



Credit Card Payment Default Prediction.

Cristanto
Steven Benny
Tri Setiawan
Ulva Dewiyanti

DataRider

Latar Belakang

HIGH DEFAULT RATE

Jumlah customer pengguna layanan kartu kredit pada September 2005 sebanyak 21000 customer. Namun, 23% dari customer tersebut mengalami gagal bayar (*default*) pada September 2005.

Sementara itu berdasarkan data FRED, default rate global dari tahun 1990 hingga 2005 berada dalam range 3% hingga 6,5%.



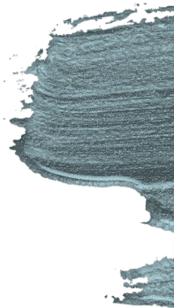
* Federal Reserve Economic Data (FRED) adalah database yang dikelola oleh divisi Riset Federal Reserve Bank of St. Louis yang memiliki lebih dari 816.000 deret waktu ekonomi dari berbagai sumber yang mencakup perbankan, bisnis/fiskal, indeks harga konsumen, nilai tukar, produk domestik bruto, suku bunga, dsb.



Lingkup Kerja

Lingkup kerja dari project terdiri dari,





23

kolom



Customer ID

Limit Balance

Sex

Education

Marriage Status

Age

Payment Status September

Payment Status Agustus

Payment Status Juli

Payment Status Juni

Payment Status Mei

Payment Status April

Bill Amount September

Bill Amount Agustus

Bill Amount Juli

Bill Amount Juni

Bill Amount Mei

Bill Amount April

Pay Amount September

Pay Amount Agustus

Pay Amount Juli

Pay Amount Juni

Pay Amount Mei

Pay Amount April

Default Status Next Month

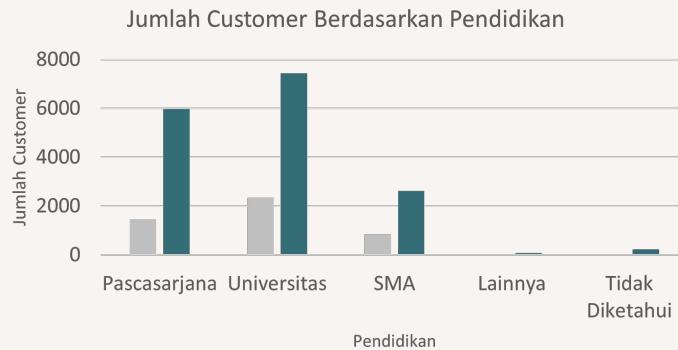


Informasi Dataset

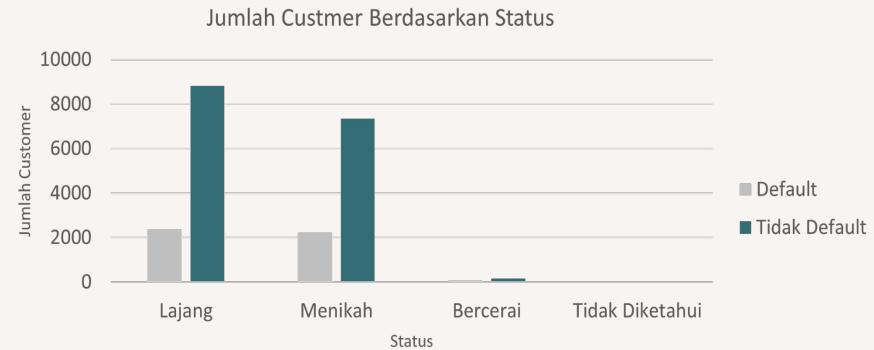
21000 BARIS & 25 KOLOM

- 01** Data terdiri dari data demografis, riwayat pembayaran April-September, tagihan April-September, status pembayaran April-September, dan status default pada September
- 02** 10 fitur kategorik yaitu Sex, Education, Marriage Status, Payment Status September, Payment Status Agustus, Payment Status Juli, Payment Status Juni, Payment Status Mei, Payment Status April, dan Default Status Next Month
- 03** 15 fitur numerik yaitu Customer ID, Limit Balance, Age, Bill Amount September, Bill Amount Agustus, Bill Amount Juli, Bill Amount Juni, Bill Amount Mei, Bill Amount April, Pay Amount September, Pay Amount Agustus, Pay Amount Juli, Pay Amount Juni, Pay Amount Mei, Pay Amount April
- 04** Default Status Next Month adalah fitur target

