

CREDIT CARD PAYMENT DEFAULT PREDICTION

Supported by:
Rakamin Academy
Career Acceleration School
www.rakamin.com

INFINITY



Alaska Napitupulu



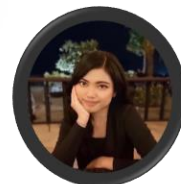
Rafsanjani Lestari Negara



Ike Setyawati



Ukhti Aprillia Vanny



Ni Made Sari Anggraini

STAGE 0

CREDIT CARD PAYMENT DEFAULT PREDICTION

Problem Statement

Bank Taiwan dibulan April hingga September 2005 memiliki 21.000 customer, dimana **22,12%** customer tersebut berstatus **default**

Our Goals

- Mendeteksi adanya kemungkinan cardholder yang akan gagal bayar
- Menurunkan tingkat default
- Membantu bank untuk tetap mendapatkan profit.

Our Objective

- Membuat sistem untuk memprediksi cardholder yang berpotensi gagal bayar
- Membuat skema keringanan pembayaran kredit supaya cardholder tetap melakukan pembayaran sebelum jatuh tempo

Business Metrics

Payment Default Rate
(Hal ini akan berdampak baik yaitu untuk mencegah pemberian credit untuk yang terdeteksi akan default / mengurangi cost)

1. Descriptive Statistics

Gunakan function info dan describe pada dataset final project kalian. Tuliskan hasil observasinya, seperti:

- A. Apakah ada kolom dengan tipe data kurang sesuai, atau nama kolom dan isinya kurang sesuai?
- B. Apakah ada kolom yang memiliki nilai kosong? Jika ada, apa saja?
- C. Apakah ada kolom yang memiliki nilai summary agak aneh? (min/ mean/ median/ max/ unique/ top/ freq)

JAWABAN

- A. semua fitur memiliki tipe kolom yang sesuai begitu juga isinya tidak ada yang aneh, semua sesuai dengan tipe yaitu int hal ini dikarenakan beberapa fitur sudah label encoder terlebih dahulu dari sumber datanya hanya saja jumlah unqiue dengan deskripsi data tidak sesuai dimana ada beberapa fitur yang jumlah uniqenya lebih seperti Education.
- B. untuk missing value sendiri tampak jelas bahwa tidak ada satupun fitur yang memiliki missing value
- C. LIMIT_BAL, PAY_AMT1 - PAY_AMT6 memiliki nilai mean > median dan selisih percentil 75 dengan max sangat jauh, BILL_AMT1 - BILL_AMT6 memiliki nilai mean > median. Uniqe yang berbeda seperti di EDUCATION dan MARRIAGE

STAGE 1

**CREDIT CARD
PAYMENT DEFAULT
PREDICTION**

2. Univariate Analysis

Gunakan visualisasi untuk melihat distribusi masing-masing kolom (feature maupun target). Tuliskan hasil observasinya, misalnya jika ada suatu kolom yang distribusinya menarik (misal skewed, bimodal, ada outlier, ada nilai yang mendominasi, kategorinya terlalu banyak, dsb). Jelaskan juga apa yang harus di-follow up saat data pre-processing.

JAWABAN

- A. SKEW = semua numerical data pada dataset tampak skew ke kanan semua seperti LIMIT_BAL, BILL_AMT1 - BILL_AMT6, PAY_AMT1 - PAY_AMT6 right skewed
- B. BIMODAL = tidak ada
- C. OUTLIERS = berdasarkan hasil boxplot ada beberapa yang Nampak outliers seperti 'limit_bal', 'age', 'bill_amt1', 'bill_amt2', 'bill_amt3', 'bill_amt4', 'bill_amt5', 'bill_amt6', 'pay_amt1', 'pay_amt2', 'pay_amt3', 'pay_amt4', 'pay_amt5', 'pay_amt6'
- D. KATEGORIK = Data telah dalam bentuk Label Encoding sehingga perlu dibenahi karena ada beberapa fitur yang masuk ke dalam one hot encoding

STAGE 1

**CREDIT CARD
PAYMENT DEFAULT
PREDICTION**

Exploratory Data Analysis

3. Multivariate Analysis

Lakukan multivariate analysis (seperti correlation heatmap dan category plots, sesuai yang diajarkan di kelas). Tuliskan hasil observasinya, seperti:

- A. Bagaimana korelasi antara masing-masing feature dan label. Kira-kira feature mana saja yang paling relevan dan harus dipertahankan?
- B. Bagaimana korelasi antar-feature, apakah ada pola yang menarik? Apa yang perlu dilakukan terhadap feature itu?

