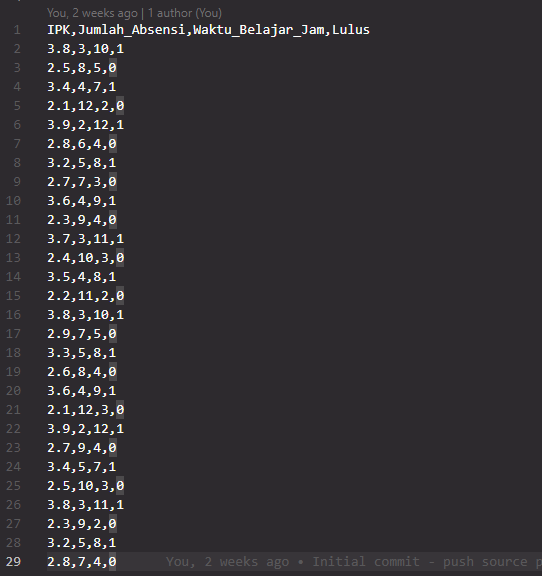
Nama : Fawzy Imam Hamdani Gunamukti

NIM : 231011401030

KLS : 05TPLE017

Penjelasan Machine Learning Pertemuan 4

Di siapkan berupa data set mahasiswa



Pada awalnya data yang di sediakan hanya 10 data, agar mendukung proses split dan akurasi data di perbanyak menjadi 200 data.

1. Membaca dan memuat data CSV ke dalam DataFrame menggunakan pandas.

2. Melakukan eksplorasi awal data dengan menampilkan struktur, tipe

data, dan contoh isi data.

3. Memeriksa nilai kosong (missing values) dan menghapus data duplikat

untuk membersihkan data.

4. Melakukan visualisasi awal dengan seaborn:

- Boxplot untuk mendeteksi outlier pada kolom IPK.

- Histogram untuk melihat distribusi nilai IPK.

- Scatterplot untuk melihat hubungan antara IPK dan Waktu Belajar

berdasarkan status kelulusan.

- Heatmap untuk melihat korelasi antar variabel numerik.

5. Menambahkan fitur baru (feature engineering):

- Rasio\_Absensi = Jumlah\_Absensi / 14

- IPK\_x\_Study = IPK \* Waktu\_Belajar\_Jam

6. Menyimpan hasil data yang sudah diproses ke file CSV baru.

7. Menyiapkan data untuk machine learning dengan memisahkan fitur (X)

dan target (y).

8. Membagi dataset menjadi tiga bagian:

- 70% untuk data training

- 15% untuk data validation

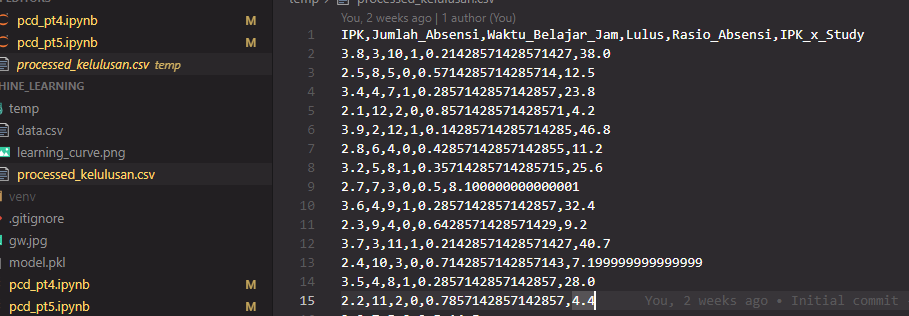
- 15% untuk data testing

9. Menampilkan ukuran data di masing-masing set (train, val, test).

Kesimpulan: Kode ini menjalankan proses lengkap mulai dari membaca,

membersihkan, memvisualisasikan, membuat fitur baru, hingga menyiapkan

data untuk pelatihan model machine learning.



Ini Adalah file hasil dari kode Pertemuan 4 yang sudah di cleanup.