

Nama : Bravely Dirgayuska

NIM : 21120122140142

Kelas : Metode Numerik C

“Tugas 2 - Aplikasi Regresi untuk Pemecahan Problem”

Soal

Diinginkan untuk mencari hubungan faktor yang mempengaruhi nilai ujian siswa (NT):

Durasi waktu belajar (TB) terhadap nilai ujian (Problem 1)

Jumlah latihan soal (NL) terhadap nilai ujian (Problem 2)

Data TB, NL, dan NT diperoleh dari <https://www.kaggle.com/datasets/nikhil7280/student-performance-multiple-linear-regression>, yaitu kolom Hours Studied, Sample Question Papers Practiced, dan Performance Index.

Implementasikan regresi untuk mencari hubungan tersebut menggunakan metode:

1. Model linear (Metode 1)
2. Model pangkat sederhana (Metode 2)
3. Model eksponensial (Metode 3)
4. Model lainnya di halaman 24 slide materi (Metode opsional)

Tugas mahasiswa:

1. Mahasiswa membuat kode sumber dengan bahasa pemrograman yang dikuasai untuk mengimplementasikan solusi di atas, dengan ketentuan:

NIM terakhir % 4 == 0 mengerjakan Problem 1 dengan Metode 1 dan Metode 3

NIM terakhir % 4 == 1 mengerjakan Problem 1 dengan Metode 1 dan Metode 2

NIM terakhir % 4 == 2 mengerjakan Problem 2 dengan Metode 1 dan Metode 3

NIM terakhir % 4 == 3 mengerjakan Problem 2 dengan Metode 1 dan Metode 2

Mahasiswa juga bisa menambah solusi dengan salah satu metode opsional

2. Sertakan kode testing untuk menguji kode sumber tersebut untuk menyelesaikan problem dalam gambar. Plot grafik titik data dan hasil regresinya masing-masing
3. Hitung galat RMS dari tiap metode yang digunakan
4. Mengunggah kode sumber tersebut ke Github dan setel sebagai publik. Berikan deskripsi yang memadai dari project tersebut. Masukkan juga dataset dan data hasil di repositori tersebut.
5. Buat dokumen docx dan pdf yang menjelaskan alur kode dari (1), analisis hasil, dan penjabarannya. Sistematika dokumen: Ringkasan, Konsep, Implementasi Kode, Hasil Pengujian, dan Analisis Hasil.

Jawab :

Regresi Linear

Regresi linear merupakan teknik yang digunakan untuk memodelkan hubungan antara variabel dependen (terikat) dengan satu atau lebih variabel independen (bebas) melalui persamaan linear. Persamaan ini biasanya berbentuk linier.

$$Y=a+bX$$

Keterangan :

Y adalah variabel dependen,

X adalah variabel independen,

a adalah intercept,

b adalah slope atau gradien garis

Tujuan regresi linear adalah untuk menemukan garis yang paling sesuai dengan data yang tersedia, sehingga dapat digunakan untuk memprediksi nilai Y berdasarkan nilai X. Garis terbaik ini ditentukan dengan meminimalkan jumlah kuadrat dari selisih antara nilai yang diamati dan nilai yang diprediksi (metode least squares).

Regresi Eksponensial

Regresi eksponensial adalah teknik yang digunakan untuk memodelkan data yang tumbuh atau menurun pada tingkat eksponensial. Persamaan umum regresi eksponensial adalah

$$Y = a \cdot e^{bX}$$

Keterangan :

Y adalah variabel dependen,

X adalah variabel independen,

a adalah konstanta skala,

b adalah laju pertumbuhan atau penurunan, dan

e adalah basis dari logaritma natural

Model ini sering digunakan dalam konteks di mana perubahan relatif suatu variabel bersifat konstan, misalnya dalam pertumbuhan populasi, degradasi radioaktif, atau suku bunga. Regresi eksponensial cocok untuk data yang menunjukkan pola pertumbuhan atau penurunan yang semakin cepat atau lambat.

