**PREDIKSI PENJUALAN MOBIL BEKAS MENGGUNAKAN METODE GAUSSIAN NAIVE BAYES CLASSIFIER**



Proyek ini disusun untuk memenuhi tugas akhir mata kuliah Penambangan Data pada Program Sarjana Teknik Informatika Fakultas Ilmu Komputer Universitas Dian Nuswantoro Semester Ganjil Tahun Akademik 2023/2023

Dosen Pengampu

Dr. Aris Marjuni, S. Si, M. Kom

TIM PROYEK:

|  |  |  |
| --- | --- | --- |
| **No.** | **NIM** | **NAMA** |
| 1. | A11.2017.10276 | Sofwan Hidayat |
| 2. | A11.2020.13205 | Athiya Nahdhiana |
| 3. | A11.2019.12030 | Anggito Budhi Prasojo |
| 4. | A11.2021.13605 | Evan faiz |

**ABSTRAKSI PROYEK**

Pasar penjualan mobil bekas kini makin diminati oleh masyarakat dan tidak kalah dengan mobil baru. Pilihan model yang ditawarkan kepada pembeli sangat beragam mulai dari model automatic / manual, mesin dengan bahan bakar diesel / bensin bahkan mesin elektrik . Proyek ini dilakukan dengan tujuan untuk mengklasifikasi model mobil mana yang paling laku terjual dan model mana yang tidak laku.

Klasifikasi dilakukan menggunakan pendekatan penambangan data dengan model Gaussian Naive Bayes Model. Dataset yang digunakan adalah dataset publik yang bersumber dari Kaggle.com. Hasil klasifikasi menunjukkan bahwa Gaussian Naive Bayes Model mampu mengidentifikasi potensi penjualan mobil bekas dengna akurasi 52%.

**A. DATASET**

Nama dataset:

bmw.xlsx

Sumber:

https://www.kaggle.com/datasets/adityadesai13/used-car-dataset-ford-and-mercedes

Informasi dataset:

10781 observations / records

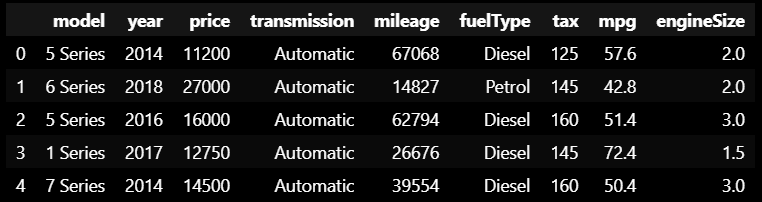
9 attributes

Informasi Atribut:

1. model : model mobil;
2. year : tahun produksi mobil;
3. price : harga mobil;
4. transmission : tipe transmisi mesin mobil;
5. mileage : ukuran mile mobil;
6. fuelType : tipe bahan bakar pada mobil;
7. tax : pajak mobil;
8. mpg : ukuran mile per galon bahan bakar;
9. engineSize : ukuran mesin mobil;

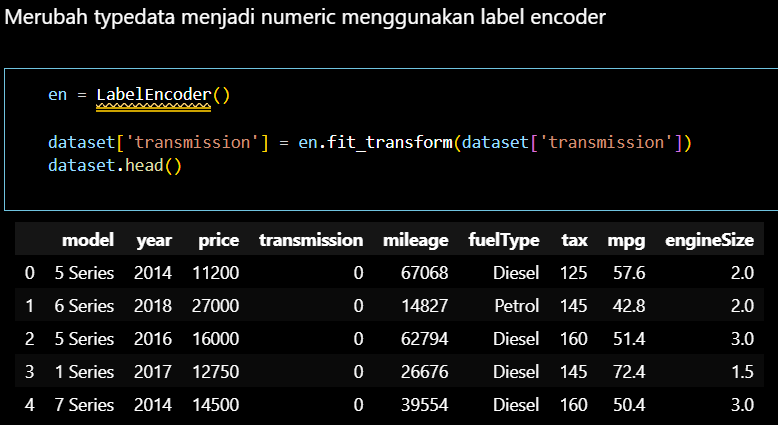
Contoh instances:

