1. **Eksplorasi Data**

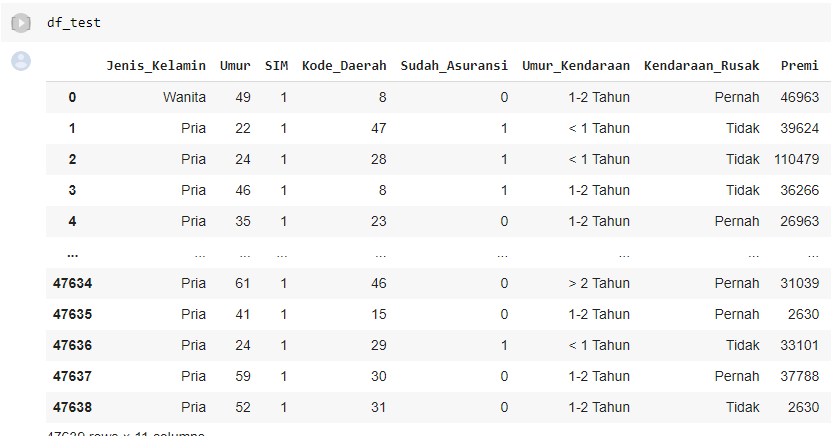
Terdapat beberapa tahap yang kami lakukkan dalam eksplor data Kendaraan, antara lain sebagai beriku :

1. **Upload dataset yang akan digunakan**

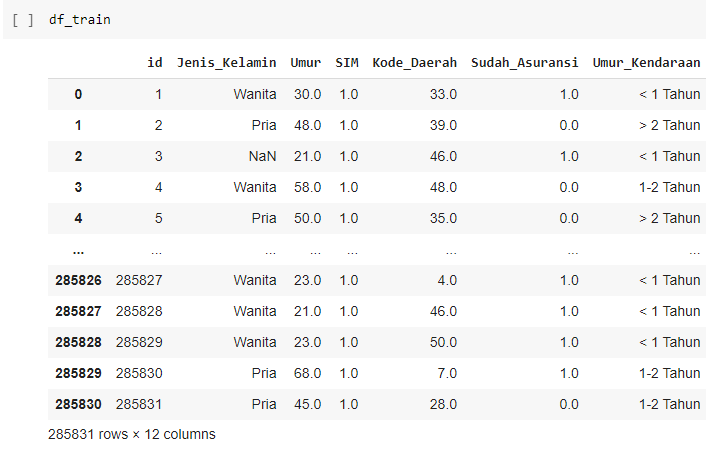


Pada proses ini, dataset yang digunakan adalah dataset kendaraan, yang dibagi menjadi dataset test dan dataset train.

1. **Keluaran dari seluruh dataset**

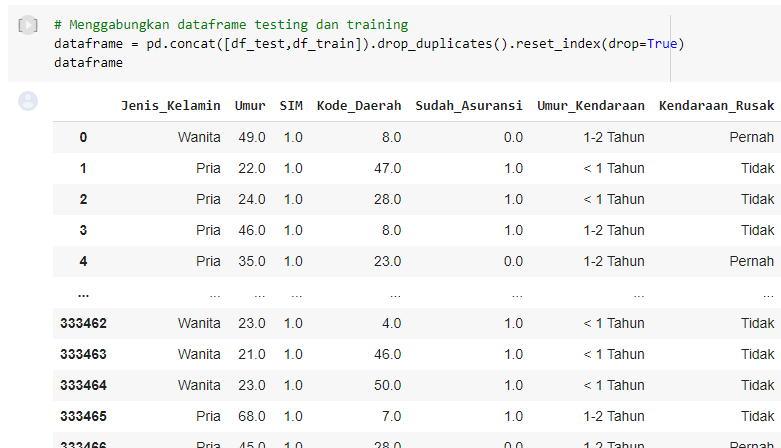








1. Menggabungkan data test dan data train menjadi satu dataframe, untuk memudahkan proses preprocessing pada tahap selanjutnya

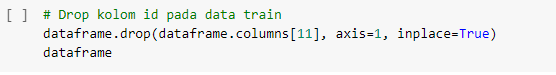




1. **Data Preprocessing**

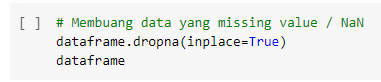
Pada taap Data Preprocessing, kami melakukkan beberapa tahap yaitu, drop kolom id, lalu clear missing value / NaN, lalu mengubah data kategorikal mejadi numerical.

