



# HOME CREDIT DEFAULT RISK

Google Colab: <https://bit.ly/FPZeniusKelompok18>



# KELOMPOK 18

FIDELA ELYSIA RISWI PUTRI

EDELIN GULTOM

SHINTIA YULITA SARI

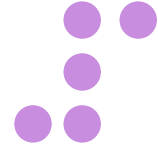
ANAK AGUNG BINTANG KRISNA DEWI

# SUMBER DATA: HOME CREDIT DEFAULT RISK

[Home Credit Default Risk | Kaggle](#)

## PROBLEM: CLASSIFICATION

01



# BUSINESS UNDERSTANDING



# HOME CREDIT

Salah satu lembaga pembiayaan di Indonesia dengan berbagai pilihan pelayanan seperti pembelian gadget, kamera, handphone, laptop, TV, furniture serta kebutuhan rumah tangga lainnya.

# HOME CREDIT



## PENERIMA KREDIT

mendapatkan sumber pembiayaan yang mudah dan cepat.



## PEMBERI KREDIT

mendapatkan keuntungan dari bunga yang dibayarkan oleh debitur.

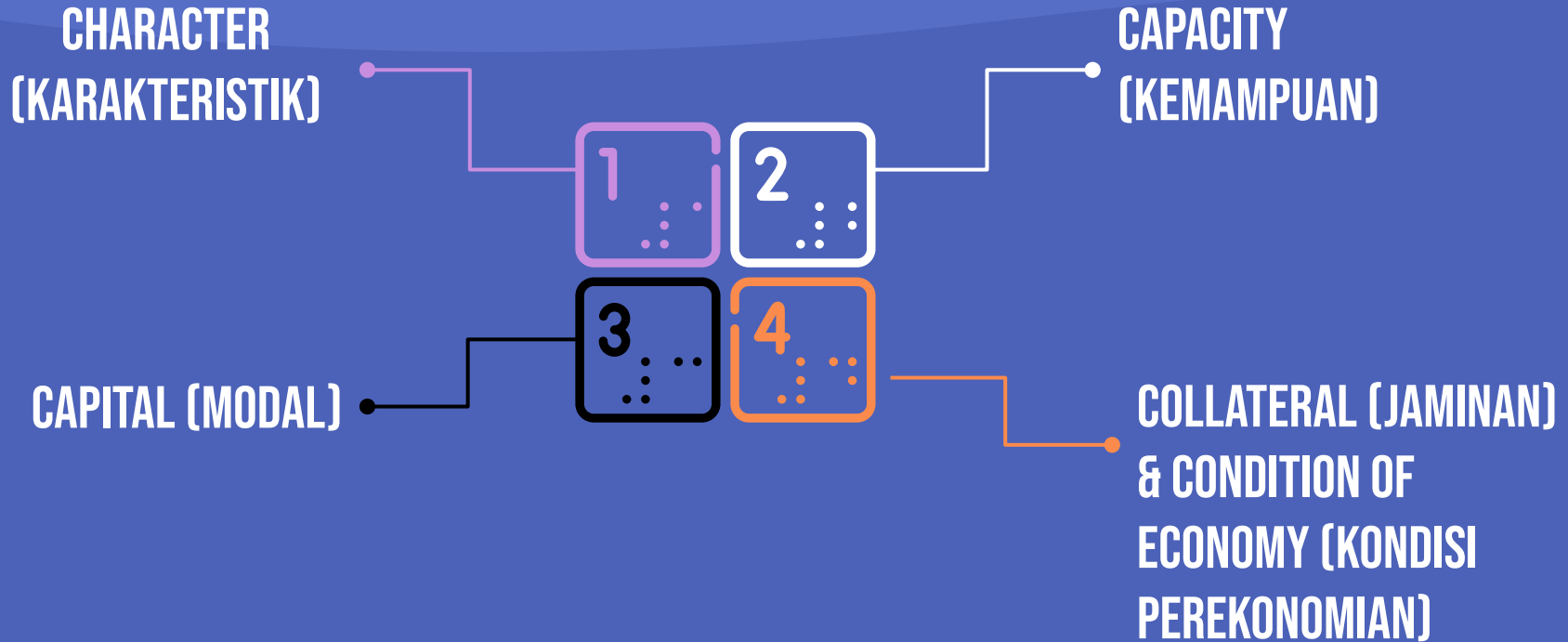
# PEMBERI KREDIT

Kreditor harus menanggung risiko ketidaklancaran pembayaran oleh debitur.

Pihak kreditor harus mengevaluasi setiap pemohon kredit sebelum keputusan penerimaan atau penolakan diambil.



# ANALISIS KREDIT





# MODEL CREDIT SCORING

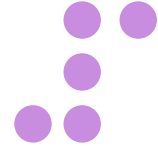
## Tujuan:

- Mengetahui faktor dan memprediksi penyebab klien gagal bayar.
- Memprediksi model machine learning mengenai pelanggan yang akan gagal bayar dengan aspek-aspek terkait untuk menghindari risiko gagal bayar atau tidak membayar tepat waktu.

