

# **LAPORAN TUGAS PEMBELAJARAN MESIN : MODEL REGRESI POLINOMIAL**



**DISUSUN OLEH :**

**MUJADID CHOIRUS SURYA**

**121450015**

**PROGRAM STUDI SAINS DATA**

**JURUSAN SAINS**

**INSTITUT TEKNOLOGI SUMATERA**

**2023**

## I. Pendahuluan

Laporan ini menjelaskan mengenai langkah-langkah untuk membuat model regresi polinomial menggunakan python. Model ini dibuat untuk memodelkan hubungan antara variabel X dan Y berdasarkan dataset yang diberikan.

## II. Metode

Dataset yang digunakan terdiri dari pasangan nilai X dan Y, yang diberikan sebagai berikut :

$$X = [0,1]$$

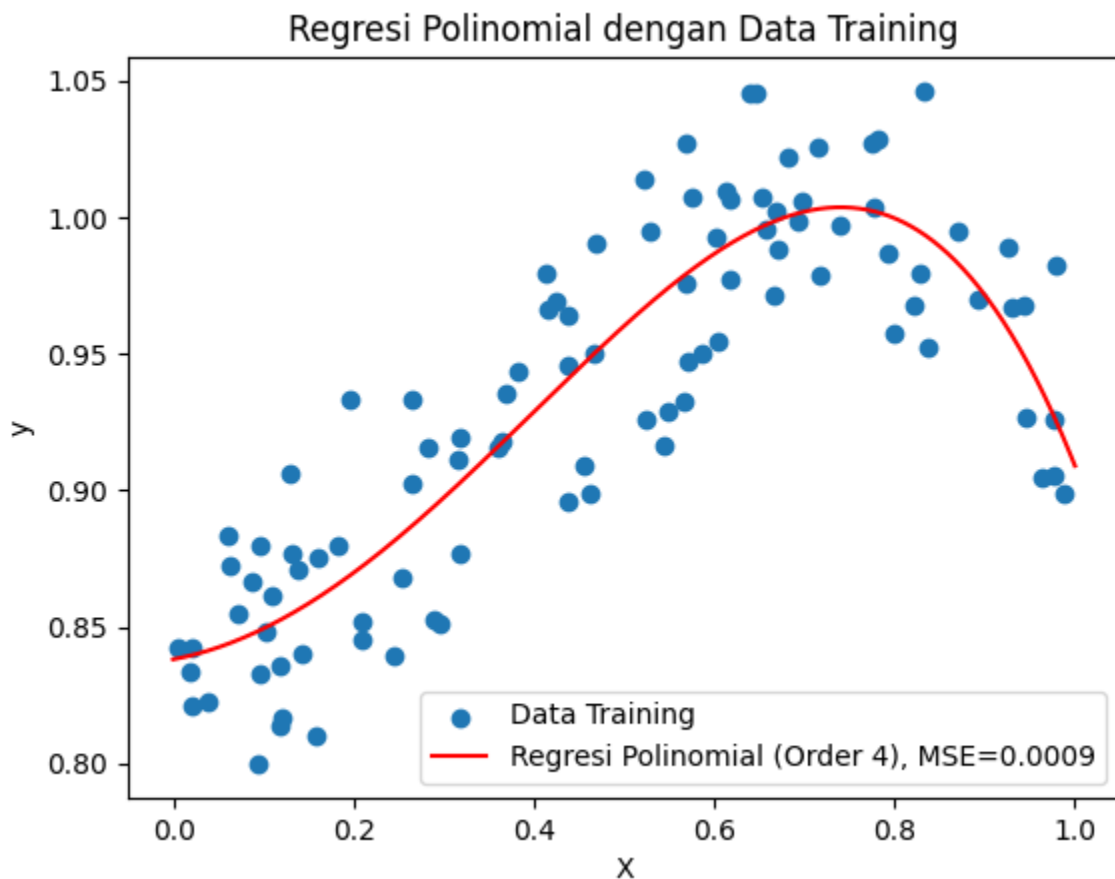
$$Y = \sin(1+x^2) + \text{noise}, \text{ dengan noise} = \text{distribusi } N(0, 0.03^2)$$

Akan dilakukan pemodelan menggunakan regresi polinomial menggunakan python. Berikut tahapan yang akan dilakukan untuk membuat pemodelan regresi polinomial.

