

PRAKTIKUM FISIKA KOMPUTASI

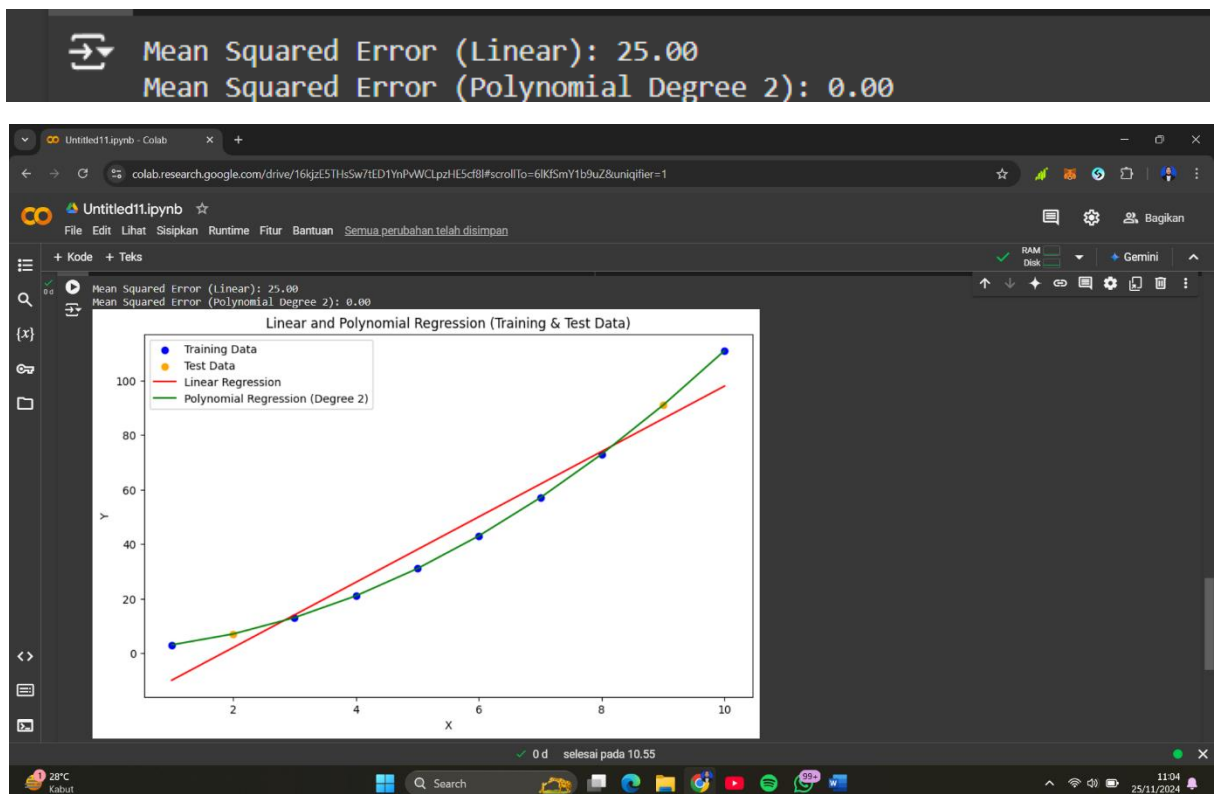
Machine Learning Regresi Linear dan Polinomial

Oleh:

Eneng Yulia Pebryanti

NIM 1227030013

1. Hasil kode program machine learning menggunakan dataset $X=[1,2,3,4,5,6,7,8,9,10]$ $Y=[3,7,13,21,31,43,57,73,91,111]$ dengan membagi dataset menjadi 80% data latih dan 20% data uji. Dan model regresi linear dan polynomial derajat 2 dan bandingkan mean squared error (MSE) beserta visualisasi hasil komputasinya



