

Tugas 3: Praktikum Mandiri 3 Machine Learning

Hisyam Wildan Alfath - 0110222206

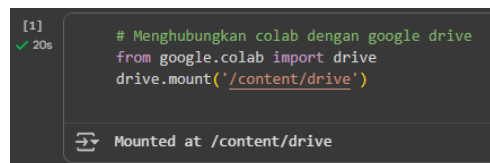
Teknik Informatika, STT Terpadu Nurul Fikri, Depok

E-mail: hisy22206ti@student.nurulfikri.ac.id

Abstract. Praktikum ini menggunakan algoritma regresi linear untuk memprediksi jumlah peminjaman sepeda berdasarkan berbagai faktor seperti musim, hari, cuaca, dan suhu. Data dibagi menjadi data latih dan data uji, lalu model dilatih menggunakan library scikit-learn. Hasil evaluasi menunjukkan model mampu memberikan prediksi yang cukup akurat, ditunjukkan dengan nilai error yang rendah dan grafik prediksi yang mendekati data aktual.

1. Praktikum Mandiri 03

- Menghubungkan Colab dengan Google Drive



```
[1]
✓ 20s
# Menghubungkan colab dengan google drive
from google.colab import drive
drive.mount('/content/drive')

Mounted at /content/drive
```

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

```
[2]
# Membaca file csv menggunakan pandas
import pandas as pd

df = pd.read_csv("/content/drive/MyDrive/Hisyam Wildan Alfath/Semester 7/Machine Learning/Pertemuan 3/Praktikum dan Tugas Mandiri 03/data/day.csv")
df.head()
```

| | instant | dteday | season | yr | mnth | holiday | weekday | workingday | weathersit | temp | atemp | hum | windspeed | casual | registered | cnt |
|---|---------|------------|--------|----|------|---------|---------|------------|------------|----------|----------|----------|-----------|--------|------------|------|
| 0 | 1 | 2011-01-01 | 1 | 0 | 1 | 0 | 6 | 0 | 2 | 0.344167 | 0.363625 | 0.805833 | 0.160446 | 331 | 654 | 985 |
| 1 | 2 | 2011-01-02 | 1 | 0 | 1 | 0 | 0 | 0 | 2 | 0.363478 | 0.353739 | 0.696087 | 0.248539 | 131 | 670 | 801 |
| 2 | 3 | 2011-01-03 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0.196364 | 0.189405 | 0.437273 | 0.248309 | 120 | 1229 | 1349 |
| 3 | 4 | 2011-01-04 | 1 | 0 | 1 | 0 | 2 | 1 | 1 | 0.200000 | 0.212122 | 0.590435 | 0.160296 | 108 | 1454 | 1562 |
| 4 | 5 | 2011-01-05 | 1 | 0 | 1 | 0 | 3 | 1 | 1 | 0.226957 | 0.229270 | 0.436957 | 0.186900 | 82 | 1518 | 1600 |

Next steps: [Generate code with df](#) [New interactive sheet](#)

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

```
[3]
# Membaca file csv menggunakan pandas
import pandas as pd

df = pd.read_csv("/content/drive/MyDrive/Hisyam Wildan Alfath/Semester 7/Machine Learning/Pertemuan 2/Praktikum dan Tugas Mandiri 02/data/500_Person_Gender_Height_Weight_Index.csv")
df
```

| | Gender | Height | Weight | Index |
|-----|--------|--------|--------|-------|
| 0 | Male | 174 | 96 | 4 |
| 1 | Male | 189 | 87 | 2 |
| 2 | Female | 185 | 110 | 4 |
| 3 | Female | 195 | 104 | 3 |
| 4 | Male | 149 | 61 | 3 |
| ... | ... | ... | ... | ... |
| 495 | Female | 150 | 153 | 5 |
| 496 | Female | 184 | 121 | 4 |
| 497 | Female | 141 | 136 | 5 |
| 498 | Male | 150 | 95 | 5 |
| 499 | Male | 173 | 131 | 5 |

500 rows x 4 columns

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

```
[3]
X = df[['season', 'yr', 'mnth', 'holiday', 'weekday', 'workingday', 'weathersit', 'temp', 'atemp', 'hum', 'windspeed']]
y = df['cnt']
```

