



**Kelompok: e-numpy**

**Stage: 2**

**Mentor: Muhammad Mirza Fahmi**

**Pukul/ Tanggal: 21.00, 3 April 2022**

**Pembagian tugas di stage ini:**

Nama: Muhammad Chosasih Mahendra, Putri Maylita

Data Cleansing

Nama: Samuel Akwila, Ilham Taufiqurrohman, Rama Putra

Feature Engineering, Laporan Stage 2

**Poin pembahasan:**

1. Outliers dan metode penghapusan outliers
2. Feature Transformation
3. Feature Selection

**Hasil Diskusi:**

1. Melihat kembali outliers, ada anomalnya atau tidak
2. Coba pairplot data yang numeric, cari outliersnya
3. Transformasi pake standardisasi karena masih skew
4. Resamplingnya pake 1:2, jangan 1:1.
5. Lihat lagi korelasi antar variable dan korelasi ke variable target. Minus tinggi juga perlu di drop dan korelasi ke variable target kecil perlu di drop juga.
6. Sebelum permodelan, cari feature importancenya dulu di stage selanjutnya.
7. Gender ganti ke One Hot Encoding
8. Pisahkan dataframe tiap stage, biar gampang otak atik.

**Tindak Lanjut:**

1. Melihat outliers lebih detail dan lihat pairplot numeric datanya.
2. Metode transformasi datanya ganti ke standardisasi.
3. Resampling nya ganti perbandingan 2:1.
4. Cek ulang korelasi dan drop fitur sampe setengahnya kalo bisa.
5. Menggunakan labeling One Hot Encoding buat kolom Gender.
6. Untuk stage 3, cari feature importance lalu baru deploy Machine Learning nya.
7. Memisahkan tiap-tiap dataframe per stage.