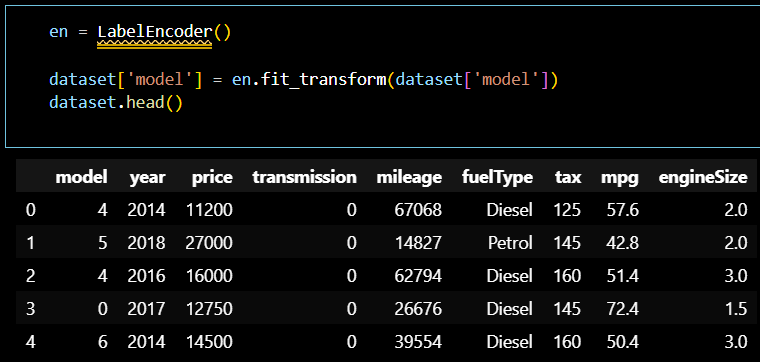
Berikut tampilan 5 data instances pertama :

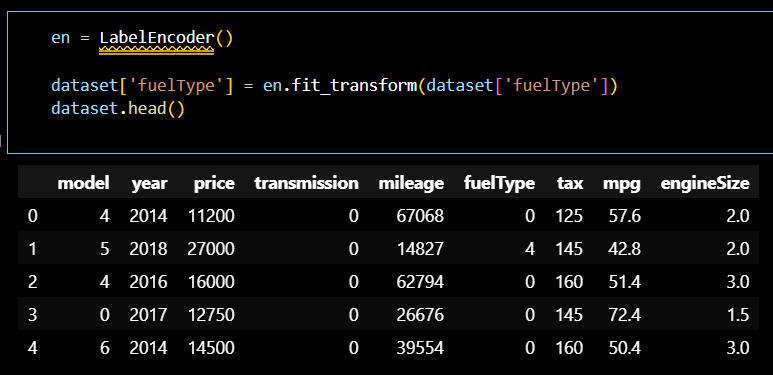


**B. DATA PREPROCESSING**

Transformasi Data (Data Transformation), data akan diubah atau ditransformasikan menjadi bentuk yang sesuai dengan metode analisis yaitu tipe data numeric. Label encoding untuk mengonversi label kata menjadi angka. Label encoding mengacu pada proses transformasi label kata menjadi bentuk numerik. Dalam hal regresi jika memuat variabel kategori dan nilainya tidak bisa difaktorisasi dalam bentuk tingkatan, dilakukan proses dummy, setiap nilai dalam variabel itu menjadi variabel lain. Data preprocessing yang dilakukan terhadap dataset bertujuan agar dataset lebih mudah untuk diproses menggunakan metode gaussian naive bayes.

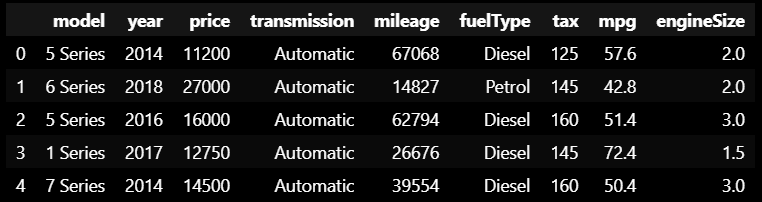




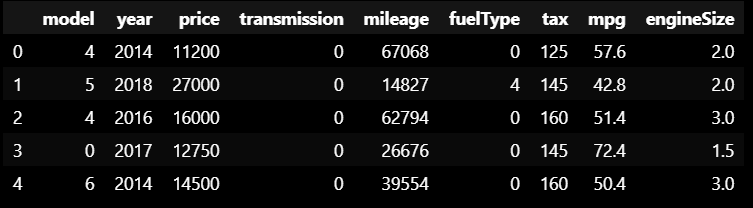


Berikut merupakan keadaan dataset sebelum dan sesudah preprocessing :

