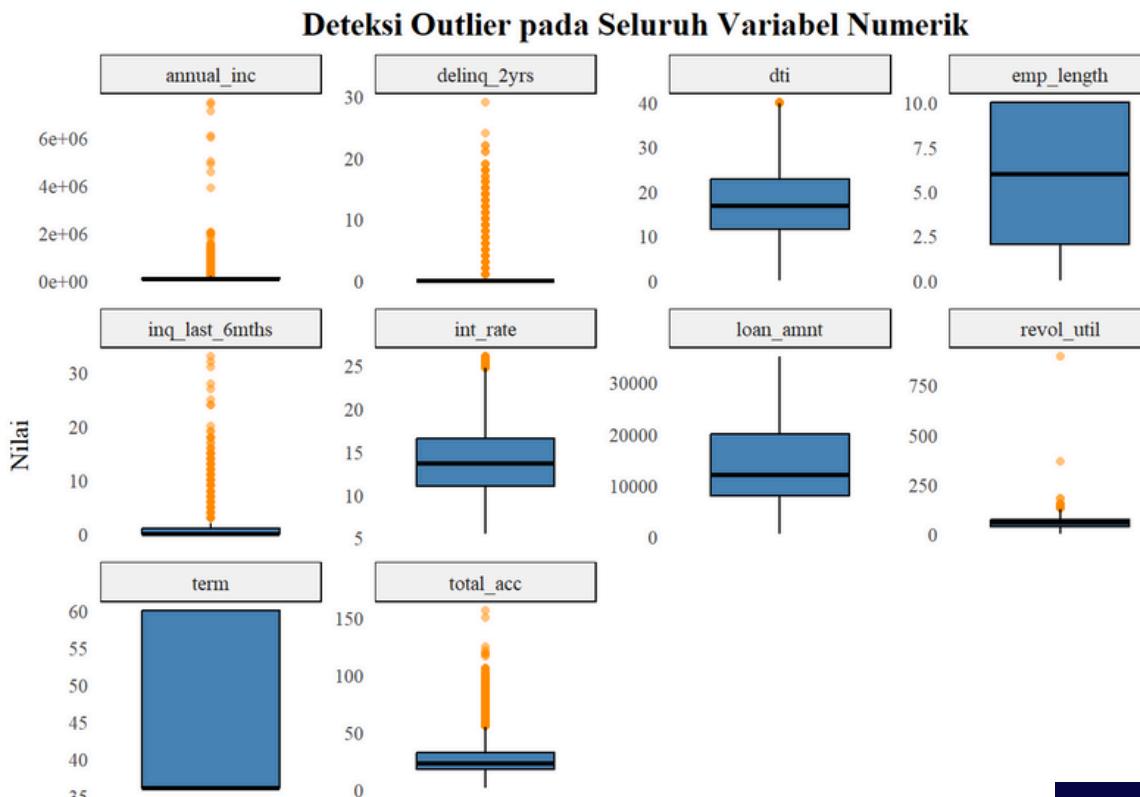
```
## # annual_inc dti emp_length home_ownership
## Min. : 1896 Min. : 0.00 Min. : 0.000 ANY : 1
## 1st Qu.: 45000 1st Qu.:11.37 1st Qu.: 2.000 MORTGAGE:235715 ##
## Median : 63000 Median :16.87 Median : 6.000 NONE : 46 ##
## Mean : 73269 Mean :17.22 Mean : 5.724 OTHER : 180 ##
## 3rd Qu.: 88863 3rd Qu.:22.78 3rd Qu.:10.000 OWN : 41667 ##
## Max. :7500000 Max. :39.99 Max. :10.000 RENT :188336 ##
## total_acc delinq_2yrs inq_last_6mths revol_util
## Min. : 1.00 Min. : 0.0000 Min. : 0.0000 Min. : 0.00
## 1st Qu.: 17.00 1st Qu.: 0.0000 1st Qu.: 0.0000 1st Qu.: 39.20
## Median : 23.00 Median : 0.0000 Median : 0.0000 Median : 57.60
## Mean : 25.07 Mean : 0.2846 Mean : 0.8048 Mean : 56.18
## 3rd Qu.: 32.00 3rd Qu.: 0.0000 3rd Qu.: 1.0000 3rd Qu.: 74.70
## Max. :156.00 Max. :29.0000 Max. :33.0000 Max. :892.30
```