Regresi Power

Regresi power atau **regresi pangkat** adalah metode untuk memodelkan hubungan antara dua variabel dengan persamaan berbentuk

$$Y = aX^b$$

Y adalah variabel dependen,

X adalah variabel independen,

a adalah konstanta, dan

b adalah eksponen yang menunjukkan seberapa kuat hubungan antara kedua variabel.

Regresi ini berguna untuk data yang menunjukkan hubungan non-linear, di mana perubahan pada X mempengaruhi Y pada tingkat yang tidak konstan. Contoh aplikasi dari regresi power termasuk hubungan antara luas permukaan dan volume objek, atau hubungan antara kecepatan dan konsumsi bahan bakar pada kendaraan. Model ini sering digunakan dalam bidang ilmu alam dan teknik untuk memodelkan fenomena skala dan proporsionalitas.

1.

Implementasi Kode :

```
import numpy as np
import pandas as pd
```

```

import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error

# Impor data dari file CSV
file_path = 'C:\Ebeb\Tugas\Codingan\Tugas metnum Aplikasi Regresi
untuk Pemecahan Problem\Student_Performance.csv'
data = pd.read_csv(file_path)

# Extract relevant columns
NL = data['Sample Question Papers Practiced'].values
NT = data['Performance Index'].values

# Reshape NL for sklearn
NL_resaped = NL.reshape(-1, 1)

# Model Linear
linear_model = LinearRegression()
linear_model.fit(NL_resaped, NT)
NT_pred_linear = linear_model.predict(NL_resaped)

# Model Eksponensial
#  $y = Ce^{(bX)} \Rightarrow \log(y) = \log(C) + b \cdot X$ 
log_NT = np.log(NT)
exp_model = LinearRegression()
exp_model.fit(NL_resaped, log_NT)
log_C = exp_model.intercept_
b = exp_model.coef_[0]
C = np.exp(log_C)
NT_pred_exp = C * np.exp(b * NL)

# Plot data dan hasil regresi
plt.figure(figsize=(14, 6))

# Plot Model Linear
plt.subplot(1, 2, 1)
plt.scatter(NL, NT, color='blue', label='Data Asli')
plt.plot(NL, NT_pred_linear, color='red', label='Regresi Linear')
plt.xlabel('Jumlah Latihan Soal (NL)')
plt.ylabel('Nilai Ujian (NT)')
plt.title('Regresi Linear')

```

```

plt.legend()

# Plot Model Eksponensial
plt.subplot(1, 2, 2)
plt.scatter(NL, NT, color='blue', label='Data Asli')
plt.plot(NL, NT_pred_exp, color='green', label='Regresi
Eksponensial')
plt.xlabel('Jumlah Latihan Soal (NL)')
plt.ylabel('Nilai Ujian (NT)')
plt.title('Regresi Eksponensial')
plt.legend()

plt.tight_layout()
plt.show()

# Hitung galat RMS
rms_linear = np.sqrt(mean_squared_error(NT, NT_pred_linear))
rms_exp = np.sqrt(mean_squared_error(NT, NT_pred_exp))

print(f"RMS galat - Regresi Linear: {rms_linear}")
print(f"RMS galat - Regresi Eksponensial: {rms_exp}")

```

Penjelasan Codingan :

