



Instruksi Tugas

customer_demography :

https://docs.google.com/spreadsheets/d/1TbsNLv8LwlKlaJggq2MOeqL5hgqJNTH_cdLDrhfUjKc/edit?gid=390105770#gid=390105770

savingmaster : <https://docs.google.com/spreadsheets/d/1OEiaai8QenH-W6mPOziADYas6I4x9-AjE3J87Qf2AbQ/edit?gid=278815876#gid=278815876>

transaction :

https://docs.google.com/spreadsheets/d/1TbsNLv8LwlKlaJggq2MOeqL5hgqJNTH_cdLDrhfUjKc/edit?gid=390105770#gid=390105770

- 1 Lakukan join antara tabel customer_demography dan saving menggunakan kolom customer_id.
- 2 Gabungkan hasil dari langkah pertama dengan tabel transaction menggunakan customer_id.
- 3 Pastikan hasil akhir dari proses join menghasilkan tepat 8.397 baris(customer pada data demography wajib masuk).
- 4 Lakukan imputasi (pengisian missing value) pada kolom-kolom berikut:
 - age (usia)
 - gender (jenis kelamin)
 - marital_status (status nikah)
 - occupation (jenis pekerjaan)
 - freq_debit
 - freq_credit
- 5 Bebas menggunakan teknik imputasi apapun, seperti:
 - Mean/median/mode/(mean, median, mode) berdasarkan kolom lain / domain knowledge
 - KNN/model lain

Dictionary Tabel

customer_id

ID unik untuk setiap pelanggan. Digunakan sebagai primary key untuk melakukan join antar tabel.

account_id

ID unik untuk setiap rekening tabungan/giro. Digunakan sebagai foreign key untuk mengaitkan transaksi dengan pelanggan.

umur

Usia pelanggan (dalam tahun). Tipe data numerik. Kolom ini penting untuk segmentasi demografi dan analisis risiko.

region

Wilayah tempat tinggal pelanggan (contoh: Jakarta, Jawa Timur, dll). Tipe data kategorikal. Bisa digunakan untuk analisis geografis.

jenis_kelamin

Jenis kelamin pelanggan, misalnya: 'Pria', 'Wanita'. Tipe kategorikal.

status_nikah

Status pernikahan pelanggan, misalnya: 'Menikah', 'Belum Menikah', 'Duda'. Tipe kategorikal.

jenis_pekerjaan

Kategori pekerjaan pelanggan.

has_sms_fin

Indikator apakah pelanggan menggunakan layanan keuangan berbasis SMS (1 = ya, 0 = tidak). Tipe numerik biner.

has_internet_fin

Indikator apakah pelanggan menggunakan layanan keuangan berbasis internet/mobile banking (1 = ya, 0 = tidak). Tipe numerik biner.

has_tabungan

Indikator apakah pelanggan memiliki rekening tabungan (1 = ya, 0 = tidak). Tipe numerik biner.

has_giro

Indikator apakah pelanggan memiliki rekening giro (1 = ya, 0 = tidak). Tipe numerik biner.

has_deposito

Indikator apakah pelanggan memiliki deposito berjangka (1 = ya, 0 = tidak). Tipe numerik biner.

has_loan

Indikator apakah pelanggan memiliki pinjaman/kredit aktif (1 = ya, 0 = tidak). Tipe numerik biner.

jenis_merchant

Kategori merchant yang paling sering digunakan pelanggan saat bertransaksi.

ratas_saldo_m1 sampai ratas_saldo_m6

Rata-rata saldo bulanan rekening pada bulan ke-1 hingga bulan ke-6, dihitung mundur dari bulan observasi terakhir.

Contoh: `ratas_saldo_m1` = saldo rata-rata bulan Mei jika Juni adalah bulan terakhir.

Tipe data numerik (dalam satuan rupiah).

cbal_base_m1 sampai cbal_base_m6

Saldo akhir rekening pada akhir bulan ke-1 hingga ke-6.

Tipe data numerik (dalam rupiah).

hold_base_m1 sampai hold_base_m6

Jumlah dana yang tertahan (misalnya karena blokir, hold sementara, dll) pada bulan ke-X.

Tipe data numerik.

abal_base_m1 sampai abal_base_m6

Saldo rata-rata harian selama bulan ke-X. Berbeda dari `ratas_saldo` yang bisa berbasis saldo akhir mingguan atau bulanan.

Tipe data numerik.

freq_debit_m1 sampai freq_debit_m6

Jumlah transaksi **debit** (penarikan dana) yang dilakukan pada bulan ke-X.

Tipe data numerik (bilangan bulat).

freq_credit_m1 sampai freq_credit_m6

Jumlah transaksi **kredit** (penambahan dana) pada bulan ke-X.

Tipe data numerik (bilangan bulat).

amt_debit_m1 sampai amt_debit_m6

Total nominal seluruh transaksi debit di bulan ke-X.

Tipe data numerik (dalam rupiah).

amt_credit_m1 sampai amt_credit_m6

Total nominal seluruh transaksi kredit di bulan ke-X.

Tipe data numerik (dalam rupiah).