Sebelum Preprocessing :



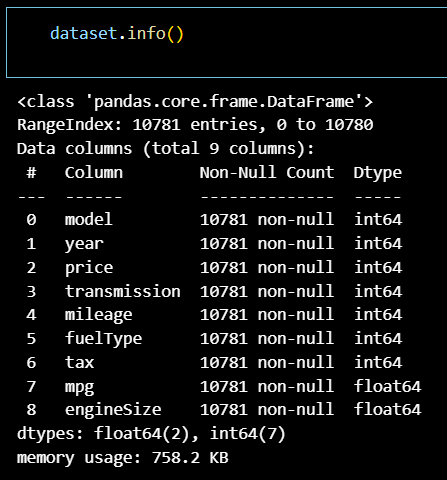
Sesudah Preprocessing :



**C. EXPLORATORY DATA ANALYSIS (EDA)**

Berikut merupakan langkah-langkah EDA terhadap dataset dan capture hasilnya , yaitu:

* Deskripsi/informasi dataset : menggunakan referensi info()



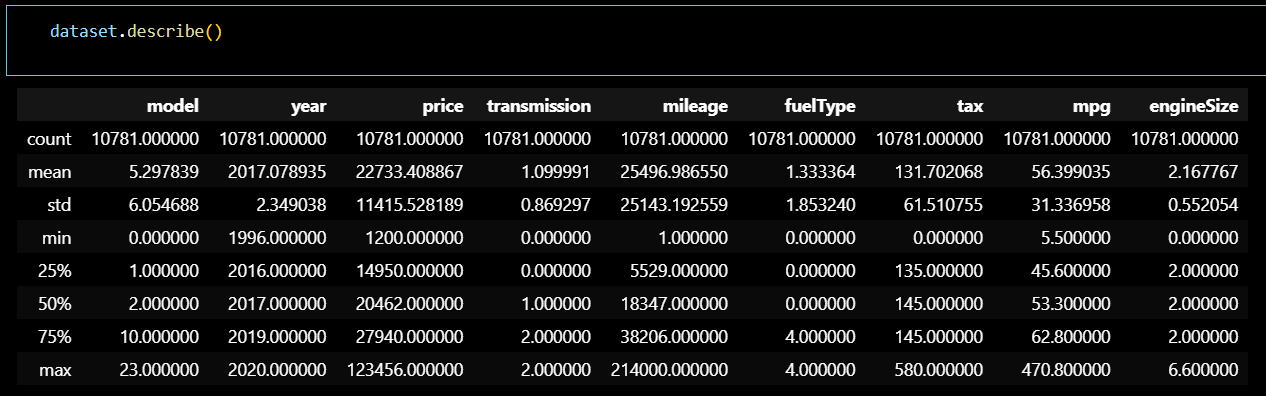
* Capture 10 data secara random : menggunakan referensi sample()



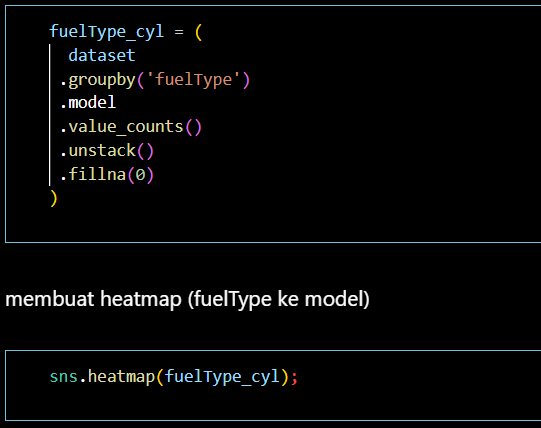
* Periksa apakah ada data yang kosong (missing) : menggunakan referensi isnull()