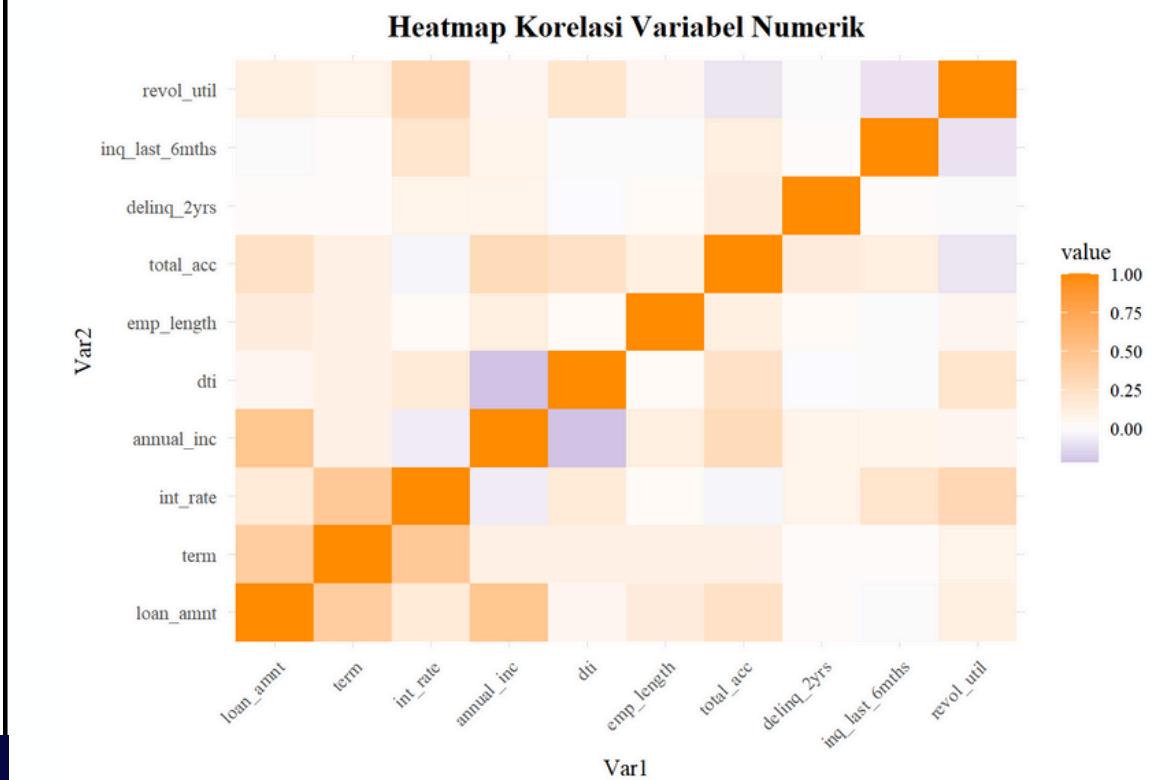
Exploratory Data Analysis (EDA)

Deteksi Outlier & Penanganan

Visualisasi menunjukkan adanya nilai ekstrem pada variabel finansial nasabah. Tindakan yang dapat dilakukan adalah menerapkan teknik **Capping (Winsorizing)** pada batas persentil ke-99. Langkah ini bertujuan untuk menjinakkan nilai ekstrem agar tidak merusak performa model, namun tetap mempertahankan data tersebut tanpa harus menghapusnya.