JAWABAN

- A. Korelasi features dengan target default_payment_next_month (memiliki korelasi positif lemah dengan LIMIT_BAL, Pay_0, Pay_1, Pay_2, Pay_3, Pay_4, Pay_5, Pay_6 yang selanjutnya akan digunakan pada model. Namun, memiliki korelasi positif cukup kuat terhadap feature apapun.
- B. Terdapat banyak sekali feature yang saling berkorelasi positif cukup kuat (BILL_AMT dengan BILL_AMT dan PAY_AMT dengan PAY_AMT) dan ada juga feature yang berkorelasi negatif lemah (AGE - MARRIAGE)

STAGE 1

**CREDIT CARD
PAYMENT DEFAULT
PREDICTION**

Exploratory Data Analysis

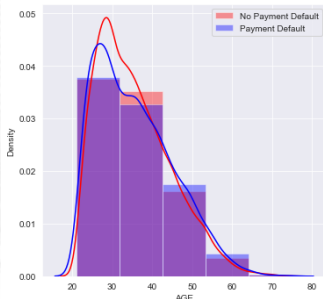
4. Business Insight

Selain EDA, lakukan juga beberapa analisis dan visualisasi untuk menemukan suatu business insight. Tuliskan minimal 3 insight, dan berdasarkan insight tersebut jelaskan rekomendasinya untuk bisnis.

JAWABAN

Setelah kami melakukan EDA ada beberapa insight yang kami menemukan

default_payment_next_month			
SEX	EDUCATION	MARRIAGE	
1	University	Married	433
		Single	535
2	Graduate_School	Single	540
		University	628
		Married	716



Dari data set yang dimiliki, kami melakukan filter agregasi untuk data default dan mencoba melihat jumlah terbanyak nasabah default, dan kami menemukan seperti gambar diatas. Hal ini bisa dijadikan acuan awal untuk **merekomendasikan nasabah baru** akan kemungkinan lebih besar default atau sebaliknya.

