



PRE-PROCESSING

Stage-1_Final Project

InsideData



TAHAPAN PRE-PROCESSING

01

DATA CLEANSING

Menyiapkan data agar dapat digunakan untuk analisis: Handle missing values, Handle duplicated data, Handle outliers, Feature transformation, Feature encoding, dan Handle class imbalance

02

FEATURE ENGINEERING

Melakukan Feature Selection dan Extraction untuk membantu proses modelling nanti.

03

GIT

Upload project di sebuah repository git. Berkolaborasi di Git jika ada perubahan version dari waktu ke waktu.

DATA CLEANSING

01

HANDLE MISSING VALUES

Data Missing Values dibuang karena hanya berjumlah 24 dari 2240 data

02

HANDLE DUPLICATED DATA

Tidak ada data yang duplikat proses dapat dilanjutkan.

03

HANDLE OUTLIER (BELUM LENGKAP)

Menggunakan 2 alternatif metode yaitu **IQR** dan **z-test** untuk mendeteksi outlier. Setelahnya Outlier akan dibuang.

DATA CLEANSING

04

FEATURE TRANSFORMATION

Transformasi fitur-fitur yang berdistribusi positive skewed dengan metode **log-transform**.

05

FEATURE ENCODING

Feature Encoding dengan metode **One Hot Encoding** karena distinct values pada fitur kategorik memiliki nilai lebih dari 2.

06

DATA SPLITTING

Memecah data guna membagi menjadi data yang digunakan untuk data latih dan data uji. Tujuannya adalah agar data uji tidak terdampak pada proses handling imbalance data.

DATA CLEANSING



07


DATA SCALLING

Scaling data dengan standardscaller agar fitur-fitur data lebih seragam.

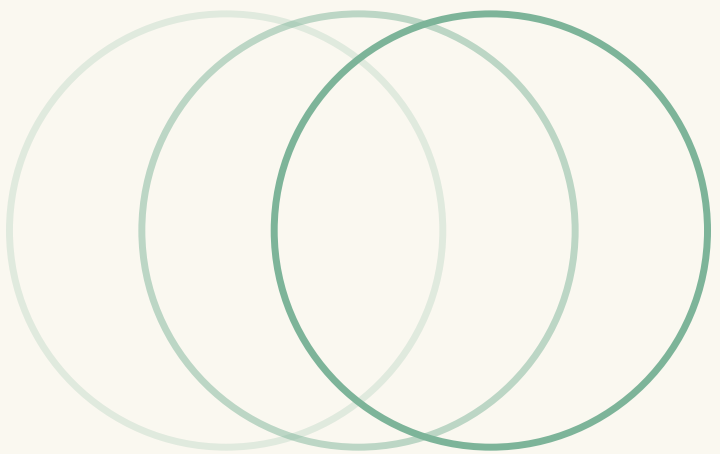
08

HANDLE CLASS IMBALANCE

Menggunakan metode **oversampling** untuk mengatasi class imbalance dikarenakan ketidakseimbangan yang jauh.



FEATURE ENGINEERING



FEATURE EXTRACTION

- Membuat kolom have child (jumlah anak kecil maupun remaja)
- Membuat kolom sum of ammount, total jumlah pembelian pelanggan (dari berbagai produk)
- Membuat kolom acc campaign, apakah pelanggan pernah menerima penawaran campaign (0 dan 1).
- Membuat kolom Age

FEATURE SELECTION

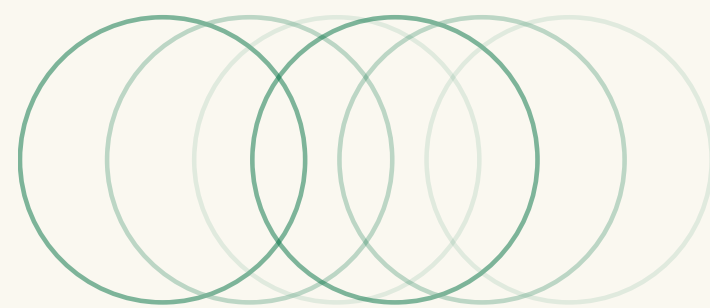
- Menghapus Kolom ID
- Menghapus Kolom Z_cost
- Menghapus kolom Revenue

FEATURE TAMBAHAN (IDE)

- Harga Produk.
- Besaran diskon.
- Tingkat Pengeluaran (Spending_Level).
- Umur Pelanggan.

GIT REPOSITORY

[Link Github](#)



TERIMAKASIH

