



Kelompok: 4

Stage: 1

Mentor: Kak Al Fath Terry

Pukul/ Tanggal: 20:00/ 17 Maret 2023

Pembagian tugas di stage ini:

- 1.Data Exploration ---> Dimas
- 2.Exploratory Data Analysis ---> Lana & Ucup
- 3.Business Insights ---> Keven & Aulia
- 4.Data Visualization & Data storytelling ---> Andre & Dila

Poin Pembahasan:

1. Membahas Hasil dari Stage 1
2. Membahas Data Train Test

Hasil Diskusi:

- 1.Tidak disarankan untuk menggunakan Oversampling dan Undersampling
2. Mencoba gunakan Logarithmic transformation (disarankan)
3. Menggunakan normality test untuk melihat/mencari tahu fitur transformation mana yang terbaik.

Hasil Diskusi:

4. Scaling —> Mengambil 1 teknik Scaling saja, seperti min max scaler dan standard scaler.
5. Dalam EDA harus melakukan featured selection method, untuk kasus dataset ini sebaiknya pakai ANOVA dan Chi Squared (best practice)
6. Di Rakamin yang diajarkan baru pearson correlation (input var numerik, out var numerik), meskipun bukan best practice.



Kelompok: 4

Stage: 1

Mentor: Kak Al Fath Terry

Pukul/ Tanggal: 20:00/ 17 Maret 2023

Fitur Transformation, selain Log transformation, bisa dicoba:

- Square root transformation
- Reciprocal transformation
- Exponential transformation
- Yeo Johnson transformation

Outlier

- Untuk kebutuhan ML sebaiknya dihapus
- Penghapusan outliers dilakukan setelah splitting (data trainnya)
- Threshold nya dari data train (karena untuk menghindari data leaked)

Korelasi

- Bounce dan exit tinggi pilih salah satu yang terbaik yaitu exit rate. (karena nilai korelasi terhadap target lebih baik exit rate -0,21) kemudian drop bounce rate.
- Exit rate dan bounce rate bisa dijadikan satu (PCA) → bisa explore dulu

Business insight

- Masukan tentang binning → melakukan treatment setiap kelompok
- Buat rekomendasi bisnis tentang returning dan new visitors
- Buat rekomendasi bisnis kenapa bulan Februari tidak ada penjualan, padahal ada hari Valentine, bisa ditambah promo tentang valentine karena pada bulan ini visual revenue-nya yang rendah.