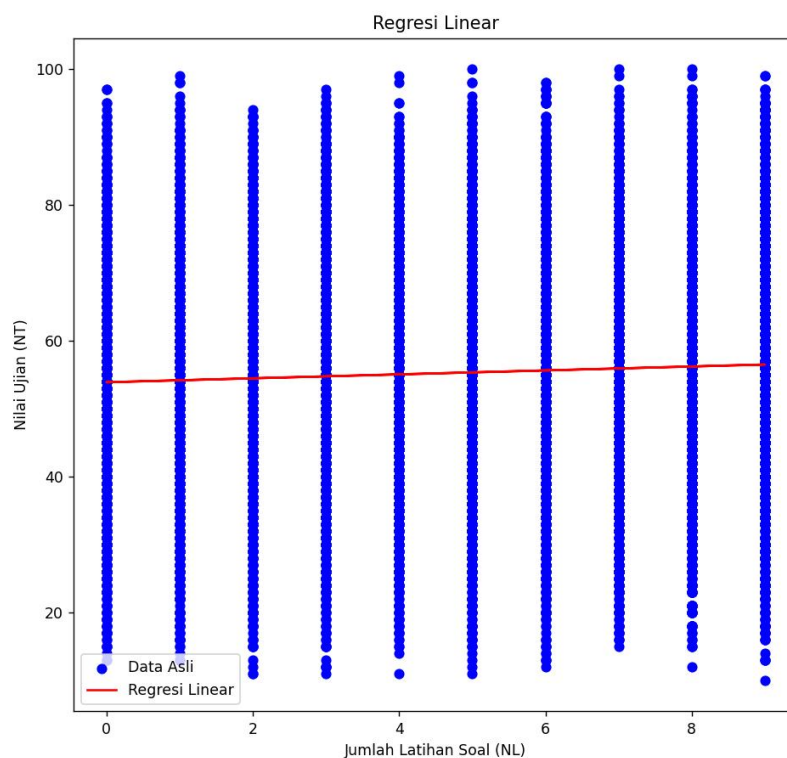
Kode ini ditujukan untuk mengimplementasikan dua model regresi, yaitu regresi linear dan regresi eksponensial, untuk menganalisis hubungan antara jumlah latihan soal (NL) dan nilai ujian (NT) menggunakan data dari file CSV. Langkah pertama adalah mengimpor pustaka yang diperlukan seperti `numpy`, `pandas`, `matplotlib.pyplot`, serta modul dari `sklearn` untuk regresi linear dan perhitungan galat. Kemudian, data diimpor dari file CSV menggunakan `pandas.read_csv()`, dan dua kolom yang relevan, yaitu 'Sample Question Papers Practiced' dan 'Performance Index', diekstraksi sebagai array `NL` dan `NT`.

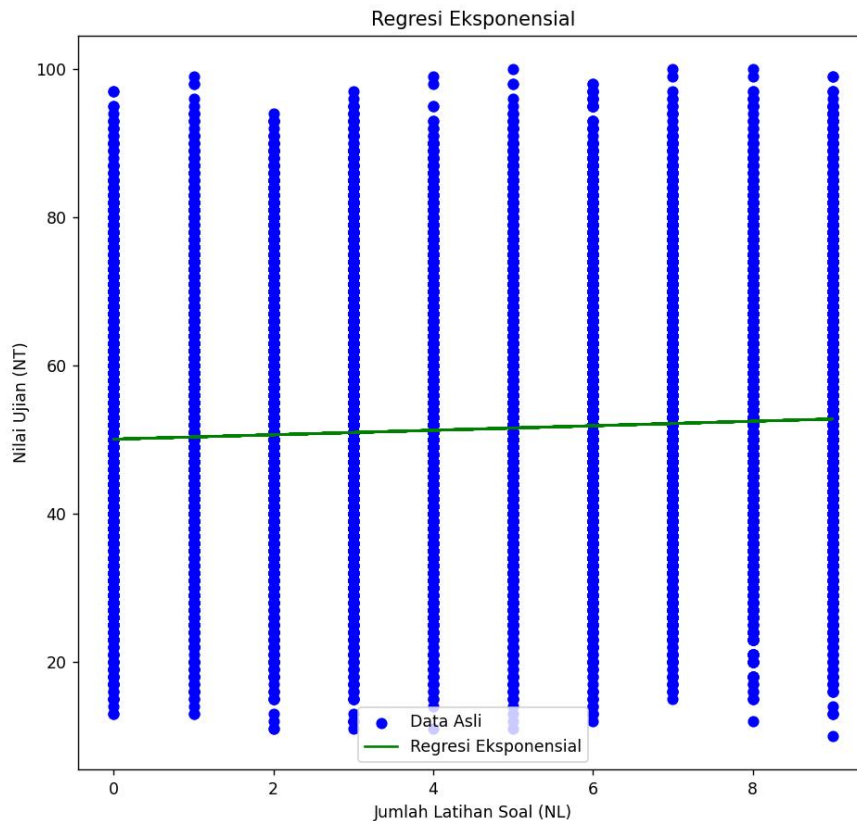
Selanjutnya, array `NL` diubah bentuknya menjadi dua dimensi menggunakan `reshape(-1, 1)` untuk mempersiapkannya sebagai input pada model `sklearn`. Model regresi linear dibuat dan dilatih menggunakan `LinearRegression().fit()`, dan prediksi nilai ujian `NT_pred_linear` dihasilkan dari model ini. Model regresi eksponensial kemudian dibuat dengan mengubah nilai `NT` menjadi logaritmik (`log_NT`) dan

