



**Kelompok: Scidata Solution**  
**Stage: 2**  
**Mentor: Kevin**  
**Pukul/ Tanggal: 20.00 / 1 Juni 2023**

**Pembagian tugas di stage ini:**

- Data Cleansing (Fahad, Rizky, Vida, Nizar)
- Feature Engineering (Alvin, Afri, Dean, Tsaniya)

**note :**

**dalam pembagian, masing-masing tetap mengerjakan semua tugas**

**Poin pembahasan:**

- Penjelasan hasil diskusi Stage 2 kelompok 2 kepada Mentor
- Koreksi dan tambahan dari mentor terhadap pengerjaan stage 2

**Hasil Diskusi:**

1. Pada data untuk outlier dibuat di masing-masing kolom saja
2. Untuk penggunaan IQR dan Z-Score dilihat dari banyaknya outlier jika  $> 10\%$  maka gunakan IQR, namun jika  $< 10\%$  gunakan Z-Score
3. Handle Imbalance disimpan dibelakang setelah Feature Selection dan yang dibuat imbalance iitu data trainnya
4. Semua data dibuat sehingga menjadi distribusi Normal
5. Pada Feature Engineering sebelum dibutkan feature selectin buat dahulu correlation map, lalu setelah itu bandingkan sehingga kolom baru yang dibuat untuk feature selection adalah hasil dari perbandingan heatmap. Kolom yang diambil Max.20
6. Untuk penggabungan data yang utuhnya di copy dahulu lalu di split dan di train sehingga hasil 70%nya itu bisa di oversampling namun jika sudah cukup, dengan data train sudah oke

**Tindak Lanjut:**

Semua anggota kelompok merevisi hasil pengerjaan sebelumnya dan menambahkan beberapa yang belum sesuai dengan instruksi yang sudah dibahas dan di diskusikan dengan mentor