**SOAL KOMPETISI DATA SAINS 2024**

**Data**

Data yang diolah adalah data sampel hasil survey kesehatan suatu perusahaan sebanyak 1336 pegawai dan 15 variabel yang diukur (terdapat pada sheet **Data 2**). Perhatikan bahwa setiap variabel memiliki jenis data yang berbeda yaitu: jenis kategorik dan numerik. Antar variabel yang diukur bisa jadi memiliki hubungan linier dan kausalitas (sebab akibat) sehingga dapat dimodelkan untuk mencapai suatu tujuan tertentu.

**Tujuan yang ingin dicapai:**

Faktor-faktor kesehatan apa yang paling berpengaruh terhadap nilai kolesterol total (CT).

**Output yang diharapkan:**

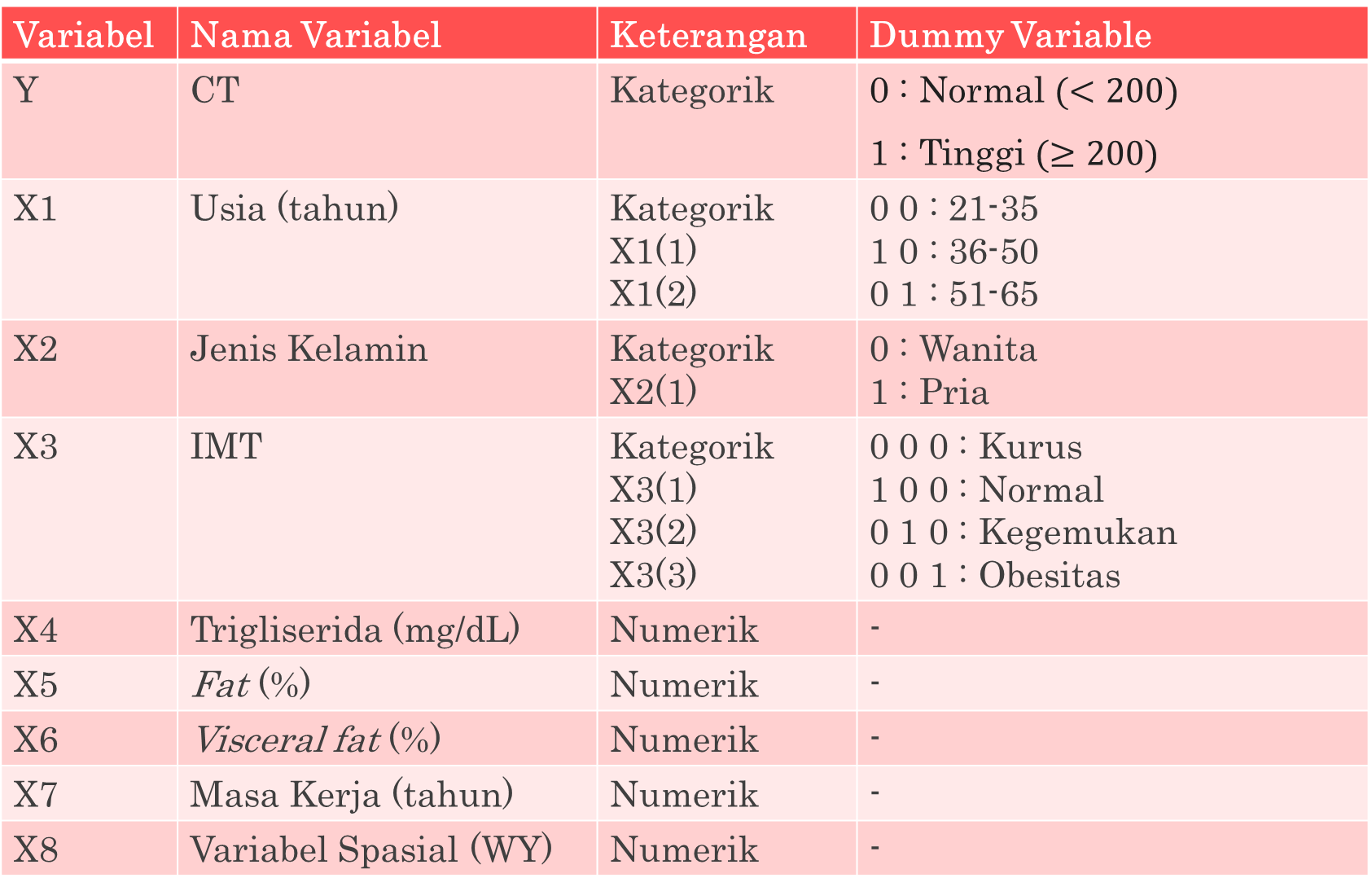
1. Statistika deskriptif dari setiap variabel-variabel yang diukur.
2. Mencari korelasi antar variabel yang akan diolah.
3. Memodelkan koresterol total dan variabel yang memengaruhinya.
4. Mendapatkan faktor-faktor apa yang paling berpengaruh terhadap koresterol total.
5. Menginterpretasikan hasil pemodelan.

**Tools:** Excel, R, Python

**Metode:** Pemodelan Regresi Logistik

**Langkah-langkah pemodelan:**

Menentukan kategori dari variabel yang digunakan:



**Prediktif Model dari Data 1**

Berdasarkan pendefinisikan variabel respon dan prediktor di atas diperoleh hasil model (**Data 1**) sebagai berikut;

**Variabel prediktor yang berpengaruh signifikan pada CT:**

Usia 51-65 tahun , IMT obesitas , TG , *fat ,* dan WY .