Korelasi antar Variabel Prediktor

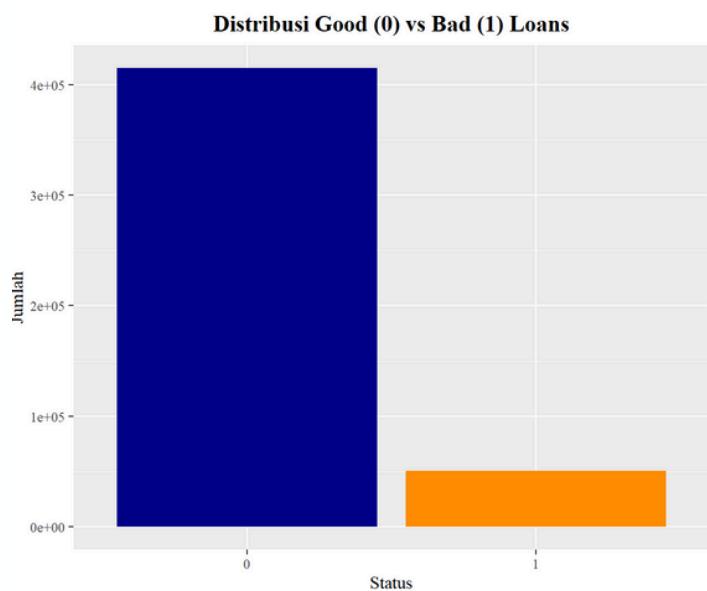


Heatmap menunjukkan bahwa terdapat korelasi positif kuat antara loan_amnt dengan term dan annual_inc, serta tidak terdapat variabel berbeda yang memiliki korelasi ekstrem (mendekati 1.0)

• Metodologi & Workflow End-to-End Solution

Handling Data Imbalance

Menerapkan teknik **SMOTE** untuk menyeimbangkan proporsi data Good dan Bad Loan agar model lebih sensitif terhadap risiko.