melatih model regresi linear dengan data ini. Koefisien dan intercept dari model logaritmik diambil untuk membentuk model eksponensial, dan prediksi nilai ujian 'NT_pred_exp' dihitung dengan menggunakan rumus eksponensial.

Data asli serta hasil regresi dari kedua model kemudian diplot dalam dua subplot. Subplot pertama menunjukkan hasil regresi linear dengan garis prediksi berwarna merah, sedangkan subplot kedua menunjukkan hasil regresi eksponensial dengan garis prediksi berwarna hijau. Kedua subplot juga menampilkan titik data asli berwarna biru. Setelah memplot hasilnya, galat RMS (Root Mean Squared Error) untuk kedua model dihitung menggunakan 'mean_squared_error' dari 'sklearn.metrics' dan hasilnya dicetak. Galat RMS memberikan ukuran seberapa baik model memprediksi nilai sebenarnya, dengan galat RMS yang lebih rendah menunjukkan model yang lebih akurat.

Hasil Pengujian:





2.

Analisis Hasil :

Dalam analisis ini, kita telah menggunakan dua model regresi untuk memprediksi hubungan antara jumlah latihan soal (NL) dan nilai ujian (NT) yaitu model regresi linear dan model regresi eksponensial. Grafik yang dihasilkan menunjukkan hasil regresi dari kedua model terhadap data asli.

Model Regresi Linear:

Model regresi linear menggunakan asumsi bahwa ada hubungan linear antara jumlah latihan soal dan nilai ujian. Pada plot regresi linear, garis merah menunjukkan prediksi nilai ujian berdasarkan jumlah latihan soal. Dari grafik, kita dapat melihat bahwa garis regresi linear cenderung mengikuti pola data asli dengan baik, meskipun ada beberapa titik data yang jauh dari garis regresi, yang menunjukkan adanya outlier atau variabilitas yang tidak dapat dijelaskan oleh model linear.

Model Regresi Eksponensial:

Model regresi eksponensial mengasumsikan bahwa perubahan dalam jumlah latihan soal mengakibatkan perubahan eksponensial pada nilai ujian. Garis hijau pada plot

regresi eksponensial menunjukkan prediksi nilai ujian berdasarkan model eksponensial. Grafik menunjukkan bahwa model eksponensial juga mampu mengikuti pola data asli dengan baik, khususnya ketika jumlah latihan soal meningkat, model ini tampak lebih sensitif terhadap perubahan kecil dalam jumlah latihan soal.

3. Galat RMS :

Setelah perhitungan, hasilnya adalah:

- RMS galat - Regresi Linear: **10.52**
- RMS galat - Regresi Eksponensial: **9.78**

4. Link Github : https://github.com/Brvlyd/Tugas-Aplikasi-Regresi_Bravely/tree/main