1. **Drop kolom id**



Code diatas merupakan suatu kode agar suatu column di drop, column yang di drop diatas adalah column 11, yaitu kolom id.

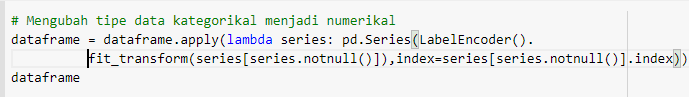
1. **Clear missing value**



Code diatas merupakan code yang memerintahkan agar melakukkan clear suatu data yang missing. Tujuannya untuk mengatasi data outlier.

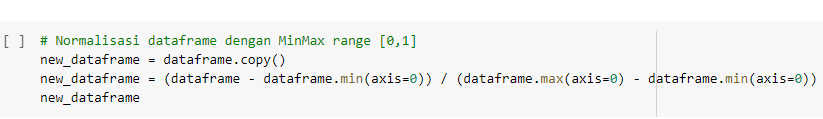
1. **Mengubah data kategorikal menjadi numerical**

Tahap ini dilakukkan agar data tersebut dapat di plotting, karena apabila ada data yang bertipe kategorikal dan numerical, maka proses plotting akan sulit dilakukkan.



1. **Normalisasi Data**

Untuk normalisasi data yang kami gunakan adalah Min Max

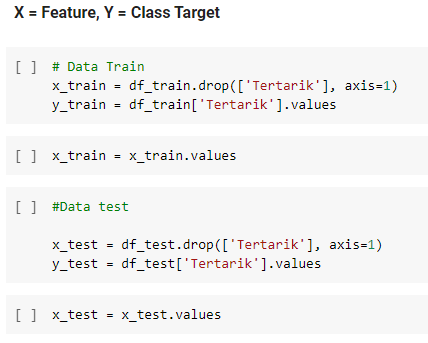


Normalisasi data ini dilakukan untuk membuat semua value yang ada pada dataframe berkisar antara 0-1.

1. **Pemodelan**

Pada tahap ini metode classification yang kami gunakan yaitu metode Naïve Bayes dan Decision Tree. Naïve Bayes merupakan sebuah metode klasifikasi berdasarkan probabilitas sederhana dan dirancang untuk dipergunakan dengan asumsi bahwa antar satu kelas dengan kelas yang lain tidak saling tergantung (independen). Pada klasifikasi Naïve Bayes, proses pembelajaran lebih ditekankan pada mengestimasi probabilitas. Sedangkan Decision Tree merupakan metode klasifikasi menggunakan struktur pohon atau struktur berhirarki. Konsep dari pohon keputusan adalah mengubah data menjadi decision tree dan aturan-aturan keputusan.

1. **Eksperimen**

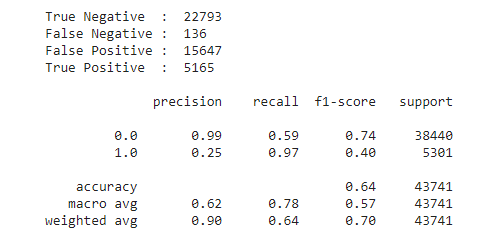
****

Untuk eksperimen, kami membagi data train dan data test menjadi 2 bagian masing-masingnya, yaitu feature dan class target untuk data train, feature dan class target untuk data test

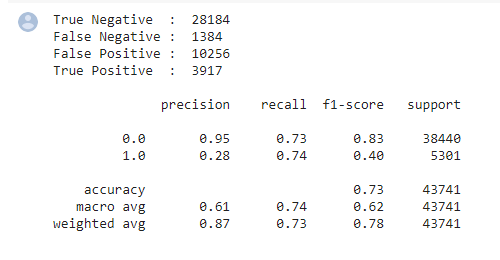
1. **Evaluasi**

Hasil evaluasi setiap metode sebagai berikut:

