

Laporan hasil praktikum mandiri

Nama Mahasiswa: Muhammad Faqih Rayya

E-mail: 0110224212@student.nurulfikri.ac.id

Link G : <https://github.com/faqihrayya0403/Machine-Learning>

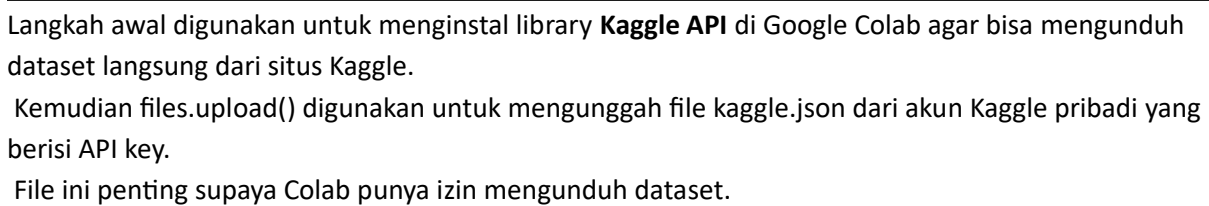
Abstrak

Praktikum ini membahas penerapan **Multiple Linear Regression** untuk memprediksi jumlah penyewaan sepeda harian menggunakan dataset *Bike Sharing Dataset* dari Kaggle. Tujuan utamanya adalah memahami hubungan antara variabel suhu (*temp* dan *atemp*), kelembapan (*hum*), serta kecepatan angin (*windspeed*) terhadap jumlah sepeda yang disewa (*cnt*). Proses analisis dilakukan menggunakan Python di Google Colab, dimulai dari instalasi Kaggle API, pengunduhan dataset, pemrosesan data, hingga evaluasi model. Hasil pengujian menunjukkan bahwa model mampu menjelaskan sekitar **49,9% variasi data penyewaan sepeda**, dengan nilai **Mean Squared Error (MSE) sebesar 2.007.059,49** dan **R² sebesar 0,499**. Dapat disimpulkan bahwa suhu memiliki pengaruh positif terbesar terhadap peningkatan jumlah penyewaan sepeda, sedangkan kelembapan dan kecepatan angin memiliki efek negatif. Model ini dapat dikembangkan lebih lanjut dengan menambahkan fitur lain seperti kondisi cuaca, hari libur, atau musim.

Kata kunci: Multiple Linear Regression, Bike Sharing Dataset, Prediksi, Python, Analisis Data

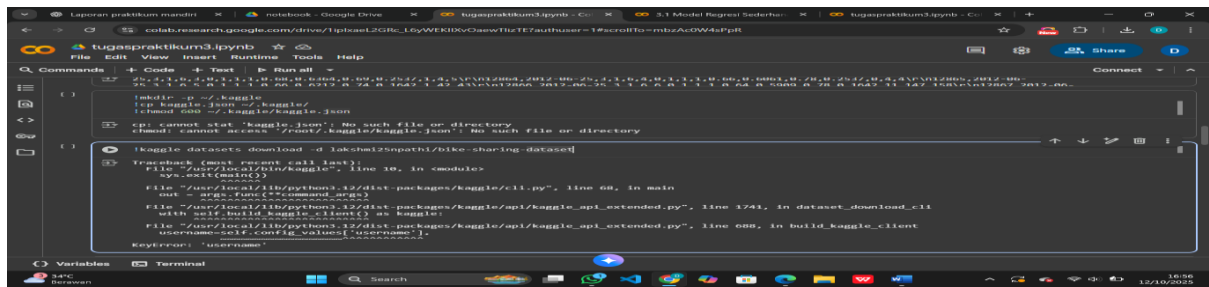
A. Tugas Praktikum 3 – Multiple Linear Regression (Bike Sharing Dataset)

1. Instalasi dan Setup Kaggle API

[illegible]

Namun, dari hasil log, file kaggle.json belum diunggah dengan benar, jadi langkah ini gagal dijalankan.