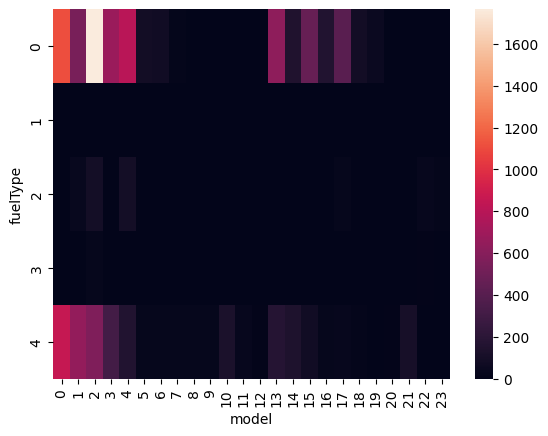


* Tampilkan informasi statistik dari dataset tersebut : menggunakan referensi describe()



* Plot correlogram (heatmap) : menggunakan referensi heatmap()

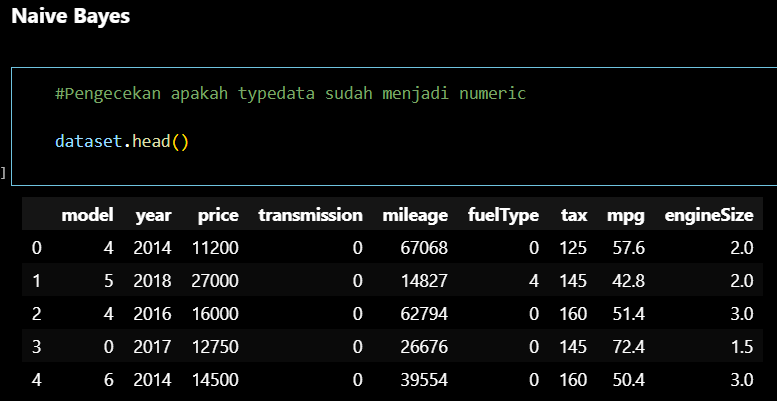




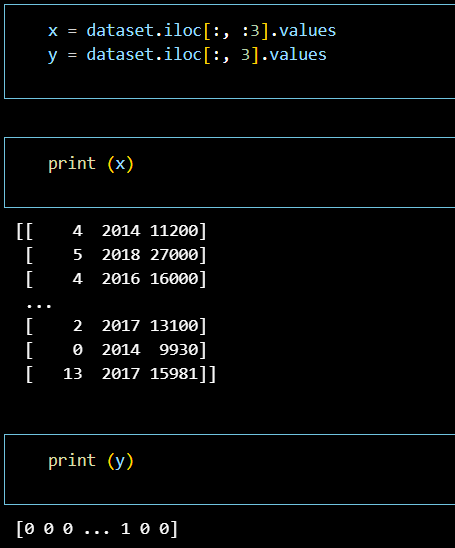
**D. PEMODELAN**

Model prediksi Gaussian Naive Bayes adalah teknik dalam machine learning yang menggunakan pendekatan probabilitas dan distribusi Gaussian atau distribusi normal.

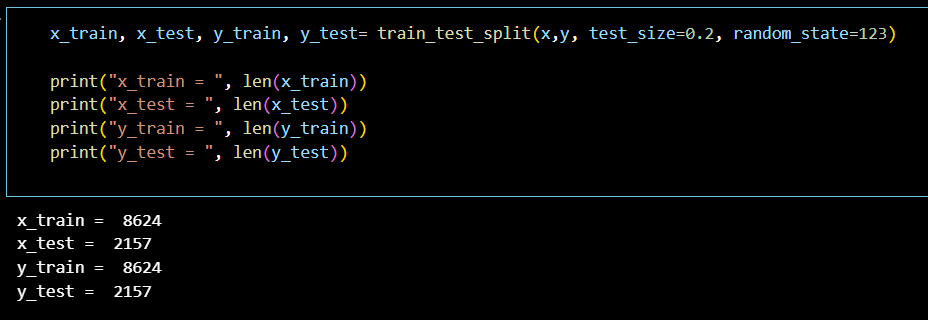
* Implementasi model Gaussian Naive Bayes pada dateset yang sudah dipreprocessing menjadi numerical value



