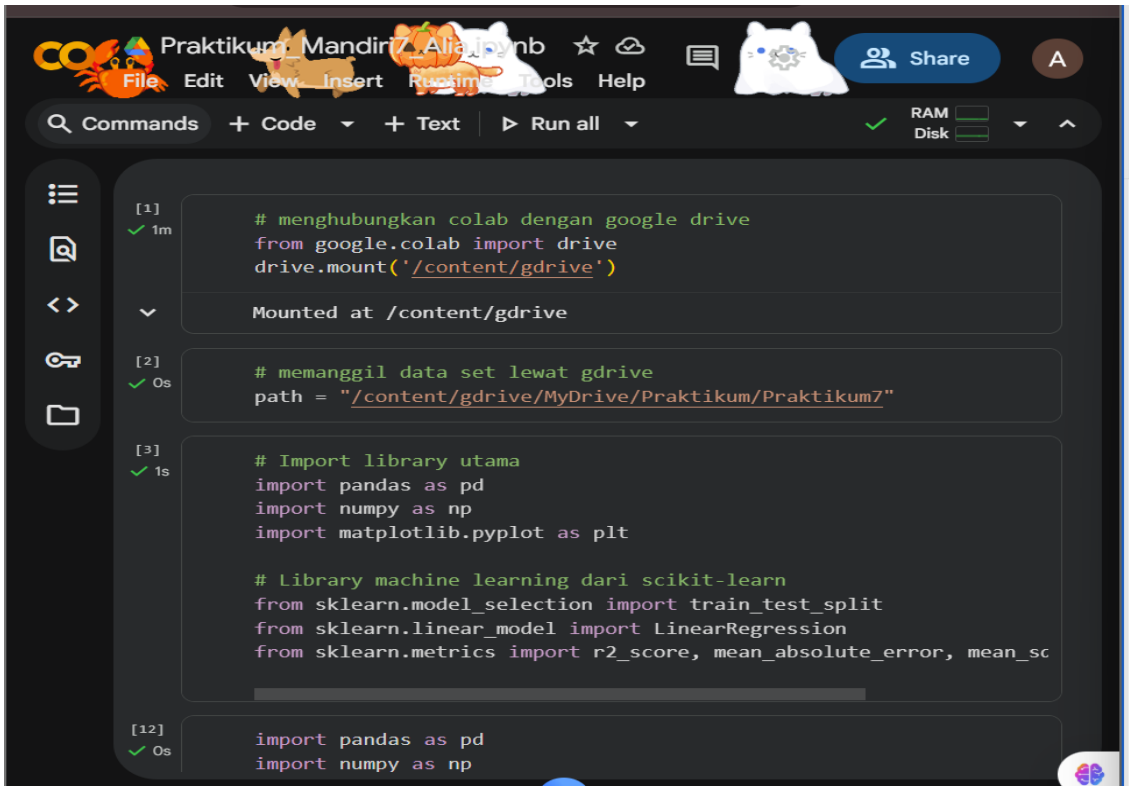
Tugas 7: Tugas Praktikum Mandiri 7 – Machine Learning

Alia Maisyarah 1 - 0110224095 [1*]

[1] Teknik Informatika, STT Terpadu Nurul Fikri, Depok

*E-mail:40110224095@student.nurulfikri.ac.id – **email mahasiswa 1**

1. Latihan Mandiri 7



1. # menghubungkan colab dengan google drive
   from google.colab import drive
   drive.mount('/content/gdrive')