Dominasi customer pada data



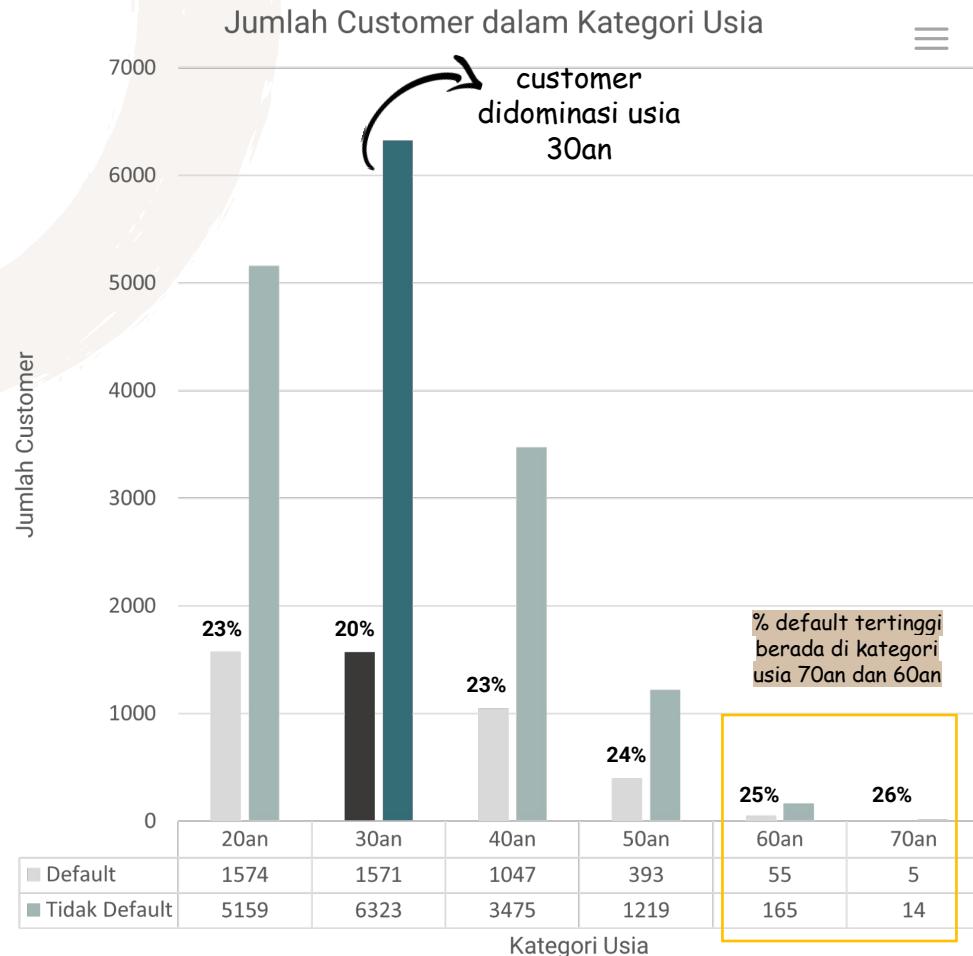
- Data didominasi oleh jenis kelamin perempuan dengan persentase sebesar 61%, namun default didominasi oleh laki-laki dengan persentase sebesar 24%
- Pendidikan didominasi oleh universitas sebanyak 47% dan diikuti oleh pascasarjana sebanyak 35%, namun default didominasi oleh universitas dan SMA sebesar 24%
- Status pernikahan didominasi oleh lajang sebanyak 53%, namun default didominasi oleh status bercerai sebesar 29%



Dominasi customer dalam usia

Customer didominasi oleh usia 30an yaitu usia antara 30 tahun – 39 tahun yaitu sebanyak 37,6% dari total customer yang ada, disusul oleh usia 20an dan kemudian 40an.