Metode: Random Forest, Logistic Regression, dan Decision tree

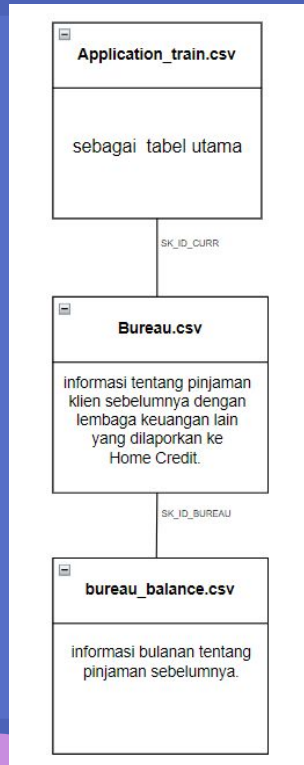


02



# DATA UNDERSTANDING

# DESKRIPSI TABEL



# EXPLORATORY DATA ANALYSIS SETIAP TABEL

## 1. APPLICATION\_TRAIN.CSV

Dataframe application\_train memiliki 122 kolom dan 67 kolom memiliki missing values

### 1. SK\_ID\_CURR

ID pinjaman dalam sampel kami

### 2. TARGET

1 = Klien dengan kesulitan pembayaran

0 = Klien tanpa kesulitan pembayaran

### 3. NAME\_CONTRACT\_TYPE

Identifikasi apakah pinjaman tunai atau bergulir.

# EXPLORATORY DATA ANALYSIS SETIAP TABEL

## 4. CODE\_GENDER

Jenis kelamin klien

## 5. FLAG\_OWN\_CAR

Tandai jika klien memiliki mobil

## 6. FLAG\_OWN\_REALTY

Tandai jika klien memiliki rumah atau flat

## 7. CNT\_CHILDREN

Jumlah anak yang dimiliki klien

## 8. AMT\_INCOME\_TOTAL

Penghasilan klien

## 9. AMT\_CREDIT

Jumlah kredit pinjaman

## 10. AMT\_ANNUITY

Anuitas pinjaman

Note: jumlah kredit / anuitas (penjelasan tentang bunga anuitas Jenis dan Cara Perhitungan Bunga Kredit - Home Credit)

# EXPLORATORY DATA ANALYSIS SETIAP TABEL

## 2. BUREAU.CSV

Dataframe bureau memiliki 17 kolom dan 7 kolom memiliki missing values.

### 1. SK\_ID\_CURR

ID pinjaman dalam sampel credit home.

### 2. SK\_BUREAU\_ID

ID yang dikodekan ulang dari kredit Biro Kredit sebelumnya.