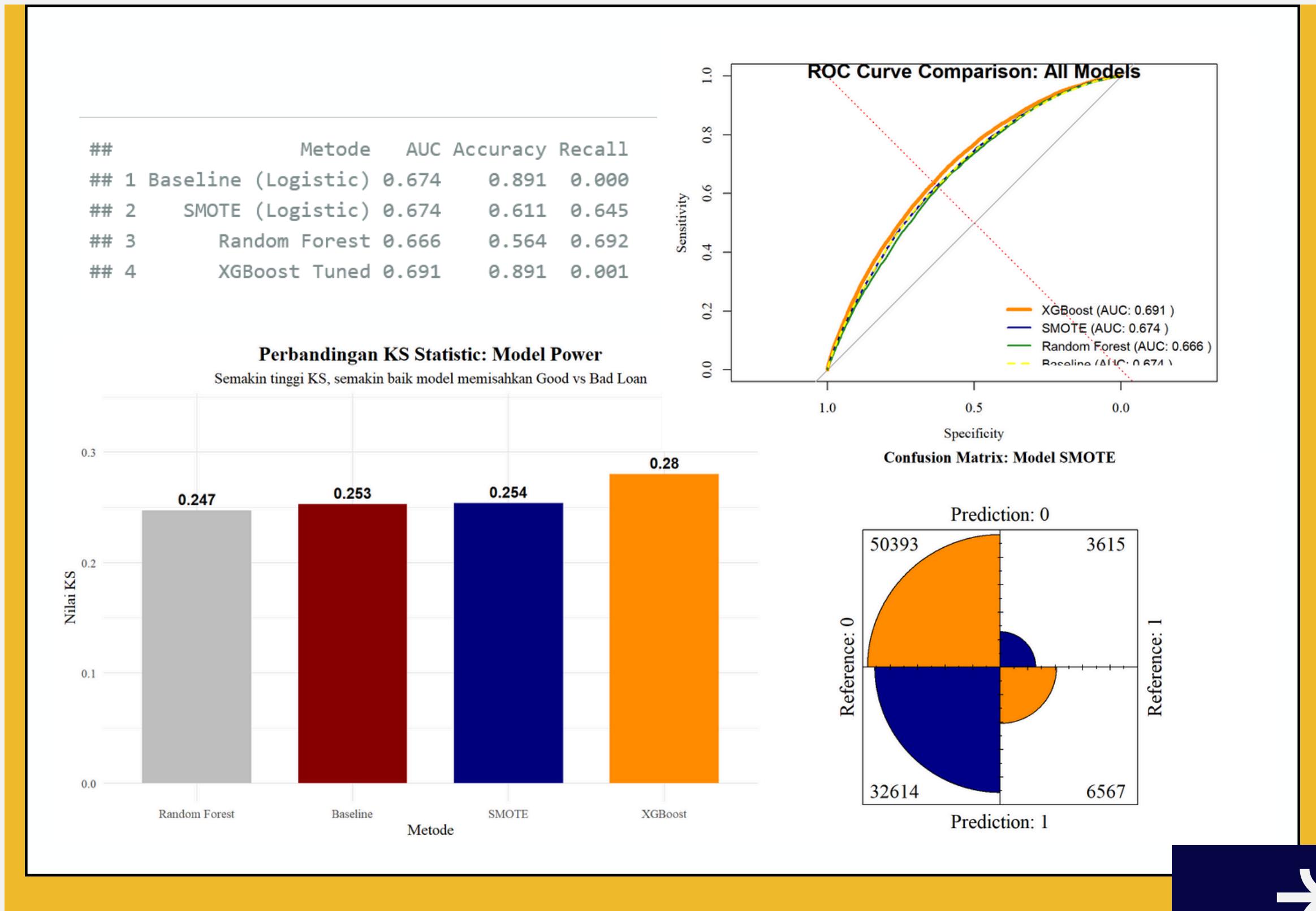


Model Development & Evaluation

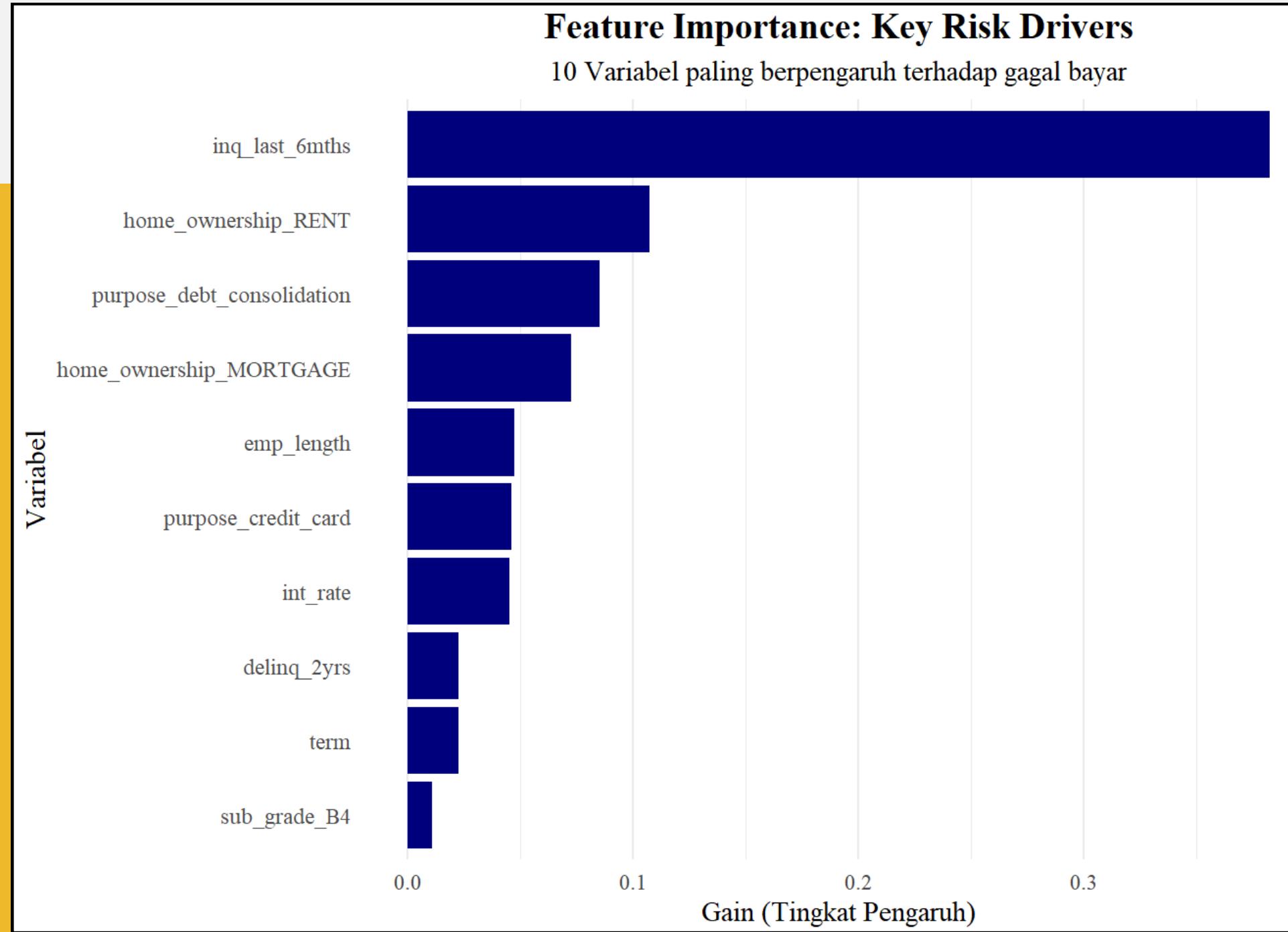
Mengembangkan beberapa algoritma pembanding, yaitu **Logistic Regression (Baseline)**, **Logistic Regression with SMOTE**, **Random Forest**, dan **XGBoost (Advanced)**. Setelah itu dilakukan evaluasi model untuk menguji performa model menggunakan **metrik akurasi, AUC, dan KS Statistic** untuk menjamin ketajaman prediksi.