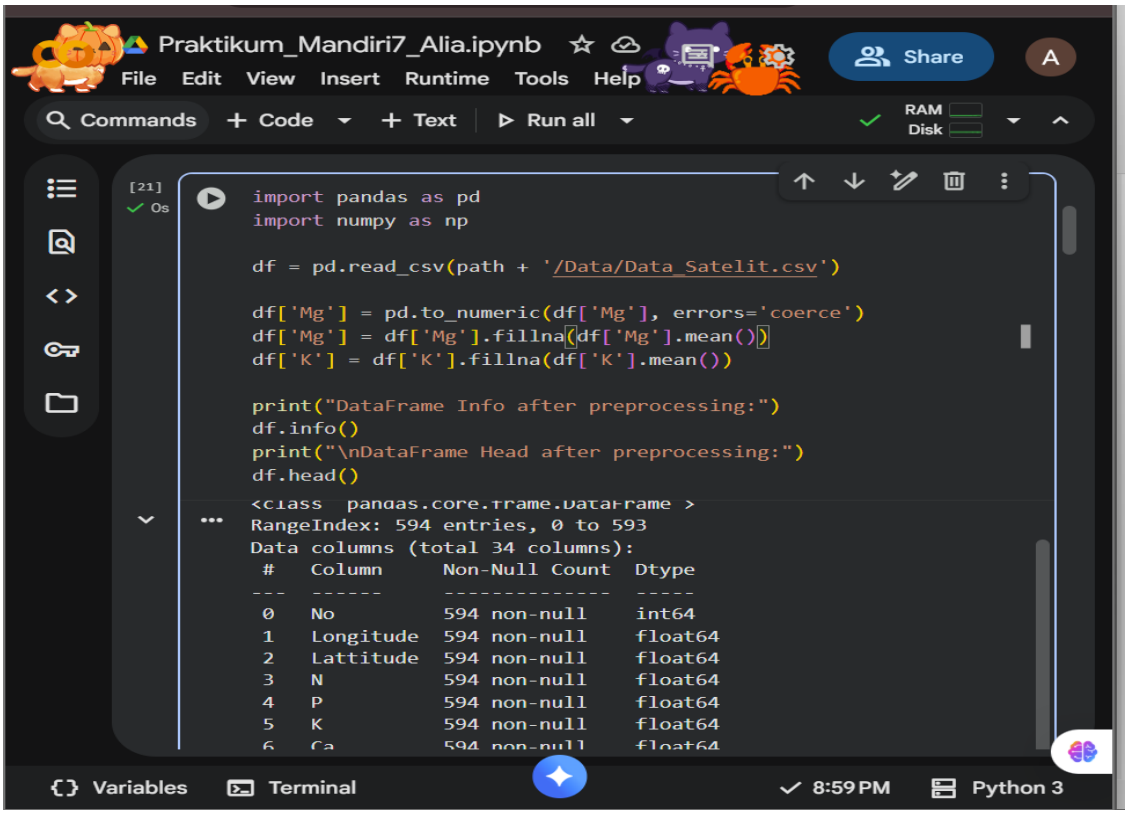
   (Agar Colab bisa mengakses file yang tersimpan di Google Drive kamu (misalnya dataset, laporan, atau file hasil analisis).Setelah baris ini dijalankan, folder Drive kamu akan muncul di /content/gdrive.)

2. memanggil data set lewat gdrive
   path = "/content/gdrive/MyDrive/Praktikum/Praktikum7"

   (Menetapkan lokasi folder tempat file dataset kamu disimpan, supaya Colab tahu di mana harus mencari file .csv tersebut.Biasanya nanti path ini digabungkan dengan nama file untuk dibaca pakai pd.read_csv(path + '/Data_Satelit.csv').

3. Import library utama
   import pandas as pd
   import numpy as np
   import matplotlib.pyplot as plt
   from sklearn.model_selection import train_test_split
   from sklearn.linear_model import LinearRegression
   from sklearn.metrics import r2_score, mean_absolute_error, mean_squared_error

