



Kelompok: Mandalika

Stage: 2

Mentor: Mas Hafizh Adi Prasetya

Pukul/ Tanggal: 16.30/03 April 2022

Pembagian tugas di stage ini:

Nama: Ardilla Safitri (Feature Encoding & Feature Selection, Extraction)

Nama: Jonse Kennedy (Handle Missing Values & Feature Selection, Extraction)

Nama: Arni Cici (Handle Outlier & Feature Selection, Extraction)

Nama: Lise Listiani (Feature Transformation & Feature Selection, Extraction)

Nama: Fakhry Husnul (Handle Class Imbalanced & Feature Selection, Extraction)

Nama: Riska Diah (Handle Duplicated Data & Feature Selection, Extraction)

Poin pembahasan:

1. Feedback (Drop CustomerID, Duplicated Data, Feature Encoding, Handle Class Imbalanced, Feature Transformation, dan Handle Outlier)
2. QnA
3. Target Stage 3

Hasil Diskusi:

1. Feedback:
 - a. CustomerID dihapus karena unique valuesnya terlalu banyak sehingga tidak bisa digunakan sebagai feature
 - b. Pada bagian Handle Duplicated data yang tidak menggunakan CustomerID data yang memiliki duplicated harus dihapus.
 - c. Pada bagian Feature Encoding, gunakan One Hot Encoding karena tidak ada kategori yang lebih tinggi dari yang lain.
 - d. Pada bagian Handle Class Imbalanced lebih baik dilakukan pada data train setelah split data dan pilih salah satu sehingga terdapat satu dataset yang akan digunakan.
 - e. Pada bagian Feature Transformation untuk setiap data skewed yang menghasilkan nilai $\log 0$, maka dapat ditambahkan dengan nilai yang kecil sehingga datanya digeser ke kanan.
 - f. Pada bagian Handle Outlier lebih baik tidak digunakan kembali pada saat data sudah displit.
2. QnA : Pada saat EDA sudah dipilih feature mana saja yang akan digunakan, apakah untuk preprocessing bisa hanya menggunakan feature yang dipilih dan drop feature yang lainnya?
3. Target Stage 3 : Buat beberapa model (3 model) dengan menggunakan dataset yang sama lalu bandingkan, serta tentukan matriksnya apa dan latih modelnya dengan Hiperparameter Tuning lalu interpretasi.

Tindak Lanjut:

1. Feedback:
 - a. Kolom CustomerID dihapus
 - b. Pada dataset yang digunakan duplicated data sudah dihapus
 - c. Pada Feature Encoding sudah digunakan One Hot Encoding
 - d. Pada Handle Class Imbalanced sudah dilakukan pada data train dan dipilih satu dataset
 - e. Pada Feature Transformation sudah sesuai
 - f. Pada Handle Outlier tidak digunakan Kembali pada dataset yang sudah di split.
2. QnA : Setelah EDA feature yang tidak akan digunakan bisa dihapus saja , sehingga feature yang digunakan pada data preprocessing hanya feature yang akan digunakan saja.
3. Target untuk Stage 3: Mempersiapkan untuk Stage 3