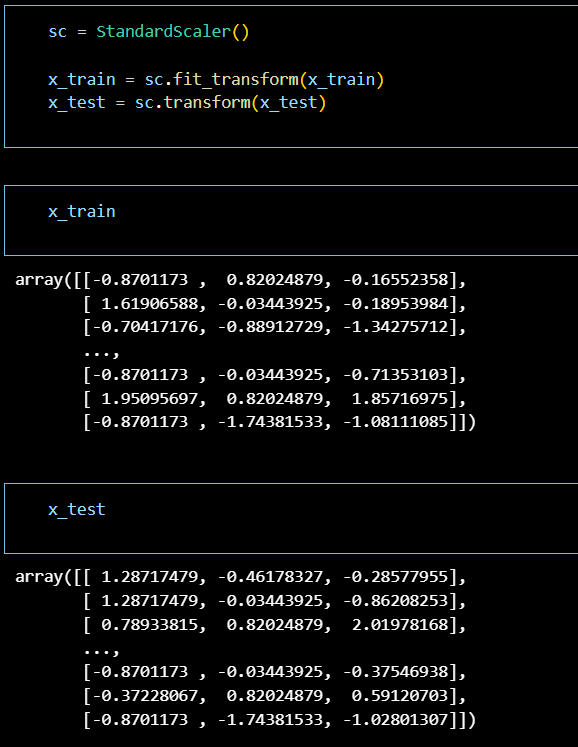
* Pendefinisian data variabel bebas dan tidak bebas