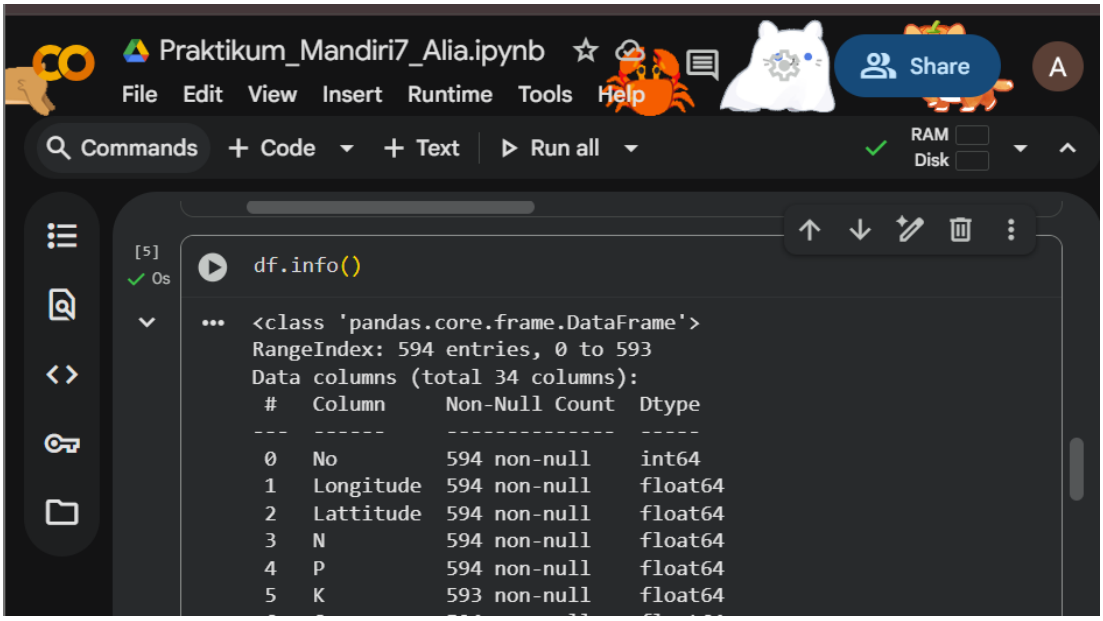
   (Bagian ini berfungsi untuk memanggil semua pustaka (library) yang diperlukan dalam proses pengolahan data, pembuatan model regresi, dan evaluasi hasilnya.)



4. import pandas as pd
   import numpy as np

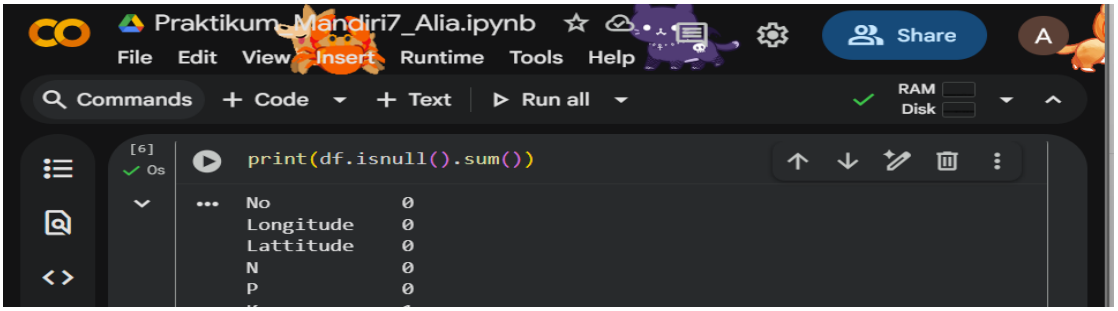
   df = pd.read_csv(path + '/Data/Data_Satelit.csv')
   df['Mg'] = pd.to_numeric(df['Mg'], errors='coerce')
   df['Mg'] = df['Mg'].fillna(df['Mg'].mean())
   df['K'] = df['K'].fillna(df['K'].mean()
   print("DataFrame Info after preprocessing:")
   df.info()
   print("\nDataFrame Head after preprocessing:")
   df.head()

(Kode ini digunakan untuk membaca file dataset dari Google Drive, lalu melakukan pra-pemrosesan data (data preprocessing))



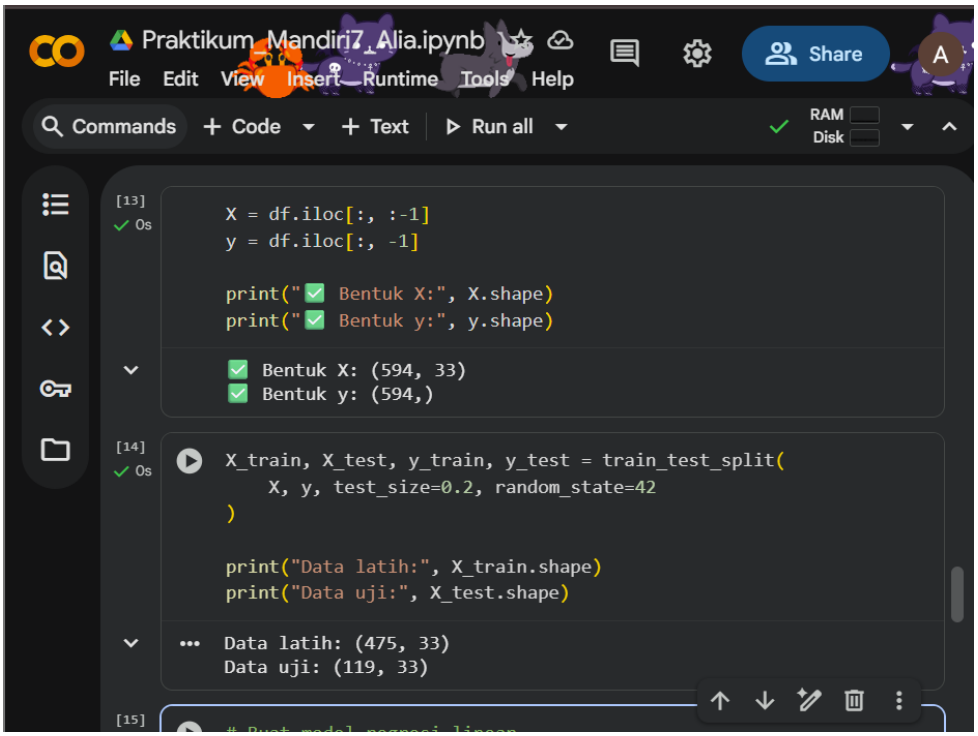
5. df.info()