### 3. CREDIT\_ACTIVE

Status kredit yang dilaporkan Biro Kredit (CB)

### 4. CREDIT\_CURRENCY

Mata uang yang dikodekan ulang dari kredit Biro Kredit

### 5. DAYS\_CREDIT

Berapa hari sebelum aplikasi saat ini klien mengajukan kredit Biro Kredit

### 6. CREDIT\_DAY\_OVERDUE

Jumlah hari yang telah jatuh tempo pada kredit CB pada saat pengajuan pinjaman terkait dalam sampel credit home

### 7. DAYS\_CREDIT\_ENDDATE

Sisa durasi kredit CB (dalam hari) pada saat pengajuan di Home Credit

# EXPLORATORY DATA ANALYSIS SETIAP TABEL

## 8. DAYS\_ENDDATE\_FACT

Beberapa hari sejak kredit CB berakhir pada saat pengajuan Kredit Rumah (hanya untuk closed credit)

## 9. CNT\_CREDIT\_PROLONG

Berapa kali kredit Biro Kredit diperpanjang

## 10. AMT\_CREDIT\_SUM

Jumlah kredit saat ini untuk kredit Biro Kredit

# EXPLORATORY DATA ANALYSIS SETIAP TABEL

## 3. BUREAU\_BALANCE.CSV

Dataframe bureau\_balance memiliki 3 kolom dan 0 kolom memiliki missing values.

### 1. SK\_BUREAU\_ID

ID kredit Biro Kredit yang dikodekan ulang (pengkodean unik untuk setiap application)

### 2. MONTHS\_BALANCE

Bulan saldo relatif terhadap tanggal aplikasi (-1 berarti tanggal saldo segar)