3. Mengunduh Dataset



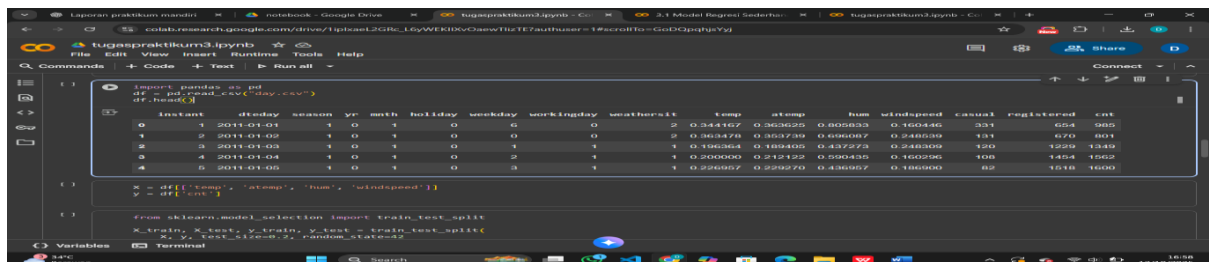
Langkah ini seharusnya mengunduh dataset “Bike Sharing Dataset” dari Kaggle. Namun, karena konfigurasi API gagal, dataset tidak berhasil diunduh otomatis.

4. Ekstraksi Dataset



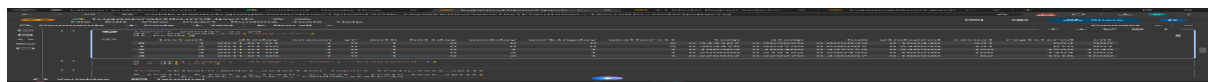
Perintah ini digunakan untuk mengekstrak file ZIP hasil unduhan, tapi karena file belum ada, proses ini juga gagal.

5. Membaca Dataset



Karena file tidak berhasil diunduh otomatis, kamu memuat dataset day.csv yang sudah ada secara manual di penyimpanan Colab. Dataset ini berisi **data peminjaman sepeda per hari**, dengan kolom seperti temp (temperatur), hum (kelembapan), windspeed (kecepatan angin), dan cnt (jumlah sepeda dipinjam).

6. Menentukan Fitur dan Target



Empat kolom (temp, atemp, hum, windspeed) dipilih sebagai **fitur input (variabel bebas)**. Kolom cnt digunakan sebagai **target output (variabel terikat)**, yaitu jumlah total sepeda yang disewa per hari.

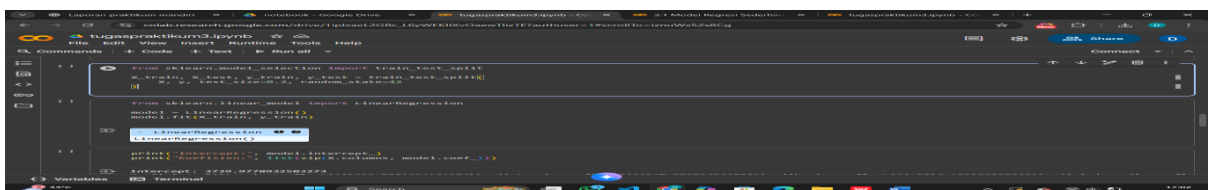
7. Membagi Data Latih dan Data Uji



Dataset dibagi menjadi:

- **Data latih (80%)** untuk melatih model.
 - **Data uji (20%)** untuk menguji kinerja model.
- random_state=42 menjaga agar pembagian data selalu konsisten setiap dijalankan.

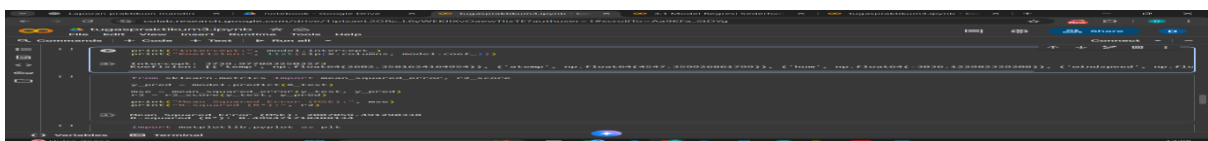
8. Membuat dan Melatih Model



Membangun model **Multiple Linear Regression** dari library scikit-learn.

fit() melatih model dengan data latih (X_train, y_train), agar model mempelajari hubungan linear antara suhu, kelembapan, angin, dan jumlah penyewaan sepeda.

9. Melihat Hasil Pelatihan (Koefisien dan Intersep)



Artinya:

- Jika suhu (temp) dan atemp naik, jumlah penyewaan sepeda meningkat.
- Jika kelembapan (hum) atau kecepatan angin (windspeed) naik, penyewaan sepeda justru menurun.
- Nilai intersep (3739.97) adalah nilai prediksi dasar saat semua variabel input bernilai nol.