(Digunakan untuk menampilkan informasi struktur DataFrame)



6. print(df.isnull().sum())

(Digunakan untuk mengecek jumlah nilai kosong (missing values) di setiap kolom pada dataset.)



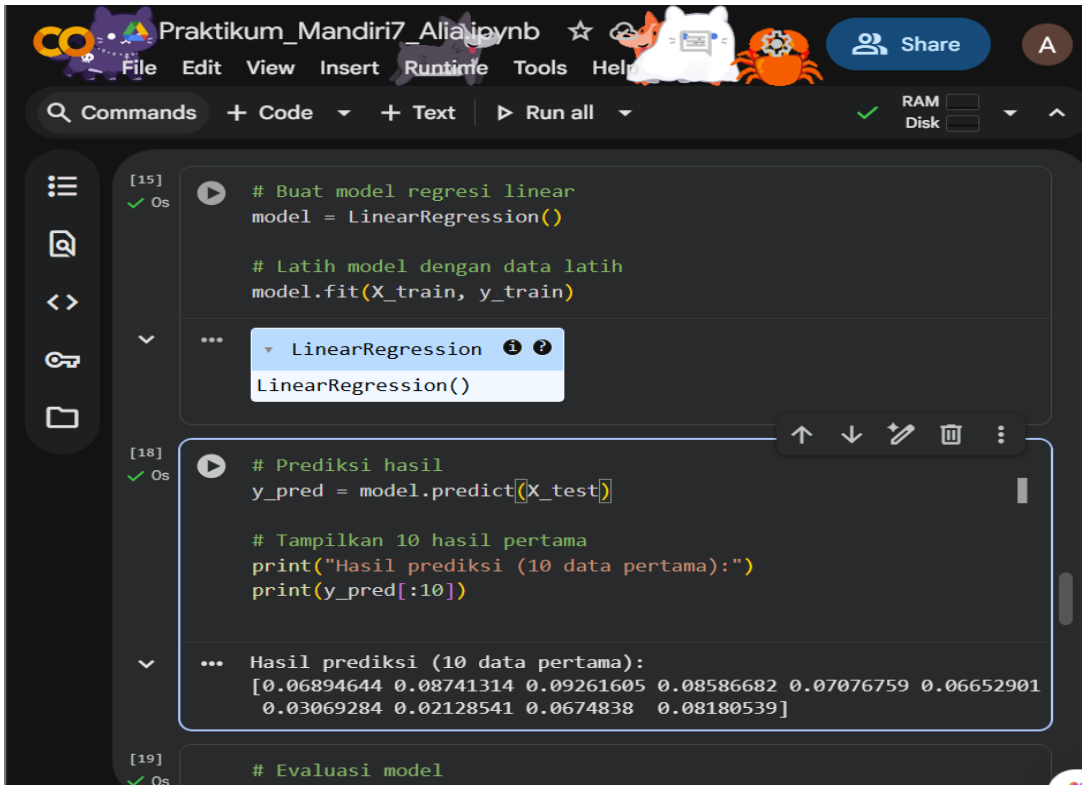
7. X = df.iloc[:, :-1]
   y = df.iloc[:, -1]

   print("✅ Bentuk X:", X.shape)
   print("✅ Bentuk y:", y.shape)

   (Memisahkan data menjadi fitur (X) dan target (y), agar model regresi tahu mana yang digunakan sebagai input dan mana yang akan diprediksi.)