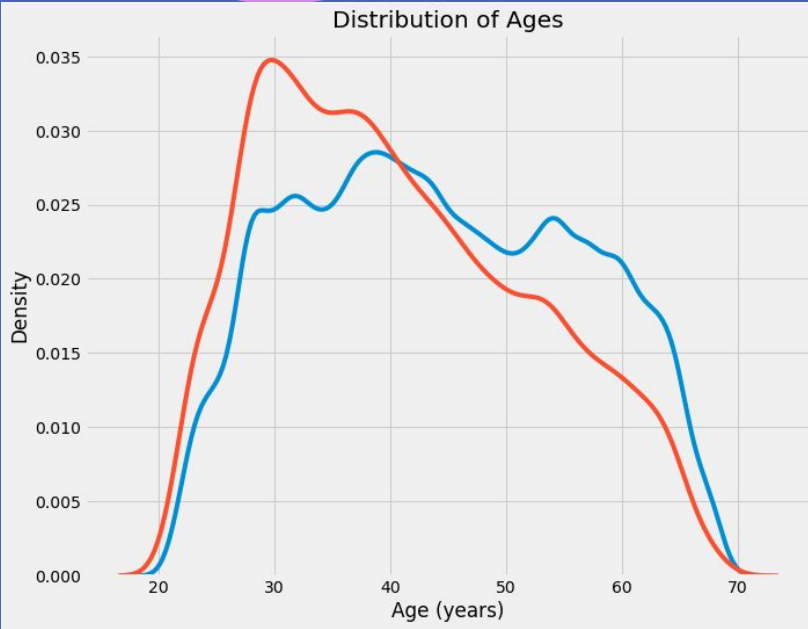
### 3. STATUS

Status pinjaman Biro Kredit selama sebulan

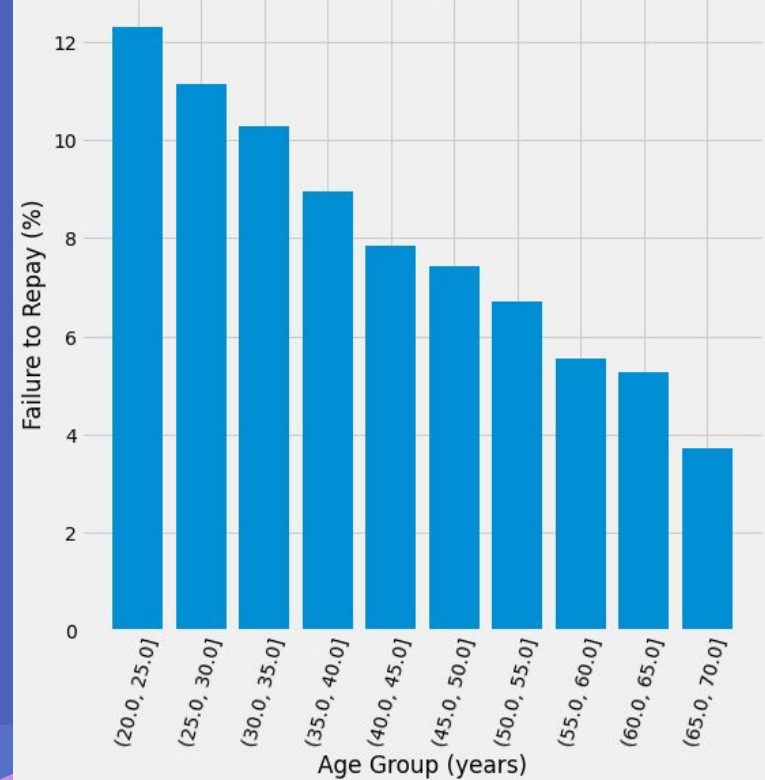


# EXPLORATORY DATA ANALYSIS

Distribution of Ages



Failure to Repay by Age Group



# EXPLORATORY DATA ANALYSIS

Berdasarkan visualisasi yang telah dilakukan dapat diketahui beberapa hal berikut :

- Client yang lebih muda lebih cenderung tidak membayar kembali pinjaman. Tingkat kegagalan untuk membayar di atas 10% untuk tiga kelompok usia termuda dan di bawah 5% untuk kelompok usia tertua.



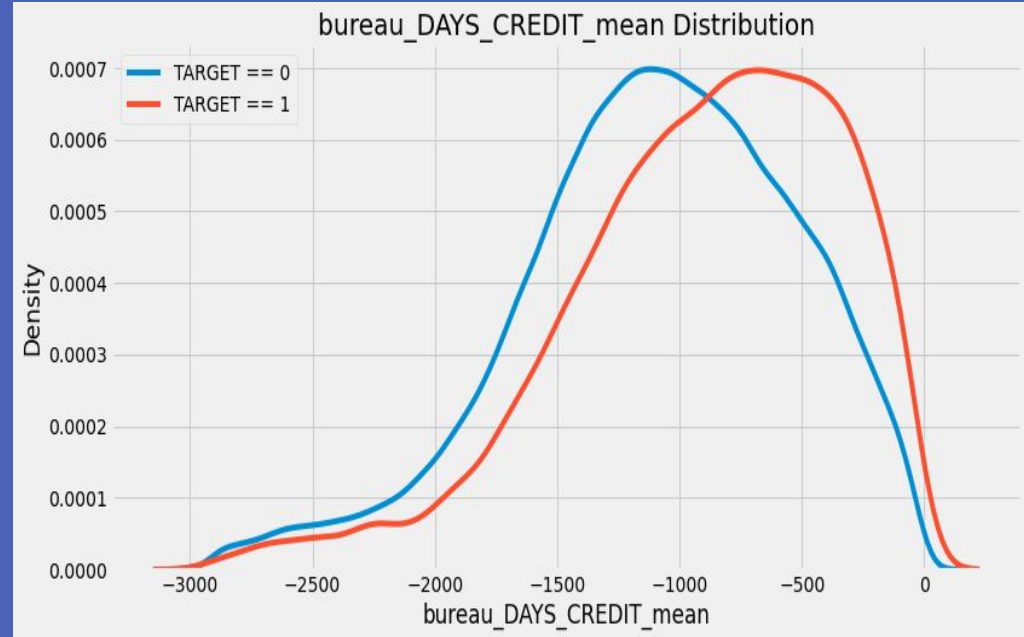
# EXPLORATORY DATA ANALYSIS

Tabel bureau kita melakukan Group by the client id, kemudian dilakukan calculate aggregation statistics meliputi ['count', 'mean', 'max', 'min', 'sum'].

