Ulasan dataset skincare:

Data yang digunakan pada project ini adalah data skincare. Data tersebut diperoleh dari google link: https://github.com/agorina91/final\_project/blob/master/Dash\_App\_and\_Models/skindataall.csv.

Dalam data skincare yang digunakan, terdapat 4 independent data, yaitu data customer, data problem, data ingredients dan data tipe produk.

1. Tabel customer: Dalam data customer memiliki 6 kolom, diantaranya kolom id\_cust, nama, jenis\_kelamin, usia, skin\_type, dan pekerjaan. Customer memiliki relationship terhadap data problem dengan hubungan 1:N (satu ke banyak) dengan nama table customer\_ke\_problem.

Untuk kolom jenis kelamin akan dilakukan random generate number 0 dan 1 dengan keterangan 0 = laki-laki dan 1 = perempuan.

Untuk kolom usia akan dilakukan random generate number dari 18-65 tahun.

Untuk kolom pekerjaan akan dilakukan random generate number 0 dan 1 dengan keterangan 0 = karyawan dan 1 = lainnya

1. Tabel problem memiliki 2 kolom yaitu kolom skin\_problem, dan alergi. tabel data problem\_ke\_ingridients dengan hubungan M:N(banyak ke banyak) memiliki 2 kolom, yaitu skin\_problem dan komposisi. Dengan menggunakan data tersebut, akan dibuat riset berdasarkan permasalahan kulit untuk mengetahui kecocokan skincare seseorang berdasarkan ingredients.

Kolom skin\_problem berupa data kategorik

Kolom alergi berupa data kategorik

1. Tabel ingredient memiliki 4 kolom, diantaranya komposisi, merk, harga dan review.

Kolom komposisi berupa data kategorik

Kolom merk/brand berupa data kategorik

Kolom harga berupa numerik dalam dollar yang akan di convert ke rupiah

Kolom review berupa data numerik rating skala 1-5

1. Tabel tipe produk, memiliki 5 kolom, diantaranya id\_kategori, nama\_kategori, tgl\_produksi, tgl\_exp , dan id\_bpom. Tabel problem ke tipe produk memiliki relasi M:N (banyak ke banyak) untuk melihat produk mana yang lebih sesuai untuk digunakan, Contoh permasalahan kulit yaitu bekas jerawat maka tipe produk yang cocok adalah acne spot.

Kolom id\_kategori berupa data integer

Kolom nama\_kategorik berupa data kategorik

Kolom tgl\_produksi akan di generate secara random dengan tahun minimal 2022

Kolom tgl\_exp akan menggunakan data tgl\_produksi ditambah tiga harus.