Namun, jika dilihat dari persentase default, terlihat bahwa customer dengan usia lansia yaitu 70an dan 60an memiliki nilai yang lebih besar jika dibandingkan dengan customer pada kategori usia lainnya.



Customer default memiliki *buying power* rendah

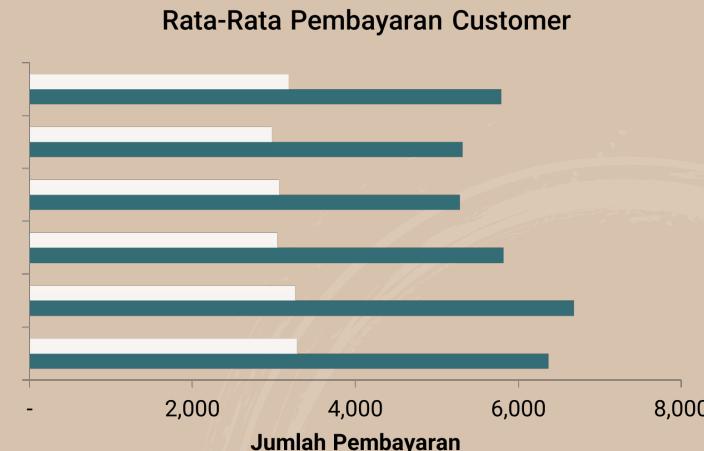
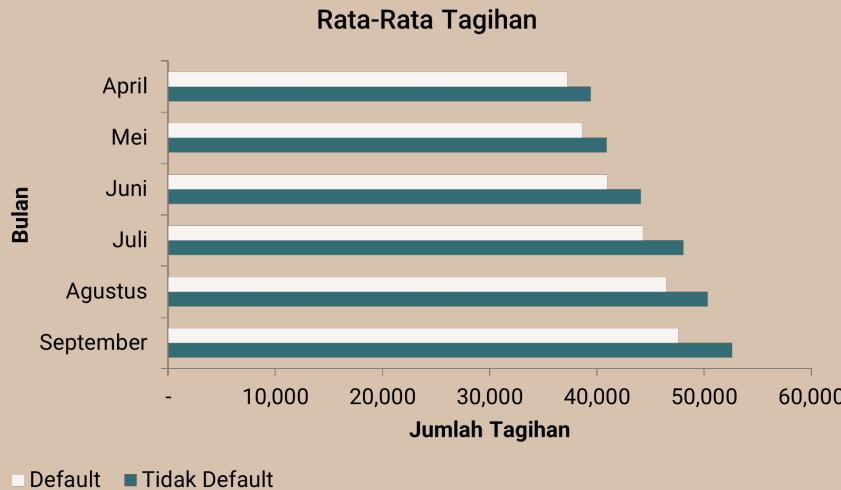
Customer default dan tidak default memiliki rata-rata limit balance yang sesuai dengan kecenderungan default customer, dimana customer default diberikan limit yang lebih rendah. Namun jika dilihat dari *bill amount/tagihannya*, terlihat bahwa rata-rata tagihan customer default dan tidak hampir sama nilainya. Sementara itu, rata-rata *pay amount/pembayaran* terlihat jauh berbeda dimana kemampuan bayar customer default hanya $\frac{1}{2}$ kali dari customer yang tidak default.

178154

Rata-rata limit balance
customer tidak default

128699

Rata-rata limit balance
customer default



Preprocessing

Pada tahap ini dilakukan beberapa proses terhadap dataset sehingga dataset menjadi bersih ketika digunakan untuk melatih model. Berikut alur preprocessing yang dilakukan secara general sebelum proses modelling dilakukan,



Eksperimen Model

Dalam tahap pemodelan dilakukan beberapa eksperimen terhadap dataset, algoritma model, dan fitur-fitur yang digunakan dalam model.