8. X_train, X_test, y_train, y_test = train_test_split(
   X, y, test_size=0.2, random_state=42
)
   print("Data latih:", X_train.shape)
   print("Data uji:", X_test.shape)

   (Membagi dataset menjadi data latih (80%) dan data uji (20%) supaya model bisa dilatih dulu, lalu diuji dengan data yang belum pernah dilihat sebelumnya untuk mengevaluasi akurasi.)



9. Buat model regresi linear
   model = LinearRegression()

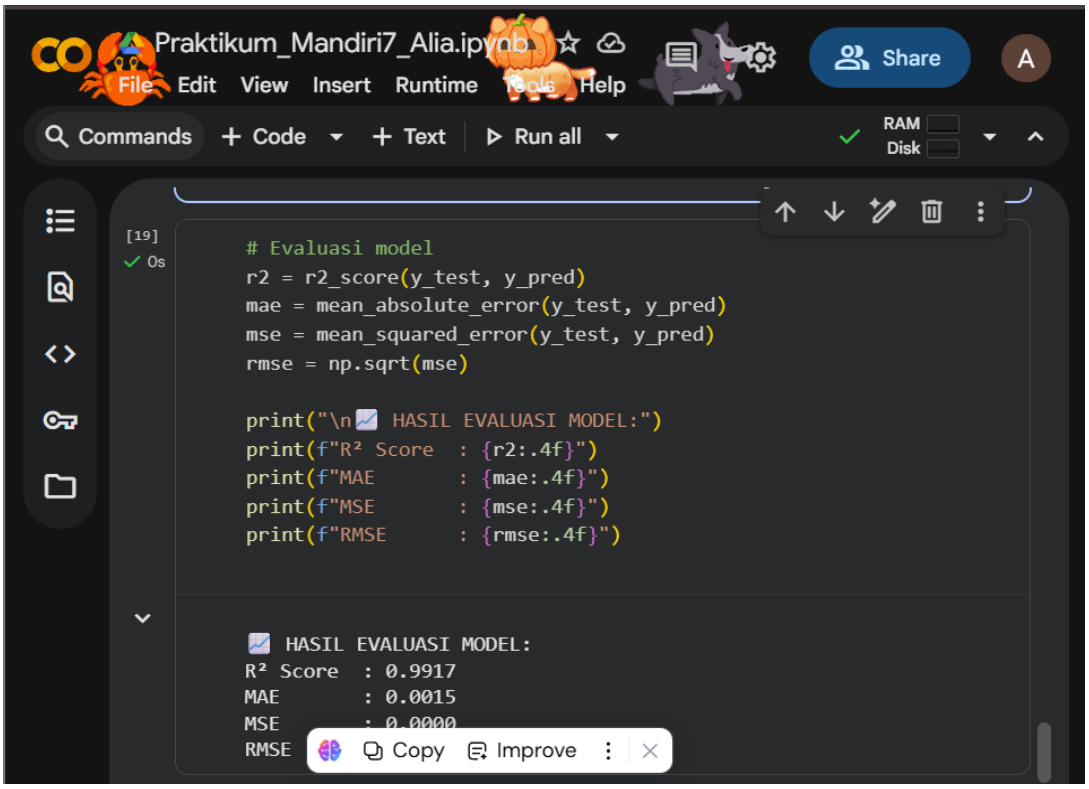
   # Latih model dengan data latih
   model.fit(X_train, y_train)

   (Membuat dan melatih model Regresi Linear menggunakan data latih (X_train, y_train). Tujuannya supaya model dapat mempelajari hubungan antara fitur dan target, sehingga nanti bisa memprediksi nilai target pada data baru.)

10. Prediksi hasil
    y_pred = model.predict(X_test)

    # Tampilkan 10 hasil pertama
    print("Hasil prediksi (10 data pertama):")
    print(y_pred[:10])

    (Kode ini digunakan untuk menguji model regresi linear dengan data uji (X_test) dan menampilkan 10 hasil prediksi pertama.)