STAGE 1

**CREDIT CARD
PAYMENT DEFAULT
PREDICTION**

Exploratory Data Analysis

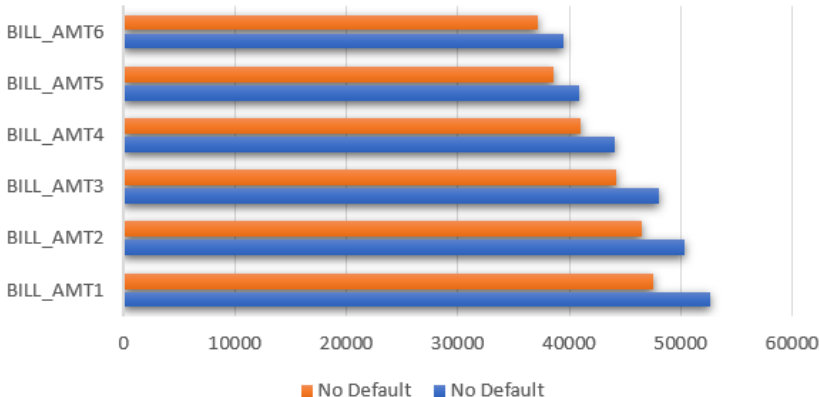
128699.177610

**Rata-rata limit balance
Customer Default**

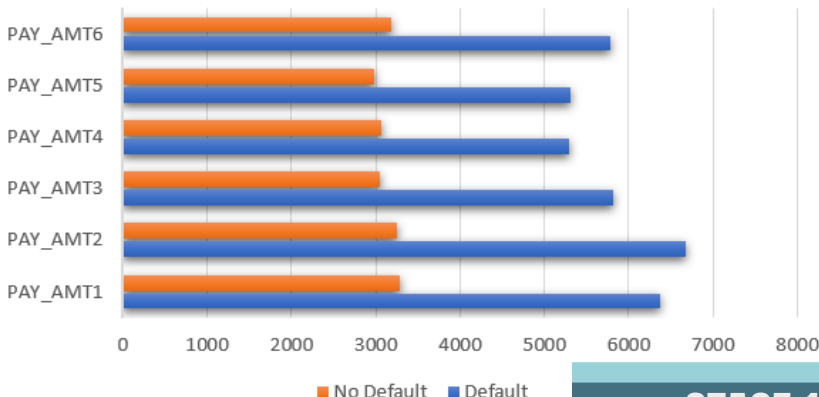
178153.592174

**Rata-rata limit balance
Customer No Default**

RATA-RATA TAGIHAN



RATA-RATA PEMBAYARAN



Customer nasabah Bank Taiwan antara default dan tidak default memiliki rata-rata limit balance dimana customer default yang lebih rendah. Namun jika dilihat dari bill amount/tagihan nasabah rata-rata tagihan customer default dan tidak default hampir sama nilainya. Sementara itu, rata-rata pay amount/pembayaran terlihat jauh berbeda.

Kemampuan bayar customer default rata-rata hanya 1/2 kali dari customer yang tidak default

STAGE 1

**CREDIT CARD
PAYMENT DEFAULT
PREDICTION**

REKOMENDASI BISNIS

1. Memberikan penawaran khusus untuk cardholder yang berpotensi default di bulan depan yaitu dengan diberikan penawaran restrukturisasi baik perpajakan tenor maupun pengurangan angsuran. Upaya ini dilakukan supaya dapat membantu mengurangi beban dari cardholder sehingga dapat membayarkan kewajibannya setiap bulan.
2. Mereminder cardholder sebelum tanggal jatuh tempo agar cardholder melakukan pembayaran pinjaman tepat waktu supaya dapat menahan atau menurunkan tingkat default dibulan berikutnya
3. Bagi cardholder yang melakukan pembayaran tepat waktu selama 1 tahun, bank akan memberikan apreasi berupa rekomendasi kenaikan limit balance dan reaktivasi credit cardnya

STAGE 1

**CREDIT CARD
PAYMENT DEFAULT
PREDICTION**

REKOMENDASI PREPROCESSING

1. Kita akan merename PAY_0 menjadi PAY_1
2. Kolom ID merupakan identifier dari tiap baris yang nilainya unik dan tidak dapat memberikan informasi apa-apa dalam analisis.
3. Terdapat nilai-nilai yang belum terdefinisi pada kolom EDUCATION, PAY_0, PAY_2, PAY_3, PAY_4, PAY_5, PAY_6, MARRIAGE (akan digabung unknown)
4. Terdapat nilai minus pada nilai min untuk kolom BILL_AMT1 - BILL_AMT6, dimana seharusnya untuk jumlah bill statement tidak ada yang bernilai negatif (akan di drop) ATAU Perlu eksplorasi lebih dalam mengenai BILL_AMT yang minus ini karena jika didrop akan mengurangi jumlah data cukup banyak, kemungkinan BILL_AMT bernilai negatif karena kelebihan bayar tagihan pada bulan sebelumnya.
5. Tampak disetiap numerikal data pada dataset memiliki outlier semua pada visualisasi Boxplot, untuk penanganan yang dilakukan yaitu dengan memfilter outlier dengan menggunakan z score, kemudian akan dilakukan pengecekan distribusi dari data setelah difilter dengan z score atau menggunakan IQR (akan dilakukan pengamatan pebandingan hasil akurasi nantinya).
6. Berdasarkan visualisasi KernelDensityEstimationplot, semua numerical data pada dataset tampak skew ke kanan semua, Penanganannya mungkin akan dilakukan log transformation untuk mengatasi distribusi yang skew tersebut setelah dilakukan (filtrasi outlier)
7. Kita akan memilih salah satu fitur baik PAY atau BILLAMT dan PAYAMT karena kedua jenis fitur sebab akibat sehingga bisa dipilih salah satunya (akan kami pertimbangkan lebih lanjut dan kemungkinan akan kami coba juga tanpa drop untk jadi bahan perbandingan akurasi kedepannya)
8. Untuk kategorik seperti SEX dan EDUCATION akan di label encoding sedangkan sisanya akan OneHotEncod

STAGE 1

**CREDIT CARD
PAYMENT DEFAULT
PREDICTION**