kolom

bureau\_DAYS\_CREDIT\_mean memiliki korelasi yang cukup tinggi diantara fitur lain yakni

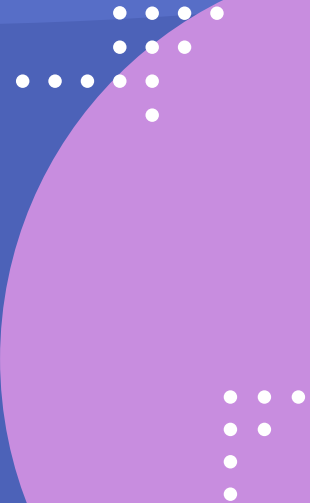
0.08972896721998116.



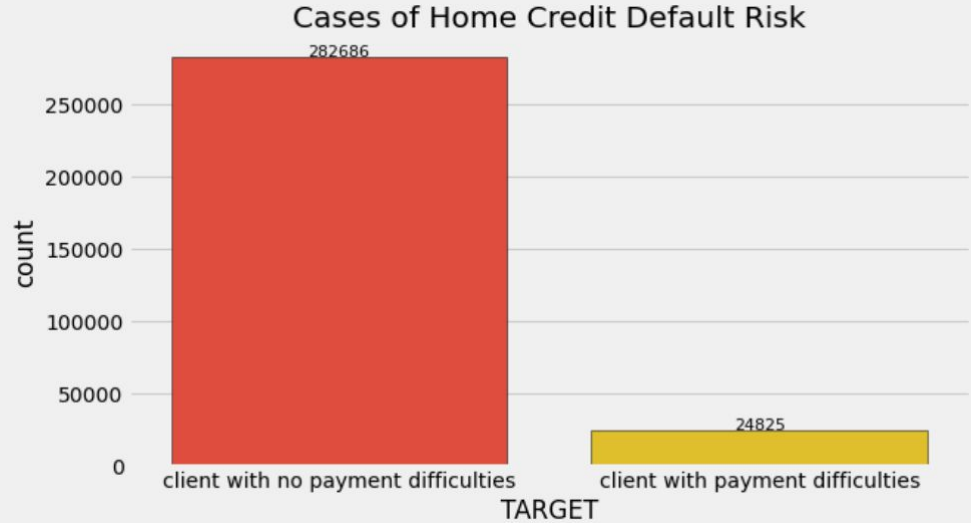
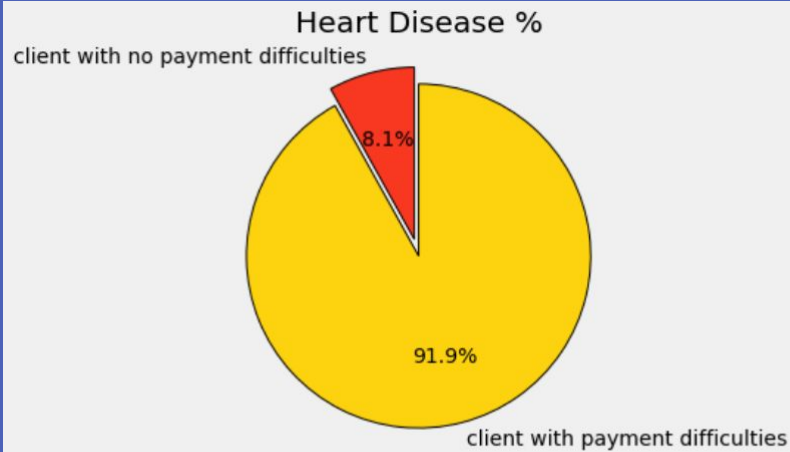
# EXPLORATORY DATA ANALYSIS