10. Evaluasi Model

```

from sklearn.metrics import mean_squared_error, r2_score

y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
print("Mean Squared Error (MSE):", mse)
print("R-squared (R^2):", r2)

Mean Squared Error (MSE): 2007059.491204338
R-squared (R^2): 0.4992146888154

import matplotlib.pyplot as plt
plt.figure(figsize=(7,5))
plt.scatter(y_test, y_pred, color='blue')
plt.xlabel('Nilai Aktual (cnt)')
plt.ylabel('Nilai Prediksi (cnt)')
plt.title('Perbandingan Nilai Aktual vs Prediksi')
plt.grid(True)
plt.show()

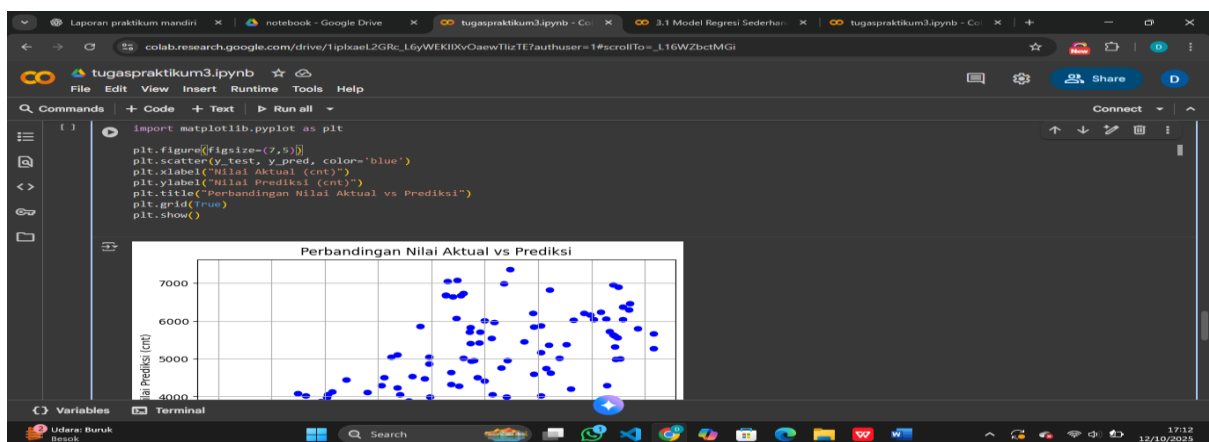
```

Model diuji pada data uji dan hasilnya:

- **MSE = 2,007,059.49** → rata-rata kesalahan kuadrat prediksi cukup besar.
- **R² = 0.499** → sekitar 49.9% variasi jumlah penyewaan sepeda dapat dijelaskan oleh keempat variabel input.

Artinya model cukup baik, tapi masih bisa ditingkatkan (misalnya dengan menambahkan fitur lain seperti cuaca atau musim).

11. Visualisasi Hasil Prediksi



Menampilkan grafik perbandingan antara nilai aktual dan prediksi:

- Titik biru yang mendekati garis diagonal menunjukkan prediksi yang akurat.
- Penyebaran data yang jauh dari diagonal menunjukkan error yang cukup besar di sebagian data.

12. Ringkasan Akhir

```

print("MSE: ", mse)
print("R-squared: ", r2)
print("Coefficients: ", model.coef_)

```

Kode ini menampilkan ulang semua hasil penting model agar mudah dibaca di satu tempat:

- Intersep

- Koefisien tiap fitur
- Nilai MSE dan R^2

Kesimpulan:

Model *Multiple Linear Regression* mampu menjelaskan sekitar **50% variasi data** peminjaman sepeda berdasarkan suhu, kelembapan, dan kecepatan angin.

Nilai koefisien menunjukkan bahwa suhu memiliki pengaruh positif terbesar terhadap jumlah penyewaan sepeda, sedangkan kelembapan dan angin cenderung menurunkan jumlah penyewaan.

Referensi

1. Bike Sharing Dataset – Kaggle.
<https://www.kaggle.com/datasets/lakshmi25npathi/bike-sharing-dataset>
2. Scikit-learn Documentation – Linear Models.
https://scikit-learn.org/stable/modules/linear_model.html
3. Jason Brownlee. (2020). *A Gentle Introduction to Multiple Linear Regression*. Machine Learning Mastery.
<https://machinelearningmastery.com/multiple-linear-regression-for-machine-learning/>
4. Google Colab Documentation.
<https://colab.research.google.com/>