1. Sila Resha Nugraha (A11.2022.14605)
2. Kemas Armando Ayyub Oktavian Azhari (A11.2020.13025)

* https://github.com/SilaResha16/UAS\_DataMining

Dataset tersebut merupakan daftar produk-produk dari **CV. Putra Andalas** dengan beberapa atribut yang menjelaskan karakteristik masing-masing produk. Setiap baris dalam dataset ini mewakili satu produk, dan kolom-kolomnya memberikan informasi spesifik tentang produk tersebut. Berikut adalah penjelasan rinci untuk setiap kolom:

1. **No**.: Nomor urut produk dalam dataset.
2. **Nama Produk**: Nama produk yang dijual.
3. **Deskripsi**: Deskripsi atau keterangan tentang ukuran atau tipe produk. Contohnya, "Sedang," "Besar," atau "Murah."
4. **Total Pesanan**: Jumlah pesanan untuk produk tersebut.
5. **Harga Barang**: Kategori harga produk, seperti "Murah," "Sedang," atau "Mahal."
6. **Gerindra, Obat, Amplas, Scotch Brite, Packing**: Karakteristik atau atribut produk yang mungkin relevan. Misalnya, untuk produk seperti "Stoel Joey," atribut ini menyatakan ukuran gerinda, obat, amplas, scotch brite, dan packing yang terkait dengan produk tersebut.
7. **Total Harga**: Total harga produk, mungkin tergantung pada beberapa faktor seperti ukuran, kategori harga, dan atribut produk lainnya.

**Permasalahan**:

Dataset yang diberikan, kita bisa menduga bahwa mungkin Anda ingin melakukan klasifikasi terhadap produk-produk berdasarkan atribut-atribut yang diberikan, seperti deskripsi, harga, ukuran, dan atribut lainnya.

**Tujuan Eksperimen**:

Tujuan eksperimen dengan menggunakan algoritma Naive Bayes Classification mungkin termasuk:

* **Klasifikasi Produk**: Mengelompokkan produk-produk ke dalam kategori atau kelas tertentu berdasarkan atribut-atribut yang ada. Contohnya, mungkin ingin mengklasifikasikan produk sebagai "Mahal," "Sedang," atau "Murah" berdasarkan harga barang.
* **Prediksi Berdasarkan Deskripsi**: Jika deskripsi produk memberikan informasi yang dapat memprediksi kategori atau harga produk, Anda mungkin ingin menguji apakah algoritma Naive Bayes dapat memprediksi dengan akurat.
* **Analisis Atribut**: Menilai seberapa penting setiap atribut (seperti gerindra, obat, amplas, scotch brite, dan packing) dalam menentukan kategori atau harga produk. Naive Bayes dapat memberikan perkiraan probabilitas bahwa suatu atribut mempengaruhi kategori produk tertentu.
* **Optimasi Proses Bisnis**: Jika ini merupakan aplikasi bisnis, tujuan eksperimen dapat mencakup meningkatkan efisiensi dalam pengelompokan atau pengelolaan stok berdasarkan hasil klasifikasi.

**Model dan Alur Tahapan Eksperimen:**

Loading Library dan Data:

1. Mengimpor library yang diperlukan seperti pandas, scikit-learn, dan seaborn.
   * Membaca dataset dari file CSV menggunakan pandas.
   * Menampilkan dataset untuk inspeksi.
2. Pembersihan Data:
   * Menghapus baris yang mengandung nilai kosong (NaN).
   * Menghapus kolom 'Nama Produk' karena tampaknya tidak diperlukan untuk analisis ini.
3. Feature Engineering & Data Integration:

* Menggunakan one-hot encoding pada kolom 'Deskripsi' untuk mengonversi fitur kategori menjadi representasi numerik.
* Mengonversi label kategori pada kolom 'Harga Barang' menjadi numerik.
* Mengonversi kolom-kolom berukuran ('Scotch Brite', 'Obat', 'Gerindra', 'Amplas', 'Packing') ke dalam format numerik menggunakan peta besaran.
* Mengonversi label pada kolom 'Total Harga' menjadi numerik menggunakan peta totalharga.

1. Pisahkan Fitur dan Label:
   * Menentukan fitur (X) dan label (y).
   * Fitur (X) adalah semua kolom kecuali 'Total Harga'.
   * Label (y) adalah kolom 'Total Harga'.
2. Pembagian Data untuk Training dan Testing:

* Membagi dataset menjadi data pelatihan (80%) dan data pengujian (20%) menggunakan train\_test\_split dari scikit-learn.

1. Pelatihan Model Naive Bayes:
   * Menginisialisasi model Naive Bayes menggunakan GaussianNB.
   * Melatih model menggunakan data pelatihan.
2. Evaluasi Performa Model:
   * Melakukan prediksi pada data pengujian.
   * Menghitung akurasi, classification report, dan confusion matrix untuk mengevaluasi performa model.
3. Visualisasi Confusion Matrix:
   * Membuat visualisasi heatmap dari confusion matrix.
4. Prediksi dengan Data Acak:
   * Membuat data acak yang mirip dengan format data latihan.
   * Melakukan prediksi pada data acak menggunakan model yang telah dilatih.
5. Tampilkan Hasil Prediksi:
   * Menampilkan hasil prediksi pada data acak, termasuk kolom 'Predicted\_Total\_Harga'.