Membandingkan **Mean Squared Error (MSE)** dari kedua model, terlihat perbedaan yang sangat signifikan. Pada model **regresi linear**, MSE-nya sebesar **25.00**, yang menunjukkan bahwa tingkat error atau kesalahan prediksi cukup besar. Ini karena regresi linear hanya mampu menghasilkan garis lurus, sehingga tidak cocok untuk data yang memiliki pola melengkung atau non-linear seperti yang ada di grafik. Garis lurus ini tidak bisa mengikuti titik-titik data dengan baik, sehingga banyak prediksi yang meleset dari nilai sebenarnya. Sebaliknya, pada model **regresi polynomial derajat 2**, MSE-nya adalah **0.00**, yang berarti error hampir nol. Hal ini menunjukkan bahwa model polynomial sangat akurat, karena dalam menyesuaikan bentuk kurva yang lebih kompleks. Model polynomial derajat 2 ini bisa melengkung sesuai dengan pola data, sehingga prediksinya jauh lebih tepat dibandingkan regresi linear.

5. Penjelasan kode program dan hasil nya

Langkah pertama, library seperti **numpy**, **matplotlib**, dan **sklearn** diimpor. Fungsinya untuk mempermudah proses seperti mengolah data, membuat grafik, dan membuat model machine learning.

```
1 import numpy as np |
2 import matplotlib.pyplot as plt
3 from sklearn.linear_model import LinearRegression
4 from sklearn.preprocessing import PolynomialFeatures
5 from sklearn.metrics import mean_squared_error
6 from sklearn.model_selection import train_test_split
```

Setelah itu, dataset dibuat. Di sini, **X** diisi dengan angka dari 1 sampai 10 sebagai data masukan, dan **Y** diisi dengan angka yang mengikuti pola tertentu. Pola yang terlihat di **Y** itu seperti kuadratik atau data yang naik lebih cepat.

```
10 X = [[1], [2], [3], [4], [5], [6], [7], [8], [9], [10]]
11 Y = [3, 7, 13, 21, 31, 43, 57, 73, 91, 111]
```

Dataset tadi kemudian dibagi menjadi dua bagian: **data training** dan **data testing**. Data training dipakai untuk melatih model, sedangkan data testing digunakan untuk menguji keakuratan model setelah selesai dilatih. Pembagian ini dilakukan secara otomatis dengan 80% untuk training dan 20% untuk testing.

```
14 X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2,
    random_state=42)
```

Model regresi linear dibuat untuk memprediksi data dengan pendekatan garis lurus. **Model regresi polynomial derajat 2** dibuat untuk menghasilkan kurva yang lebih sesuai dengan pola data

```
17 linear_model = LinearRegression()
18 linear_model.fit(X_train, Y_train)
19
20 # Creating a polynomial regression model of degree 2
21 poly_features_2 = PolynomialFeatures(degree=2)
22 X_train_poly_2 = poly_features_2.fit_transform(X_train)
23 poly_model_2 = LinearRegression()
24 poly_model_2.fit(X_train_poly_2, Y_train)
```

Setelah model dibuat, prediksi dilakukan menggunakan data masukan **X**. Hasil prediksi dari kedua model (linear dan polynomial) disiapkan untuk dibandingkan. Data **X** diurutkan lebih dulu supaya garis grafik nantinya terlihat lebih rapi.

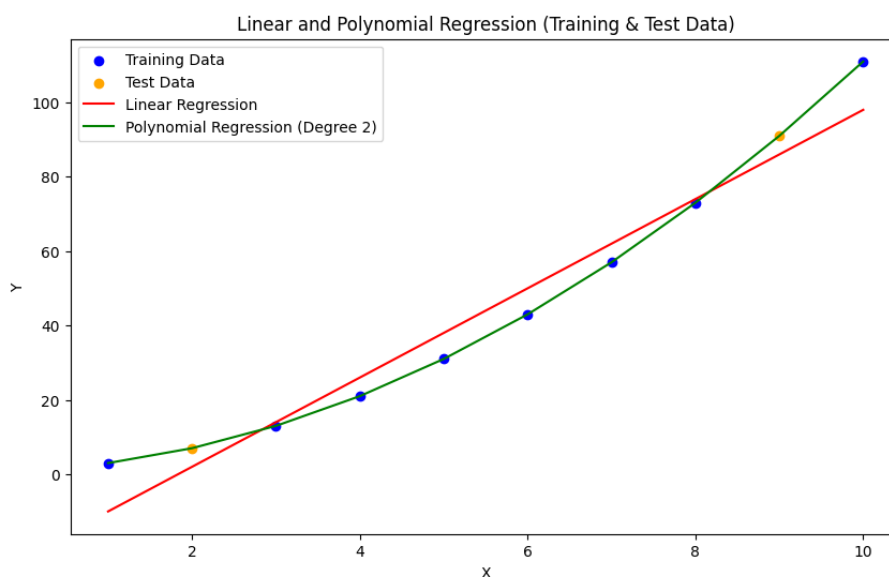
```
27 X_sorted = np.sort(X, axis=0) # Sorting X for smooth plotting
28 Y_pred_linear_all = linear_model.predict(X_sorted)
29 Y_pred_poly_2_all = poly_model_2.predict(poly_features_2.transform(X_sorted))
```