Berdasarkan visualisasi yang telah dilakukan dapat diketahui beberapa hal berikut :

- angka negatif yang lebih besar menunjukkan bahwa pinjaman sudah dilakukan jauh sebelum aplikasi pinjaman saat ini. Klien yang mengajukan pinjaman lebih jauh di masa lalu berpotensi lebih mungkin untuk membayar kembali pinjaman di Home Credit.



# TARGET VARIABLE VISUALIZATION (TARGET) ::

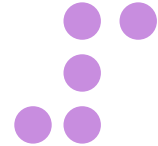


# EXPLORATORY DATA ANALYSIS

Berdasarkan visualisasi dan EDA yang telah dilakukan dapat diketahui beberapa hal berikut :

- Terdapat sekitar 8.1% pelanggan yang kesulitan melakukan pembayaran atau sekitar 24.825 pelanggan.

03



# DATA PREPARATION

# MENGHITUNG JUMLAH PINJAMAN KLIEN SEBELUMNYA

Pandas Operations:

groupby	Mengelompokkan dataframe menurut kolom.
agg	Melakukan perhitungan pada data yang dikelompokkan seperti mengambil rata-rata kolom.
merge	Mencocokkan aggregated statistics dengan klien yang sesuai.



# TAHAP MEMILIH FITUR

01



Menghitung Koefisien Korelasi Pearson (r-value) antara variabel dan target.

02



Aggregating Numeric Columns menggunakan 'mean', 'max', 'min', 'sum', 'count'.

03



Hitung korelasi semua nilai baru dengan target. Mengurutkan korelasi berdasarkan besarnya (nilai absolut) yang diurutkan dan fungsi lambda.

04



Categorical Variables akan memanfaatkan penghitungan jumlah nilai dari setiap kategori dalam setiap variabel kategoris.

05



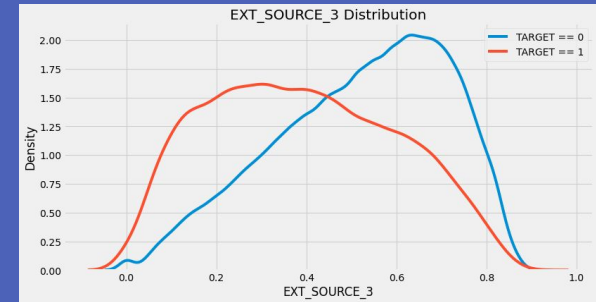
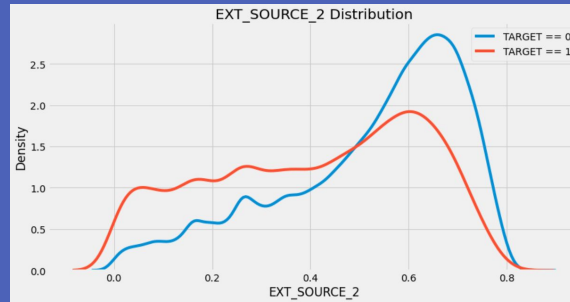
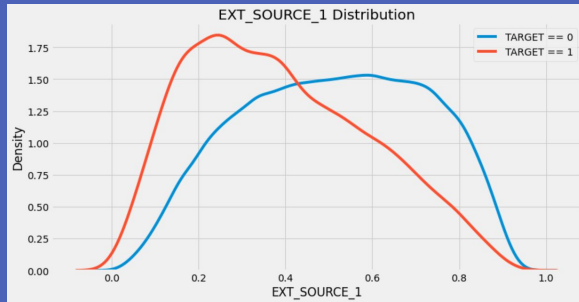
Merge dengan tabel utama (application\_train.csv). Kemudian cek missing values.