1 Dataset

- dataset original
- dataset setelah outlier dihapus
- dataset yang distandarisasi
- dataset setelah outlier dihapus, distandarisasi, dan adanya one hot encoding

2 Algoritma Model

- Logistic Regression
- KNN
- Decision Tree
- Random Forest
- AdaBoost
- XGBoost

4 Nilai metrik evaluasi dengan AUC terbesar

Algoritma	Accuracy	Precision	Recall	AUC	F1
XGBoost	0.82	0.67	0.37	0.77	0.48
	0.81	0.67	0.34	0.77	0.46
	0.82	0.66	0.39	0.78	0.49
AdaBoost	0.81	0.67	0.34	0.77	0.45

Best model yaitu **XGBoost** dan telah di oversampling, dimana model tersebut memiliki nilai AUC dan Recall tertinggi.

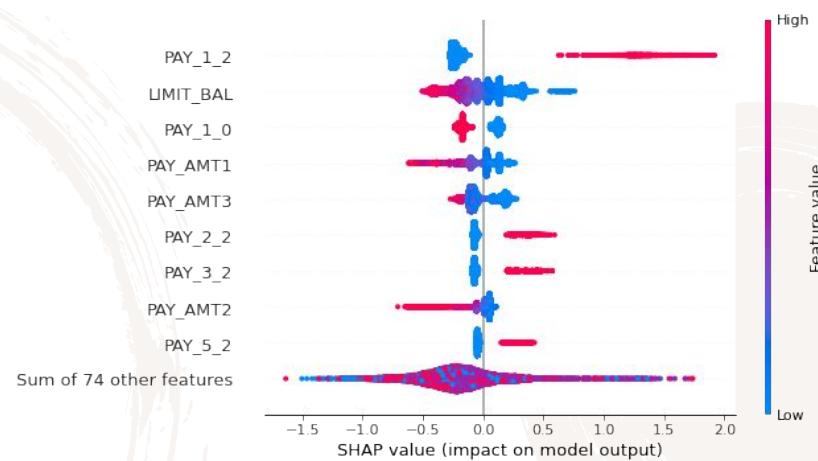
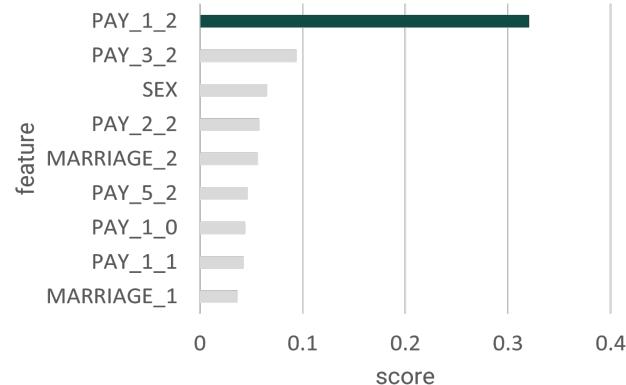
Features Importance

DARI BEST MODEL

9 Features importance tertinggi dari model terbaik berdasarkan SHAP value dan feature importance score didominasi fitur status pembayaran, sehingga dapat disimpulkan bahwa **status pembayaran merupakan fitur yang berkontribusi cukup besar dalam menentukan customer akan default atau tidak di bulan berikutnya**.