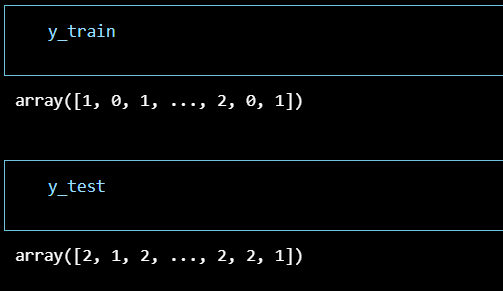


* Pendefinisian data training dan testing

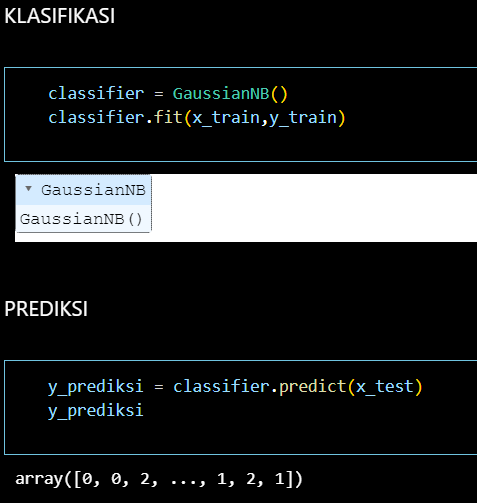


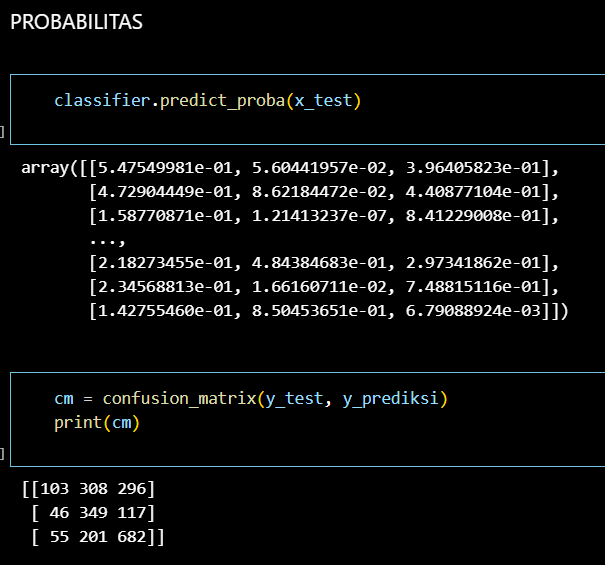
Standarisasi fitur dengan menghapus rata-rata dan penskalaan ke varians satuan. Ini berarti kita mengubah fitur sehingga memiliki rata-rata 0 dan varians 1.

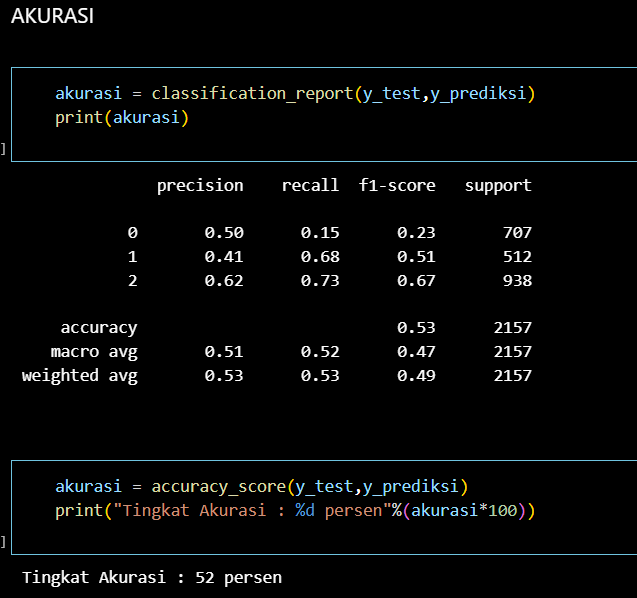




* Hasil eksperimen dan pengukuran/evaluasi model









**E. KESIMPULAN**

Naive Bayes adalah salah satu algoritma yang sering digunakan dalam pengenalan pola, analisis teks, klasifikasi dokumen, dan banyak aplikasi lainnya. Algoritma Naive Bayes mengandalkan asumsi dasar yang cukup sederhana, yaitu bahwa semua atribut (fitur) yang digunakan dalam klasifikasi adalah independen satu sama lain. Oleh karena itu, istilah "naive" digunakan, karena dalam dunia nyata, atribut seringkali tidak benar-benar independen. Namun, asumsi ini mempermudah perhitungan matematis dan sering kali menghasilkan hasil yang cukup baik, terutama dalam kasus klasifikasi teks. Rumus dasar Naive Bayes adalah berdasarkan pada Teorema Bayes, yang menghubungkan probabilitas suatu peristiwa dengan probabilitas peristiwa lain yang berhubungan. Dalam konteks klasifikasi, Naive Bayes digunakan untuk menghitung probabilitas bahwa sebuah sampel data termasuk dalam suatu kelas tertentu.

**F. KONTRIBUSI ANGGOTA**

Tuliskan persentase kontribusi masing-masing anggota dalam pekerjaan tugas ini. Kontribusi diukur dari peran aktif anggota, khususnya dalam kerja sama tim untuk menyelesaikan proyek ini.

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **NIM** | **NAMA** | **KONTRIBUSI DAN KEAKTIFAN (%)** |
| 1. | A11.2017.10276 | Sofwan Hidayat | 25% |
| 2. | A11.2020.13205 | Athiya Nahdhiana | 25% |
| 3. | A11.2019.12030 | Anggito Budhi Prasojo | 25% |
| 4. | A11.2021.13605 | Evan faiz | 25% |