

PRE-TEST DATA ANALYST

A. Kategorisasi Data

1. Pengolahan data

- Buatlah data dummy sebanyak 1 juta data kemudian masukkan ke dalam database seperti PostgreSQL atau tools database yang lainnya. Untuk table head di dalam databasenya sesuai dengan contoh JSON di bawah ini.

```
{
  "ACCOUNT_NUMBER": "100002",
  "TRX_TYPE": "Ewallet",
  "AMOUNT": -20000,
  "FLAG_DB_CR": "D",
  "INFORMATION": "GOPAY GP-6287860594619",
  "TRX_DATE": 1466812800000,
  "REFERENCE_NUMBER": "274678",
  "CURRENCY": "IDR",
  "TRX_TIME": "10:03:15",
  "CIF": "1234567890",
  "FITUR": "Gopay"
}
```

(note: untuk dummy data minimal 100.000 - 500.000 data)

- Setelah data sudah disiapkan, lakukan preprocessing data (***data cleaning***) terhadap data yang sudah di generate seperti (menghilangkan tanda baca, mengonversi huruf ke huruf kecil, dan menghapus kata-kata berhenti).

(note: jika data yang digenerate sudah bersih tidak usah melakukan processing lagi).

- Lakukan ekstraksi fitur teks seperti TF-IDF atau Word Embeddings.

(note: jika data yang digenerate sudah bersih tidak usah melakukan ekstraksi lagi).

- Lakukan pembagian data menjadi dua yaitu data latih dan data uji.
- Pilih model machine learning yang paling gampang anda implementasikan dalam case tersebut seperti (CNN, KNN, Naive Bayes, RNN, dan lain sebagainya).
(note: sebagai nilai tambahan, tambahkan implementasi model lain, sebagai pembandingan hasil yang di dapatkan)
- Lakukan pelatihan model dan validasi model

2. Setelah melakukan pengolahan data, petakan setiap transaksi ke setiap kategori (Kategori berdasarkan kriteria Anda sendiri).

Ex: jika "TRX_TYPE" : "Ewallet" dan "FITUR": "Gopay" masuk ke dalam kategori "SHOPPING"

- #### 3. Kemudian hitung akumulasi ('AMOUNT') tiap kategori dari data yang sudah di kategorisasi dalam range mingguan/bulanan/tahunan, perhitungkan juga apabila terjadi transaksi reversal.

4. Manfaatkan PL/pgSQL pada PostgreSQL atau PL/SQL pada Oracle
5. Gunakan tools seperti Python, Jupyter Notebook atau Google Collaboration

Informasi tambahan:

CIF: Merupakan identifier unik setiap nasabah

AccountNumber: Merupakan nomor rekening nasabah

Jenis-jenis transaksi yang **harus** ada di dummy data pada field "INFORMATION"

"WITHDRAWAL TRANSFER (ATM)", "GOPAY", "DEPOSIT TRANSFER (ATM)", "DEPOSIT CROSS BORDER", "WITHDRAWAL TRANSFER", "REVERSAL", "BIAYA ADMIN", "PAYMENT CC", "PAYMENT TELCO PASCABAYAR"

Jika ada tambahan lain yang ingin ditambahkan sesuai kreatifitas sendiri, dipersilahkan (dan akan mendapat nilai tambah).