Setiap model dievaluasi dengan menghitung **Mean Squared Error (MSE)**. Angka MSE ini menunjukkan seberapa besar rata-rata kesalahan prediksi. Model regresi linear menghasilkan MSE yang lebih besar, artinya prediksinya kurang akurat. Sebaliknya, model polynomial memiliki MSE lebih kecil, sehingga dianggap lebih tepat.

```
32 mse_linear = mean_squared_error(Y_test, linear_model.predict(X_test))
33 mse_poly_2 = mean_squared_error(Y_test, poly_model_2.predict(poly_features_2.
    transform(X_test)))
34
35 print(f"Mean Squared Error (Linear): {mse_linear:.2f}")
36 print(f"Mean Squared Error (Polynomial Degree 2): {mse_poly_2:.2f}")
```

Grafik dibuat untuk menunjukkan hasil dari kedua model. Data asli ditampilkan sebagai titik-titik (biru untuk data training, oranye untuk data testing), garis merah menunjukkan prediksi dari regresi linear, dan garis hijau mewakili prediksi dari regresi polynomial. Hasil ini menunjukkan bahwa kurva polynomial lebih mengikuti pola data dibanding garis linear.

```
39 plt.figure(figsize=(10, 6))
40 plt.scatter(X_train, Y_train, color='blue', label='Training Data') # Training
    data
41 plt.scatter(X_test, Y_test, color='orange', label='Test Data') # Test data
42 plt.plot(X_sorted, Y_pred_linear_all, color='red', label='Linear Regression')
    # Linear regression line
43 plt.plot(X_sorted, Y_pred_poly_2_all, color='green', label='Polynomial
    Regression (Degree 2)') # Polynomial regression line
44
45 plt.xlabel('X')
46 plt.ylabel('Y')
47 plt.title('Linear and Polynomial Regression (Training & Test Data)')
48 plt.legend()
49 plt.show()
```



Penjelasan hasil grafik

Pada grafik yang ditampilkan Pertama, ada titik-titik berwarna biru yang menunjukkan data training, yaitu data yang digunakan untuk melatih model, dan titik berwarna oranye yang menunjukkan data testing, yang dipakai untuk menguji seberapa baik model bekerja. Garis merah (regresi linear) itu hanya membuat garis lurus. Walaupun bisa cocok dengan beberapa data, garis ini terlihat kurang tepat karena data yang sebenarnya membentuk pola melengkung, bukan garis lurus. Jadi, model ini kurang bisa menangkap pola yang ada. Garis hijau (regresi polynomial derajat 2) lebih fleksibel. Garis ini bisa mengikuti pola data yang melengkung, jadi lebih akurat dalam mencocokkan data yang ada.

Berdasarkan Mean Squared Error (MSE) yang muncul di atas grafik, hasilnya adalah:

- MSE untuk regresi linear adalah 25.00.
- MSE untuk regresi polynomial derajat 2 adalah 0.00.

Angka MSE ini menunjukkan seberapa besar kesalahan prediksi model. MSE yang lebih kecil berarti model tersebut lebih akurat. Karena MSE model polynomial adalah 0, artinya model ini sangat tepat dalam memprediksi data. Sementara MSE model linear lebih besar (25), menunjukkan bahwa model linear memiliki banyak kesalahan dalam memprediksi data.