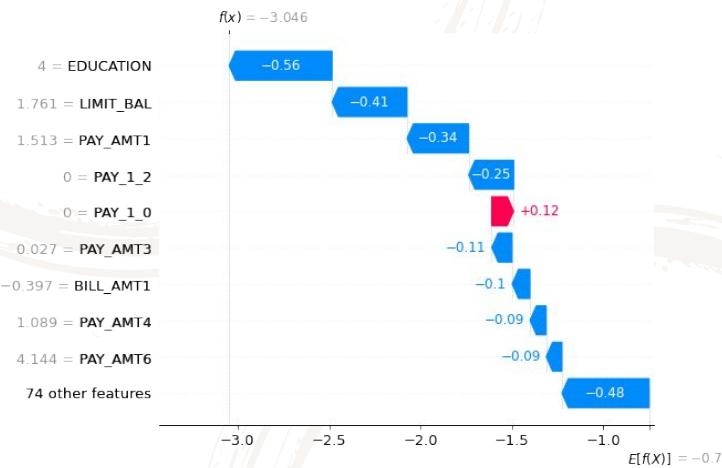
Feature importance tertinggi pada grafik feature importance score yaitu **PAY_1_2** yang merupakan status pembayaran di bulan September, dimana customer gagal bayar selama 2 bulan berturut-turut. Fitur ini nantinya **dapat digunakan dalam penentuan customer prioritas yang akan dimitigasi** untuk menurunkan default rate pada bulan berikutnya.

Top 9 Feature Importance Score of Best Model

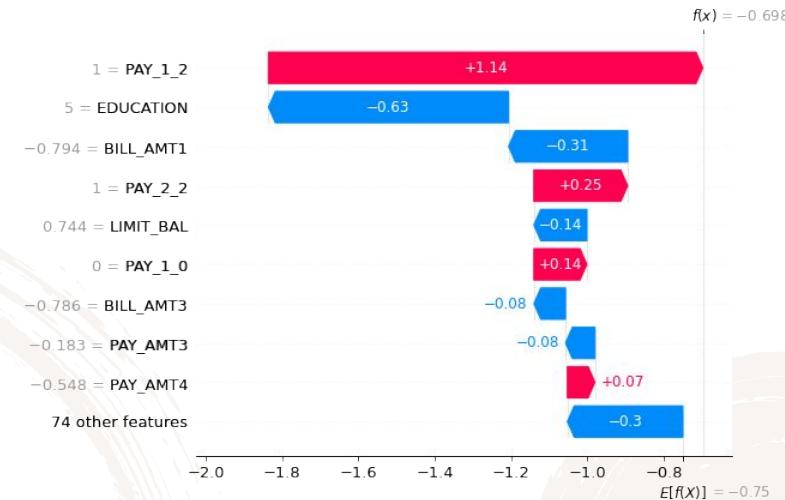


Contoh SHAP Value

FITUR PENTING MODEL



Grafik di atas adalah salah satu contoh SHAP value untuk customer yang **tidak default**.



Grafik di atas adalah salah satu contoh SHAP value untuk customer yang **default**, terlihat bahwa PAY_1_2 memberikan kontribusi yang besar dalam penentuan customer tersebut default.



4830 customer (23% dari total)
default pada September 2005

jika model diaplikasikan

1884 customer (39% dari total
default) berhasil diprediksi default
dengan tepat pada bulan berikutnya

jika semua dimitigasi

Jika 1884 customer
berhasil dimitigasi, maka
default rate menjadi

$$\frac{2946}{21000} * 100\% =$$

14%

jika hanya beberapa customer
dimitigasi

asumsikan 1000 customer
default yang berhasil diprediksi
dan dengan status pembayaran
di September adalah telat 2
bulan saja yang dimitigasi,
maka default rate menjadi

$$\frac{3830}{21000} * 100\% =$$

18%

Simulasi & Strategi

Berikut adalah strategi dan simulasi terhadap
rekomendasi yang diberikan,

21000
TOTAL
CUSTOMER

39%
RECALL
MODEL

Rekomendasi Bisnis

1. Customer default dapat diberikan penawaran berupa,
 - Mengubah jadwal pembayaran seperti perubahan jangka waktu kredit dan masa tenggang
 - Mengubah struktur pembiayaan seperti besar tagihan, bunga pinjaman, besar limit balance, dsb
2. Proses mitigasi customer default dapat dimulai dari customer yang memiliki status pembayaran yang terlambat selama 2 bulan pada September 2005



Thank You

DataRider