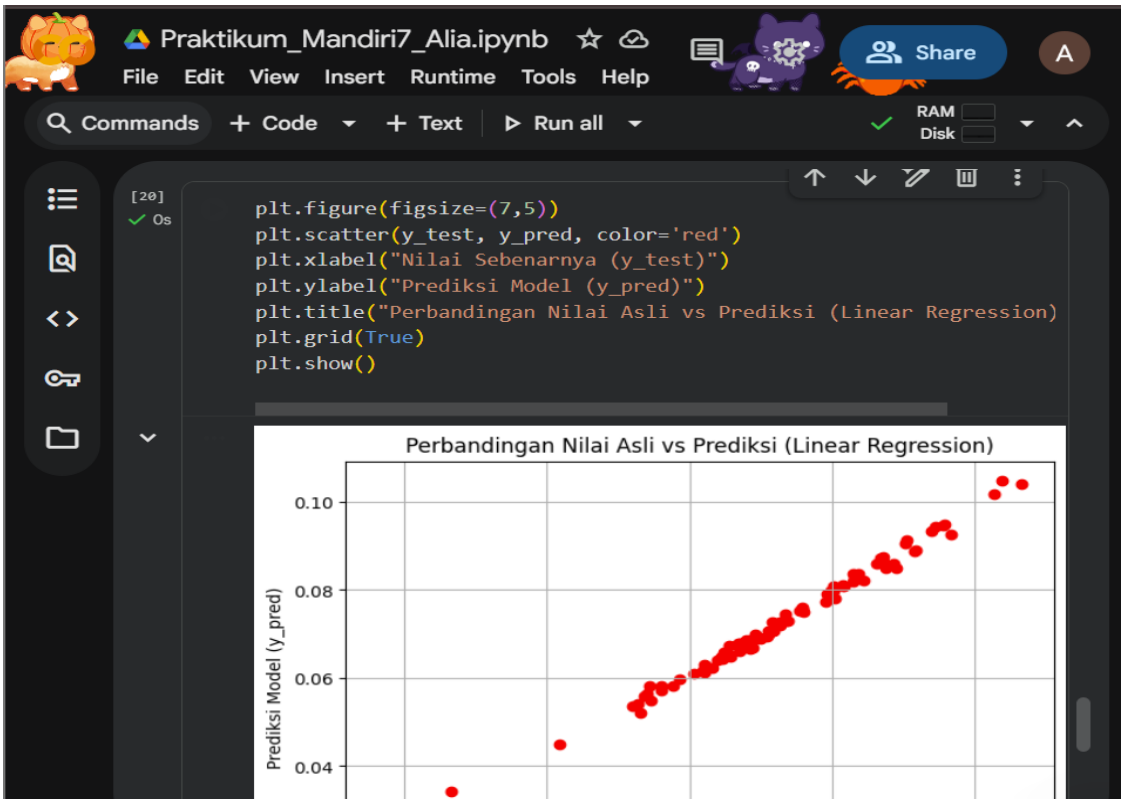


11. # Evaluasi model

```
r2 = r2_score(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)

print("\n📈 HASIL EVALUASI MODEL:")
print(f"R² Score  : {r2:.4f}")
print(f"MAE       : {mae:.4f}")
print(f"MSE       : {mse:.4f}")
print(f"RMSE      : {rmse:.4f}")
```

(Kode ini digunakan untuk mengukur seberapa baik kinerja model regresi linear dengan menghitung nilai R^2, MAE, MSE, dan RMSE, sehingga dapat diketahui tingkat akurasi dan kesalahan prediksi model terhadap data uji.)



12. plt.figure(figsize=(7,5))
    plt.scatter(y_test, y_pred, color='red')
    plt.xlabel("Nilai Sebenarnya (y_test)")
    plt.ylabel("Prediksi Model (y_pred)")
    plt.title("Perbandingan Nilai Asli vs Prediksi (Linear Regression)")
    plt.grid(True)
    plt.show()

    (Kode ini digunakan untuk menampilkan visualisasi hasil prediksi model regresi linear dalam bentuk grafik scatter plot, agar dapat dilihat perbandingan antara nilai sebenarnya dan nilai prediksi, serta mengetahui seberapa baik model mengikuti pola data aslinya.)

LINK GITHUB UPLOAD TUGAS :

https://github.com/AliaMaisyarah14/PRAKTIKUM_ML/tree/bb7e9033f96628aff17011ad6a9a4247721306b2/Praktikum7_AliaMaisyarah/Praktikum7