Tugas 3: Praktikum Mandiri 3 Machine Learning

**Hisyam Wildan Alfath - 0110222206**

Teknik Informatika, STT Terpadu Nurul Fikri, Depok

E-mail: [hisy22206ti@student.nurulfikri.ac.id](mailto:hisy22206ti@student.nurulfikri.ac.id)

**Abstract.** Praktikum ini melatih kemampuan membaca, mengolah, dan membagi dataset menggunakan pandas dan scikit-learn di lingkungan Python. Mulai dari membaca file CSV yang berisi data kompleks terkait atribut cuaca dan pengguna, kemudian membagi data menjadi subset training, validation, dan testing untuk persiapan pemodelan machine learning. Visualisasi dan eksplorasi data menggunakan statistik deskriptif serta grafik juga dilakukan untuk memahami karakteristik data. Praktikum mandiri memperdalam pengalaman dengan dataset nyata yang lebih besar dan beragam, serta mengaplikasikan teknik pembagian data secara mandiri agar diperoleh model yang siap diuji dan dioptimalkan. Keseluruhan proses ini menyiapkan dasar kuat dalam alur kerja analisis data dan machine learning yang efektif dan terstruktur.

1. Praktikum Mandiri 03

* Menghubungkan Colab dengan Google Drive

A screen shot of a computer

AI-generated content may be incorrect.

Kode ini memanggil modul drive dari google.colab untuk menghubungkan (mount) Google Drive ke lingkungan Google Colab, dan outputnya adalah muncul pesan “Mounted at /content/drive” yang menandakan folder Google Drive sekarang sudah bisa diakses pada path tersebut di Colab.

* Membaca file CSV dari Google Drive di Colab

A screenshot of a computer

AI-generated content may be incorrect.

Kode ini menggunakan library pandas untuk membaca file CSV yang tersimpan di Google Drive ke dalam sebuah DataFrame. Path file pada parameter pd.read\_csv() sudah menggunakan direktori /content/drive/MyDrive/..., yang artinya folder Google Drive sudah berhasil di-mount pada path tersebut di Colab. Setelah pemanggilan fungsi df.head(), ditampilkan 5 baris awal dari data CSV yang berisi berbagai kolom seperti instant, dteday, season, dan lainnya.

* Membaca file csv menggunakan pandas

A black and orange background

AI-generated content may be incorrect.

Kode ini mengimpor library pandas sebagai pd, lalu menggunakan fungsi pd.read\_csv untuk membaca file CSV dari Google Drive berdasarkan path yang diberikan. Outputnya adalah DataFrame yang menampilkan data dari file CSV tersebut, berupa tabel dengan kolom Gender, Height, Weight, dan Index serta berisi 500 baris data.

* Penjelasan Pemilihan Fitur dan Target



* Variabel X diisi dengan beberapa kolom dari DataFrame yaitu: season, yr, mnth, holiday, weekday, workingday, weathersit, temp, atemp, hum, dan windspeed. Kolom-kolom ini berfungsi sebagai variabel input yang akan digunakan untuk memprediksi hasil.​
* Variabel y diisi dengan kolom cnt, yang merupakan variabel target atau hasil yang akan diprediksi oleh model.
* Penjelasan Split Data Menggunakan train\_test\_split



* Library train\_test\_split dari modul sklearn.model\_selection diimpor untuk membagi dataset menjadi dua bagian, yaitu data latih (training) dan data uji (testing).​
* Perintah train\_test\_split(X, y, test\_size=0.2, random\_state=42) akan membagi data input (X) dan target (y) sehingga 80% data digunakan untuk pelatihan (training) dan 20% untuk pengujian (testing).​
* Parameter random\_state=42 digunakan agar proses pembagian data ini konsisten setiap kali script dijalankan, sehingga hasil pembagian data tidak berubah-ubah
* Melatih Model Regresi Linear

**A screenshot of a computer program

AI-generated content may be incorrect.**

Kode ini menggunakan library scikit-learn untuk membuat dan melatih model regresi linear dengan data latih. Input dari proses ini adalah fitur-fitur (X\_train) sebagai variabel independen dan y\_train sebagai target, sedangkan outputnya adalah model LinearRegression yang sudah terlatih dan siap digunakan untuk prediksi jumlah berdasarkan data fitur.

* Menghitung dan Menampilkan Evaluasi Model

A screenshot of a computer program

AI-generated content may be incorrect.

Kode ini menggunakan library scikit-learn untuk melakukan prediksi pada data uji dan menghitung nilai evaluasi berupa MAE, MSE, dan R2. Input berupa data fitur uji (X\_test) dan target uji (y\_test), sedangkan outputnya adalah nilai MAE sebesar 617.39, MSE sebesar 691035.80, serta skor R2 sebesar 0.827, yang menampilkan performa model hasil prediksi terhadap data uji.

* Visualisasi Hasil Prediksi Regresi Linear

A screen shot of a computer screen

AI-generated content may be incorrect.

Kode ini memanfaatkan matplotlib untuk menampilkan grafik sebar antara nilai aktual dan prediksi hasil regresi linear. Input grafik adalah data aktual (y\_test) dan hasil prediksi model (y\_pred), sedangkan outputnya berupa plot scatter biru untuk membandingkan prediksi vs aktual dan garis merah sebagai garis ideal jika prediksi benar-benar akurat.

Kesimpulannya, model regresi linear yang telah dibuat mampu memprediksi nilai dengan cukup baik karena hasil prediksi mendekati nilai aktual. Dari visualisasi grafik, semakin dekat titik biru pada garis merah, maka model semakin akurat dalam melakukan prediksi. Evaluasi menggunakan metrik MAE, MSE, dan R2 juga menunjukkan performa yang cukup baik sehingga model ini sudah bisa digunakan untuk memprediksi data sejenis dengan hasil yang mudah dipahami.

Link Praktikum 03:

<https://github.com/HisyamWildan/TI03_HisyamW.A_0110222206/blob/main/Pertemuan%202/Praktikum%20dan%20Tugas%20Mandiri%2002/notebooks/praktikum02.ipynb>

Tugas Mandiri 03: