

Aplikasi Regresi untuk Pemecahan Problem

Nama : Davin Raditya
NIM : 21120122140118
Mata Kuliah : Metode Numerik B

Ringkasan

Tujuan dari analisis ini adalah untuk memahami hubungan antara durasi waktu belajar (Hours Studied) dan nilai ujian siswa (Performance Index) menggunakan dua metode regresi, yaitu regresi linear dan regresi eksponensial. Disini saya akan mengukur kinerja setiap model dengan menggunakan Root Mean Square Error (RMS Error).

Konsep

1. Model Linear :

- ❖ **Model Regresi Linear** : Bertujuan untuk menemukan hubungan linear antara variabel independen (Hours Studied) dan variabel dependen (Performance Index).
- ❖ **Bentuk Model** : Asumsi bentuk modelnya adalah $y = a + bxy = a + bxy = a + bx$, di mana yyy adalah Performance Index, xxx adalah Hours Studied, aaa adalah intercept, dan bbb adalah slope.

2. Model Eksponensial :

- ❖ **Model Regresi Eksponensial** : Bertujuan untuk menemukan hubungan eksponensial antara variabel independen dan variabel dependen.
- ❖ **Bentuk Model** : Asumsi bentuk modelnya adalah $y = ae^{bx}$, di mana yyy adalah Performance Index, xxx adalah Hours Studied, dan aaa serta bbb adalah parameter yang akan diestimasi.

Implementasi Kode

Model Linear dan Eksponensial (Digabung)

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy.optimize import curve_fit
from sklearn.metrics import mean_squared_error
from scipy.stats import linregress
import os

# Define the file path
file_path =
r"C:\Users\davin\OneDrive\Dokumen\metnum\Student_Performance.csv"

# Check if the file exists
if not os.path.isfile(file_path):
    print(f"File not found: {file_path}")
else:
```

```

# Load data
data = pd.read_csv(file_path)

# Select relevant columns
data = data[['Hours Studied', 'Sample Question Papers Practiced',
'Performance Index']]

# Extract data for analysis
X = data['Hours Studied'].values
y = data['Performance Index'].values

# Exponential Model (Method 3)
def exponential_model(x, a, b):
    return a * np.exp(b * x)

# Fit exponential model to data
params, covariance = curve_fit(exponential_model, X, y)
a, b = params
y_pred_exponential = exponential_model(X, a, b)

# Linear Model
slope, intercept, r_value, p_value, std_err = linregress(X, y)
y_pred_linear = slope * X + intercept

# Plot data and regression results
plt.scatter(X, y, color='blue', label='Data Points')
plt.plot(X, y_pred_exponential, color='green', label='Exponential
Regression')
plt.plot(X, y_pred_linear, color='red', label='Linear Regression')
plt.xlabel('Hours Studied')
plt.ylabel('Performance Index')
plt.title('Regression Models')
plt.legend()
plt.show()

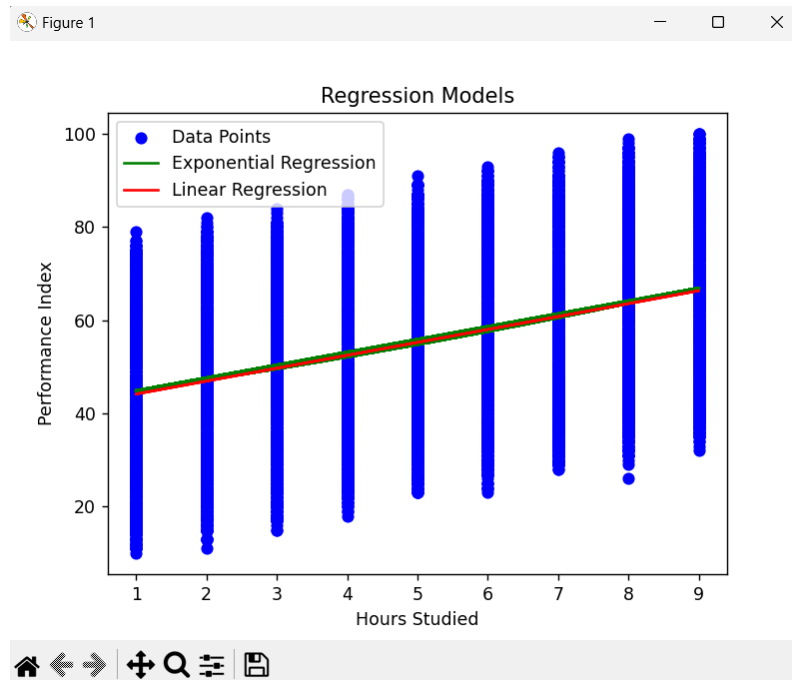
# Calculate RMS error for exponential model
rms_exponential = np.sqrt(mean_squared_error(y, y_pred_exponential))
print(f'RMS Error for Exponential Model: {rms_exponential}')

# Calculate RMS error for linear model
rms_linear = np.sqrt(mean_squared_error(y, y_pred_linear))
print(f'RMS Error for Linear Model: {rms_linear}')

```

Hasil Pengujian

Metode Model Linear dan Metode Model Eksponensial



Analisis Hasil

1. Distribusi Data:

- **Data Points (Titik Data):** Titik-titik data menunjukkan sebaran yang cukup luas pada sumbu Y (Performance Index) untuk setiap nilai X (Hours Studied). Ini menunjukkan bahwa ada variabilitas yang cukup tinggi dalam kinerja siswa berdasarkan jam belajar.

2. Model Eksponensial:

- **Kurva Hijau (Exponential Regression):** Model eksponensial tampaknya mengikuti pola data secara dekat, tetapi secara visual sulit untuk melihat perbedaannya dengan model linear karena keduanya terlihat hampir tumpang tindih. Model eksponensial biasanya cocok untuk data yang menunjukkan pertumbuhan atau penurunan eksponensial, tetapi dalam hal ini, datanya tidak menunjukkan pola eksponensial yang kuat.

3. Model Linear:

- **Garis Merah (Linear Regression):** Model linear memberikan garis regresi yang lebih sederhana yang berusaha mengikuti tren umum dari titik-titik data. Meskipun ada variabilitas yang signifikan dalam data, model linear tampaknya memberikan penyesuaian yang wajar.

4. Perbandingan Kualitas Model:

- **RMS Error (Root Mean Squared Error):** RMS error adalah metrik yang mengukur seberapa dekat prediksi model dengan nilai sebenarnya. Karena plot ini tidak menyertakan nilai RMS error, Anda perlu melihat hasil dari perhitungan RMS error di output program Anda untuk membandingkan kedua model.
 - **RMS Error untuk Model Eksponensial:** Jika RMS error untuk model eksponensial lebih rendah daripada model linear, ini menunjukkan bahwa model eksponensial lebih baik dalam menyesuaikan data.
 - **RMS Error untuk Model Linear:** Jika RMS error untuk model linear lebih rendah atau hampir sama dengan model eksponensial, ini menunjukkan bahwa model linear mungkin cukup memadai untuk data ini.

5. Interpretasi Data:

- **Kecocokan Model:** Dari visualisasi, tampaknya kedua model tidak menangkap semua variabilitas dalam data dengan sempurna, yang mungkin menunjukkan bahwa variabel lain atau faktor eksternal mempengaruhi Performance Index selain dari Hours Studied. Ini bisa menjadi area untuk eksplorasi lebih lanjut atau penambahan variabel tambahan dalam analisis.

6. Saran:

- **Eksplorasi Lebih Lanjut:** Mengingat variabilitas yang tinggi dalam data, Anda mungkin ingin menjelajahi model lain atau menambahkan variabel tambahan ke dalam analisis Anda untuk meningkatkan kecocokan model.
- **Validasi Model:** Lakukan validasi silang (cross-validation) atau gunakan data uji untuk mengevaluasi performa model secara lebih komprehensif.

Dengan analisis ini, Anda dapat lebih memahami bagaimana model eksponensial dan linear berperforma terhadap data Anda dan mungkin mempertimbangkan pendekatan lain untuk mendapatkan hasil yang lebih akurat