- 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

```
[4] ✓ 1s from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

- 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

```
[5] ✓ 0s from sklearn.linear_model import LinearRegression
model = LinearRegression()
model.fit(X_train, y_train)
```

LinearRegression

LinearRegression()

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

```
[6] ✓ 0s
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score

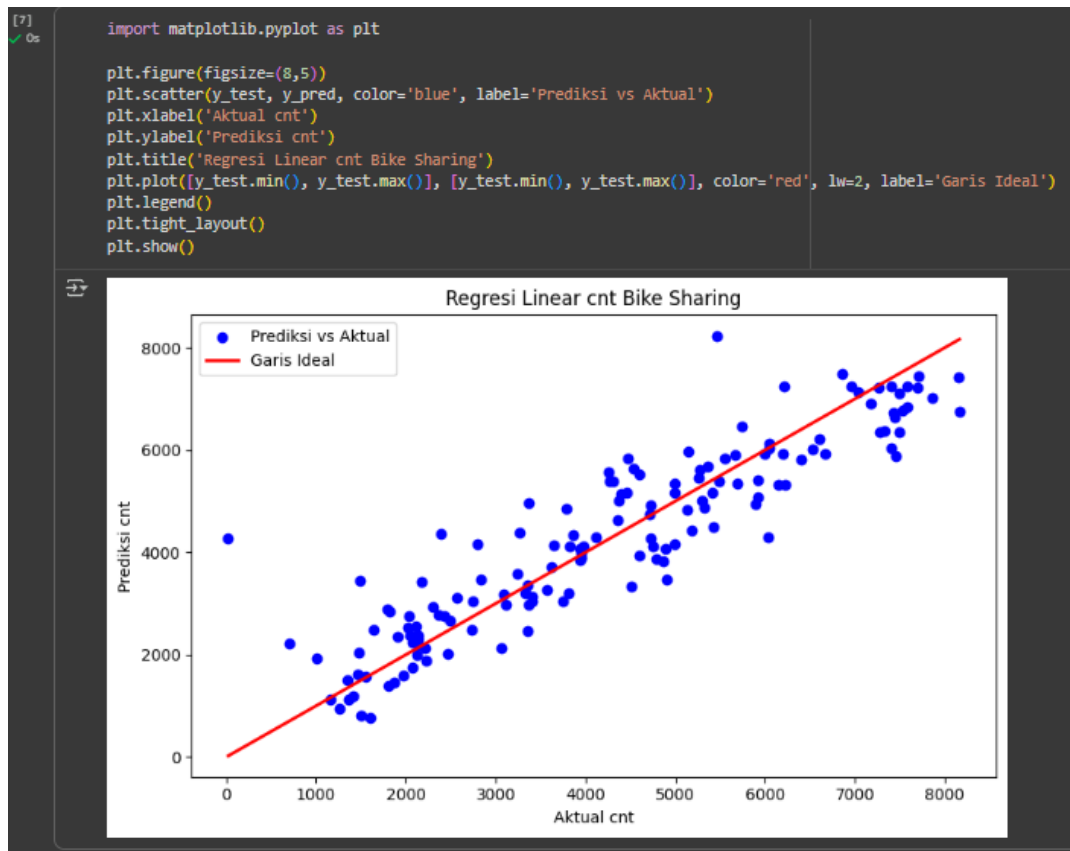
# Prediksi pada data test
y_pred = model.predict(X_test)

# Hitung metrik evaluasi
mae = mean_absolute_error(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
print('MAE:', mae)
print('MSE:', mse)
print('R2:', r2)
```

MAE: 617.3930656443376
MSE: 691035.0082022651
R2: 0.8276670090367212

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



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%203/Praktikum%20dan%20Tugas%20Mandiri%2003/notebooks/praktikum03.ipynb

Tugas Mandiri 03:

https://github.com/HisyamWildan/TI03_HisyamW.A_0110222206/blob/main/Pertemuan%203/Praktikum%20dan%20Tugas%20Mandiri%2003/notebooks/latihan03.ipynb