1. Naïve Bayes

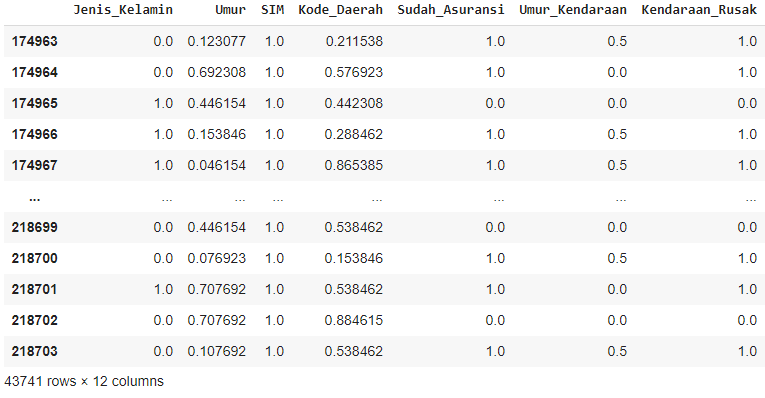


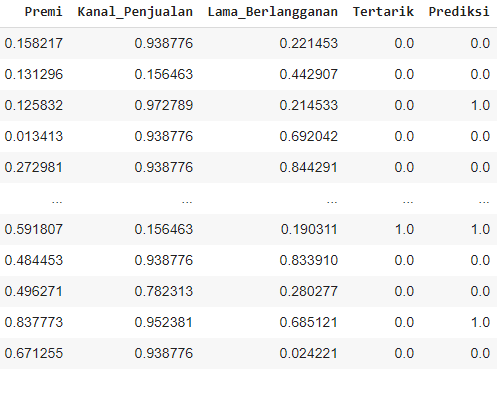
1. Decision Tree



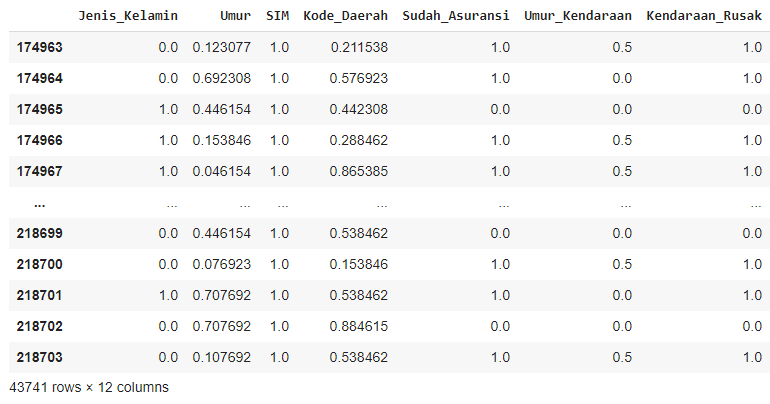
Kesimpulan: dapat dilihat bahwa Algoritma Decision Tree memiliki nilai akurasi lebih tinggi yaitu 0.73 jika dibandingkan dengan nilai akurasi yang dimiliki oleh algoritma Naïve Bayes yaitu 0.64.

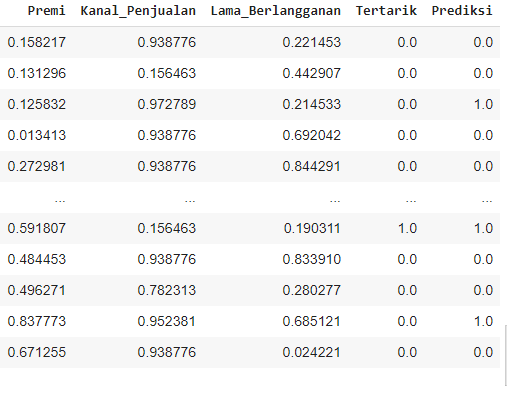
1. Naïve Bayes





1. Decision Tree





Namun walaupun nilai dari masing-masing precision, recall dan f1-score berbeda yang menghasilkan sebuah akurasi yang berbeda, namun prediksi kedua metode menunjukkan kesamaan. Bisa dilihat dari dataframe dengan kolom yang ditambahkan yaitu kolom prediksi.

**LINK GITHUB**

<https://colab.research.google.com/drive/1G7snLs9DpZuX76M3U89KgZAjSc_OBHsm?usp=sharing>