06



Mencari variabel yang memiliki korelasi lebih besar dari 0,8 dengan variabel lain.

# KORELASI ANTARA VARIABEL DAN TARGET ::::



menunjukkan bahwa risiko kesulitan pembayaran tinggi ketika peringkat dari sumber eksternal 1, 2, 3 rendah. Demikian pula, peluang pembayaran lebih tinggi ketika peringkat dari sumber-sumber ini lebih tinggi.

# PERHITUNGAN TARGET

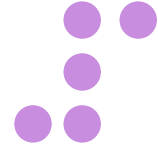
## Perhitungan Target



0: klien melakukan pembayaran tepat waktu

1: klien memiliki kesulitan pembayaran

04



# MODELING

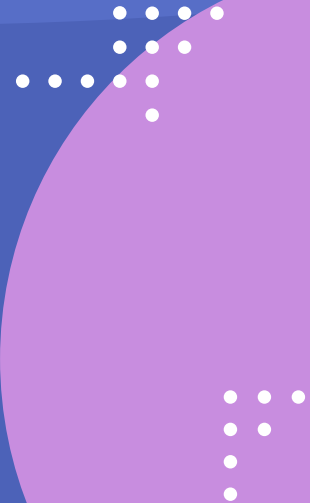
# METRIK

Metrik yang digunakan adalah metrik confusion untuk mengevaluasi akurasi klasifikasi. Terdiri dari:

- (TN) True Negative
- (FP) False Positive
- (FN) False Negative
- (TP) True Positive

# TRAIN TEST SPLIT

Train/test split adalah salah satu metode yang dapat digunakan untuk mengevaluasi model dengan membagi dataset menjadi dua bagian, yakni bagian yang digunakan untuk training data dan untuk testing data dengan proporsi tertentu (train size = 60%, test size = 40%)



# LOGISTIC REGRESSION

## Evaluasi dengan Classification Report

Dari classification report disebutkan bahwa model memiliki akurasi 92%

## Evaluasi dengan AUC

Score AUC nilainya sekitar 50%



# RANDOM FOREST

## Evaluasi dengan Classification Report

Dari classification report disebutkan bahwa model memiliki akurasi 92%

## Evaluasi dengan AUC

Score AUC nilainya sekitar 50%





# DECISION TREE

## Evaluasi dengan Classification Report

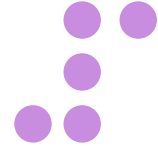
Dari classification report disebutkan bahwa model memiliki akurasi 92%

## Evaluasi dengan AUC

Score AUC nilainya sekitar 53,7%



05



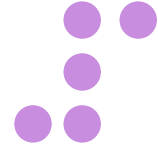
# CONCLUSION

# KESIMPULAN

Berdasarkan evaluasi yang telah dilakukan pada ketiga model, model terbaik adalah Decision Tree karena memiliki akurasi dan AUC tertinggi di antara ketiga model.

Dalam pemodelan ini, kelompok kami menggunakan semua variabel yang ada pada data karena dirasa variabel dalam data memengaruhi pelanggan yang Target.

06



# DEPLOYMENT

**Model di-deploy  
menggunakan  
aplikasi  
Streamlit**



**Streamlit**

# TAMPILAN HASIL DEPLOY

← → ↻ ⓘ localhost:8501

🏠 ☆ 📱 B ⋮

🌊 RUNNING... Stop ☰

## Home Credit Risk

What is your contract type?

Cash loans ▼

You selected: Cash loans

Select gender

M ▼

You selected: M

Select type

N ▼

You selected: N

# TERIMA KASIH!

**CREDITS:** This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#), and infographics & images by [Freepik](#)

Please keep this slide for attribution