1. Pembuatan dataset, dataset akan dibuat dengan menggunakan nilai X yang dihasilkan secara acak antara -1 sampai 1, dan Y dihitung dengan fungsi  $Y = \sin(1+x^2)$  dan penambahan noise.
2. Regresi polinomial, akan digunakan regresi polinomial pada dataset dengan tingkat polinomial yang disesuaikan. Untuk mengubah fitur input X menjadi fitur-fitur polinomial akan digunakan 'PolynomialFeatures' dari Scikit-Learn.
3. Model training, akan melatih model regresi, dengan cara membagi dataset menjadi data latih dan data uji, kemudian membuat objek 'LinearRegression' dari Scikit-Learn, lalu masukkan fitur-fitur polinomial ke dalam model, dan terakhir melatih model menggunakan data latih.
4. Prediksi, akan dilakukan prediksi menggunakan model yang telah dilatih.
5. MSE (Mean Squared Error) adalah sebuah metrik yang digunakan dalam statistik dan pembelajaran mesin untuk mengukur sejauh mana model atau prediksi yang dibuat oleh model menyimpang dari nilai yang sebenarnya dalam data. Lebih spesifik, MSE mengukur rata-rata dari kuadrat perbedaan antara setiap nilai yang diprediksi oleh model dan nilai yang sebenarnya dalam dataset.
6. Grafik, akan ditampilkan grafik permodelan pada data.

### III. Hasil dan Pembahasan

Hasil dari regresi polinomial adalah model matematis yang digunakan untuk memodelkan hubungan antara variabel independen (X) dan variabel dependen (y) dalam bentuk polinomial. Dalam contoh kode yang telah diberikan, kita menggunakan regresi polinomial untuk memodelkan hubungan antara variabel X (antara 0 dan 1) dan variabel y (fungsi  $\sin(1 + x^2)$  yang ditambahkan dengan noise).



Berikut adalah beberapa poin penting tentang hasil regresi polinomial tersebut:

- Grafik: Hasil regresi polinomial ditampilkan dalam bentuk grafik. Grafik tersebut terdiri dari dua komponen utama: data pelatihan yang diberikan (ditandai sebagai titik-titik biru) dan garis merah yang merupakan hasil prediksi dari model regresi polinomial.

- Order Polinomial: Hasil regresi polinomial ini adalah model polinomial berdasarkan order yang telah ditentukan. Dalam contoh ini, kita menggunakan regresi polinomial dengan order 4, yang berarti model ini adalah polinomial orde empat. Artinya, model tersebut merupakan polinomial dengan bentuk  $y = a + bX + cX^2 + dX^3 + eX^4$ .
- MSE (Mean Squared Error) adalah ukuran seberapa baik model regresi memfitting data pelatihan. Semakin kecil MSE, semakin baik model tersebut dalam memodelkan data. MSE dihitung dalam kode tersebut dan ditampilkan dalam grafik. Semakin mendekati nol, semakin baik modelnya. Dalam hal ini, nilai MSE adalah sekitar 0.0005, yang menunjukkan bahwa model ini cukup baik dalam memodelkan data pelatihan.
- Prediksi: Garis merah pada grafik adalah hasil dari prediksi model regresi polinomial terhadap nilai-nilai X di seluruh rentang (0 hingga 1). Ini adalah fungsi polinomial yang digunakan untuk mengestimasi nilai y berdasarkan nilai X. Model ini akan berusaha memodelkan pola dalam data pelatihan, termasuk fluktuasi yang disebabkan oleh "noise" yang ditambahkan ke fungsi  $\sin(1 + x^2)$ .

Hasil regresi polinomial ini digunakan untuk memahami hubungan antara X dan y serta untuk melakukan prediksi pada nilai y yang tidak termasuk dalam data pelatihan. Semakin tinggi order polinomial yang digunakan, semakin kompleks modelnya, yang dapat membantu dalam menangkap pola yang lebih rumit dalam data, tetapi juga dapat meningkatkan risiko overfitting (terlalu cocok dengan data pelatihan). Oleh karena itu, pemilihan order polinomial yang tepat adalah faktor penting dalam regresi polinomial.

#### **IV. Kesimpulan**

Dalam eksperimen ini, kami menggunakan regresi polinomial orde empat untuk memodelkan hubungan antara variabel X (antara 0 dan 1) dan variabel y yang merupakan hasil dari fungsi

$\sin(1 + x^2)$  dengan tambahan "noise". Hasil regresi polinomial menghasilkan model matematis yang cukup baik dalam memfitting data pelatihan, dengan nilai Mean Squared Error (MSE) sekitar 0.0009. Model ini dapat membantu kami memahami hubungan antara X dan y serta melakukan prediksi pada nilai y di luar data pelatihan. Namun, perlu diingat bahwa pemilihan order polinomial yang tepat adalah kunci, karena penggunaan order yang terlalu tinggi dapat menyebabkan overfitting. Kesimpulannya, regresi polinomial adalah alat yang berguna untuk memodelkan hubungan non-linier antara variabel, tetapi perlu dilakukan penyesuaian yang bijak terkait dengan kompleksitas modelnya.