

Tugas 1: Membaca Dataset

Rafa Al Razzak - 0110224155

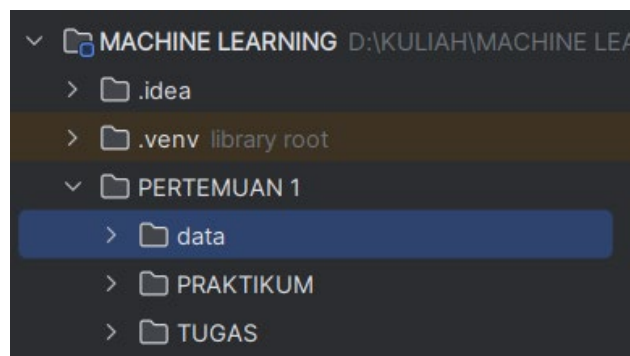
Teknik Informatika, STT Terpadu Nurul Fikri, Depok

E-mail: 0110224155@studetnt.nurulfikri.ac.id

Abstract. Membaca sebuah dataset CSV menggunakan bahasa pemrograman Python dengan Jupyter Notebooks

1. Membuat Struktur Folder

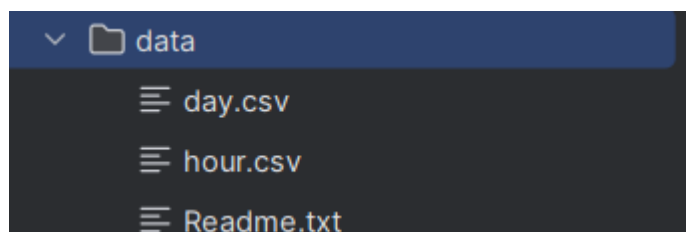
Agar file lebih terorganisasi dan mudah digunakan, disarankan untuk membuat struktur folder seperti pada Gambar 1.



Gambar 1 Struktur Folder

2. Memasukan File Data

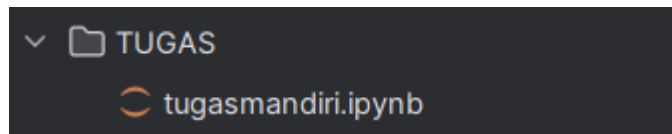
Download file data dari sumber yaitu Elena Nurul Fikri kelas *Machine Learning* masukan dalam folder data agar rapi.



Gambar 2 Memasukan File Data

3. Membuat File Notebook

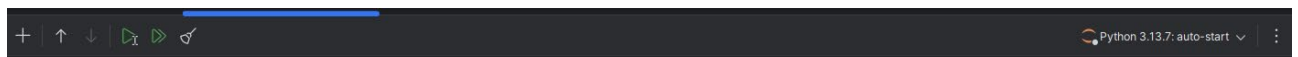
Pada Jupyter, file *notebook* berekstensi .ipynb, buat dalam folder tugas, dengan nama tugasmandiri.ipynb.



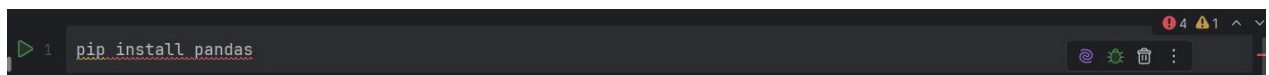
Gambar 3 Buat File Notebook

4. Membuat Kode Program

Kita akan membaca sebuah dataset yang bersumber pada folder data dan file hour.csv, kita perlu menginstall perpustakaan yang bernama *pandas*. Untuk menginstall *pandas*, kita perlu membuat sel dengan cara menekan tombol tambah pada atas kiri dan mengetik *pip install* pada sel kosong tersebut.



Gambar 4 Buat Sel Baru

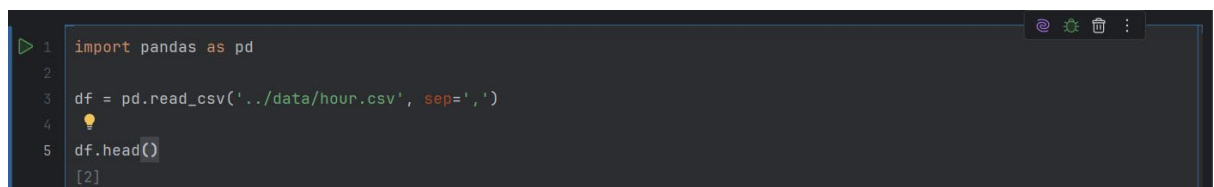


Gambar 5 Perintah Untuk Install Pandas

Setelah mengetikkan perintah tersebut, kita bisa menekan tombol mulai pada bagian kiri sel yang telah kita isi perintah untuk menginstall *pandas*.

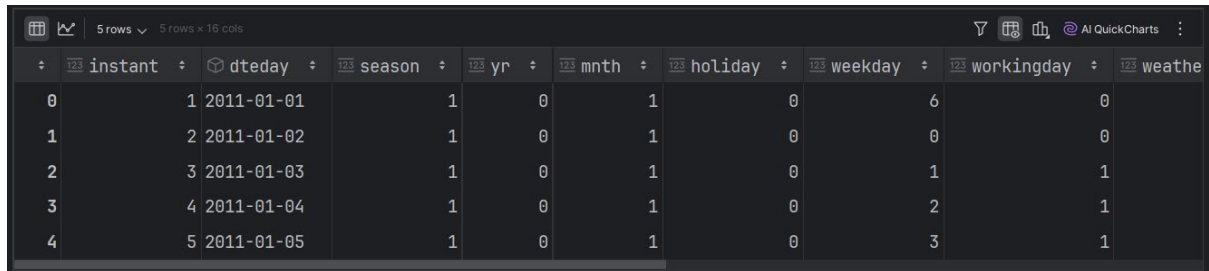
Setelah selesai menginstall *pandas*, kita bisa membuat sel baru dan memasukkan kode Python sesuai dengan tugas yang diberikan.

Pada baris pertama merupakan perintah untuk mengimport perpustakaan *pandas* sebagai variabel *pd*, selanjutnya kita buat sebuah variable dengan nama *df*, assign variable tersebut dengan hasil dari pembacaan pada file *../data/hour.csv*, setelah itu kita tampilkan 5 hasil pertama dari variable *df* dengan perintah *df.head()*



Gambar 6 Kode Program Untuk Menampilkan Dataset

Akan tampil sebuah tabel dengan jumlah 5 jumlah data sesuai dengan yang kita panggil, dengan kolom *index*, *instant*, *teday*, *season*, *yr*, *mnth*, *hr*, *holiday*, *weekday*, *workingday*, *weathersit*, *temp*, *atemp*, *hum*, *windspeed*, *casual*, *registered*, *cnt*. Data yang ditampilkan bervariasi sesuai dengan data pada dataset *hour.csv*



	instant	dteday	season	yr	mnth	holiday	weekday	workingday	weathe
0	1	2011-01-01	1	0	1	0	6	0	
1	2	2011-01-02	1	0	1	0	0	0	
2	3	2011-01-03	1	0	1	0	1	1	
3	4	2011-01-04	1	0	1	0	2	1	
4	5	2011-01-05	1	0	1	0	3	1	

Gambar 7 Hasil Dari Kode Program

5. Kesimpulan

Dataset hour.csv berhasil dibaca menggunakan pustaka pandas pada Python dan ditampilkan dalam bentuk DataFrame.

Hasil tersebut menunjukkan bahwa kode yang digunakan berjalan sesuai harapan dan data dapat diolah dengan jelas serta terstruktur.

Link GitHub: <https://github.com/sttnf/machine-learning>