• Model Performance Comparison

Berdasarkan evaluasi yang dilakukan, diperoleh keputusan bahwa model terbaik yang dapat digunakan adalah **XGBoost**. Keputusan memilih XGBoost didasarkan pada performanya yang superior dalam metrik AUC dan KS Statistic, yang menunjukkan daya pisah risiko paling tajam. Selain itu, kemampuan algoritma ini dalam menangani hubungan non-linear antar variabel finansial memberikan akurasi yang lebih stabil dibandingkan model linear konvensional.



• Identifikasi Key Risk Drivers



Berdasarkan grafik Feature Importance yang dihasilkan, kita dapat mengidentifikasi variabel yang paling krusial dalam menentukan risiko nasabah:



- **inq_last_6mths**

Variabel ini menempati posisi teratas dengan nilai Gain tertinggi (mendekati 0.4). Ini menunjukkan bahwa perilaku nasabah dalam mengajukan pinjaman dalam 6 bulan terakhir adalah indikator terkuat untuk memprediksi risiko gagal bayar.

- **home_ownership_RENT**

Status tempat tinggal dengan cara menyewa menjadi faktor berpengaruh kedua. Hal ini mengindikasikan adanya korelasi antara stabilitas aset tempat tinggal dengan kepatuhan pembayaran kredit.

- **purpose_debt_consolidation**

Tujuan pinjaman untuk konsolidasi utang juga muncul sebagai salah satu pendorong utama risiko yang dipertimbangkan oleh model.

Data menunjukkan bahwa perilaku nasabah (**jumlah inquiry**) dan stabilitas ekonomi (**status rumah**) jauh lebih menentukan risiko gagal bayar dibandingkan variabel pendukung lainnya.

• Rekomendasi Strategis Kebijakan

Untuk memitigasi risiko kerugian dan menjaga kesehatan portofolio kredit, kami merekomendasikan:

- **Pengetatan Aturan Inquiry**

Melakukan pengecekan ekstra atau memberikan penalti skor bagi calon nasabah yang memiliki jumlah aplikasi kredit (inquiry) tinggi dalam 6 bulan terakhir.

- **Diferensiasi Plafon & Limit**

Menerapkan kebijakan limit kredit yang lebih konservatif bagi nasabah dengan status tinggal sewa guna menjaga stabilitas arus kas perusahaan.

- **Skrining Tujuan Pinjaman**

Melakukan audit pendapatan yang lebih ketat bagi nasabah yang mengajukan pinjaman dengan tujuan konsolidasi utang atau pembayaran kartu kredit.

- **Digitalisasi Scoring**

Mengintegrasikan model XGBoost ini ke dalam sistem persetujuan otomatis (Auto-Approval) untuk meningkatkan efisiensi operasional pada aplikasi kredit yang berisiko rendah.





Thankyou!



**In God we trust, all
others must bring data.**

Contact Me!



+62 853 3547 8546



fatikhanailans@gmail.com



<https://www.linkedin.com/in/fatikhanailans/>



@fatikhanailans_