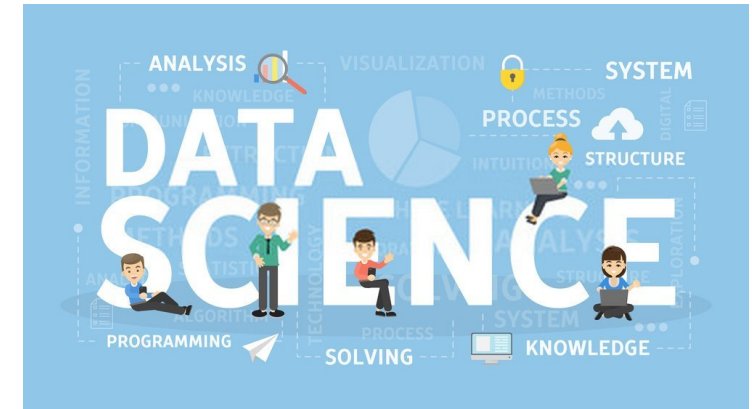


# Pengantar Data Science

## Prodi : Sistem Informasi

1. Konsep regresi, apa yang disebut MODEL pada Data Science
2. Regresi Linear (dasar konsep dan matematika)
3. Regresi Non Linear (dasar konsep dan matematika)
4. Contoh ilustrasi penggunaan Regresi



1. Pengantar Data Science
2. Pengenalan Berbagai Metoda DS
3. Pengalan Tools untuk Data
4. Pengenalan Tools Data Science
5. Data Understanding
6. Data Preparation (1)
7. Data Preparation (2)
8. Ujian Mid

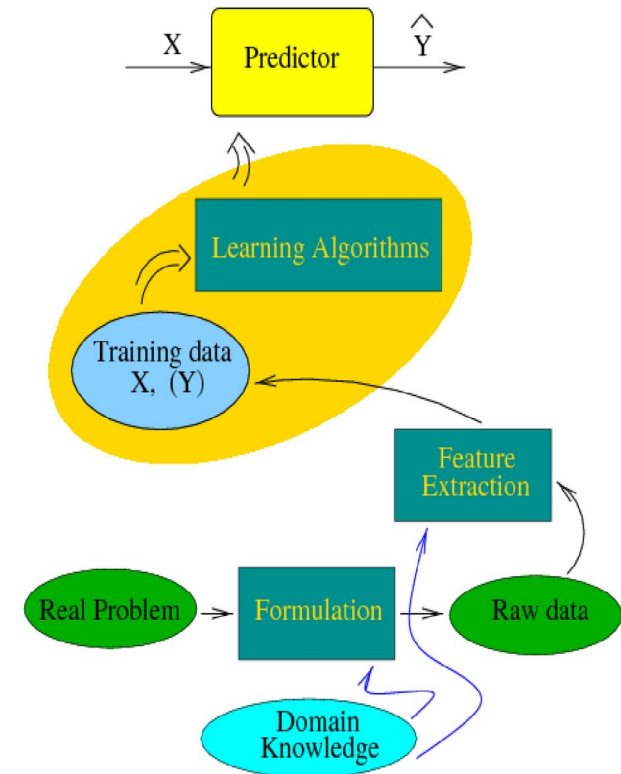
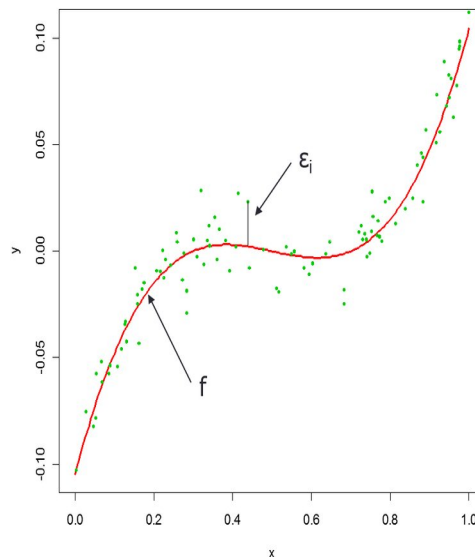
9. **Membangun Model (1)**
10. Membangun Model (2)
11. Membangun Model (3)
12. Membangun Model (4)
13. Menggunakan dan menguji Model
14. MLOps dan Pipelining
15. Mengenal teknologi deployment
16. Ujian Akhir



# Konsep Regresi

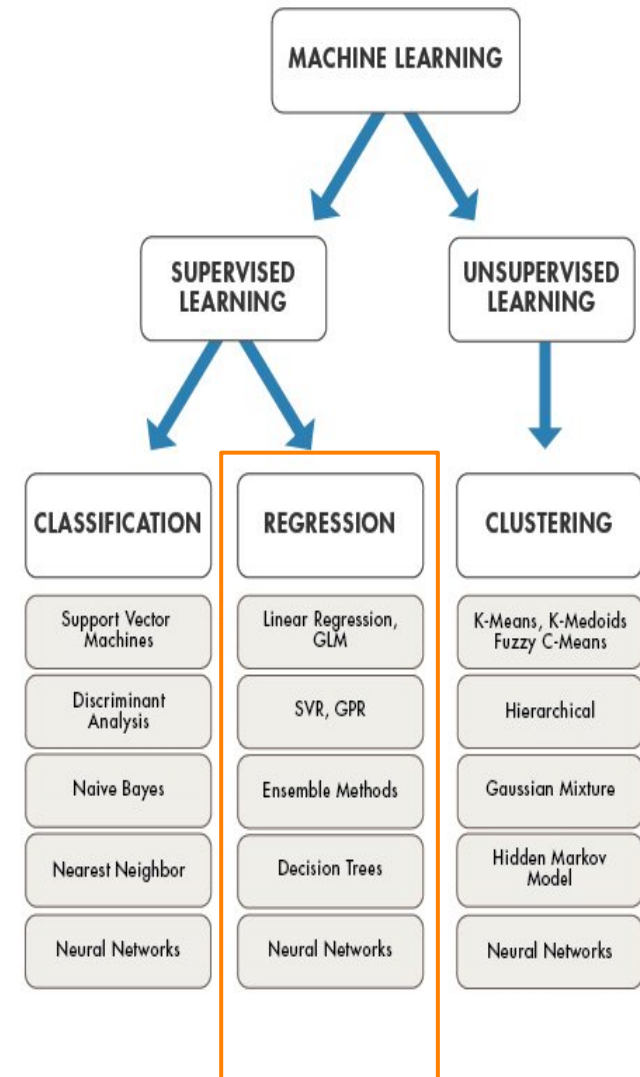
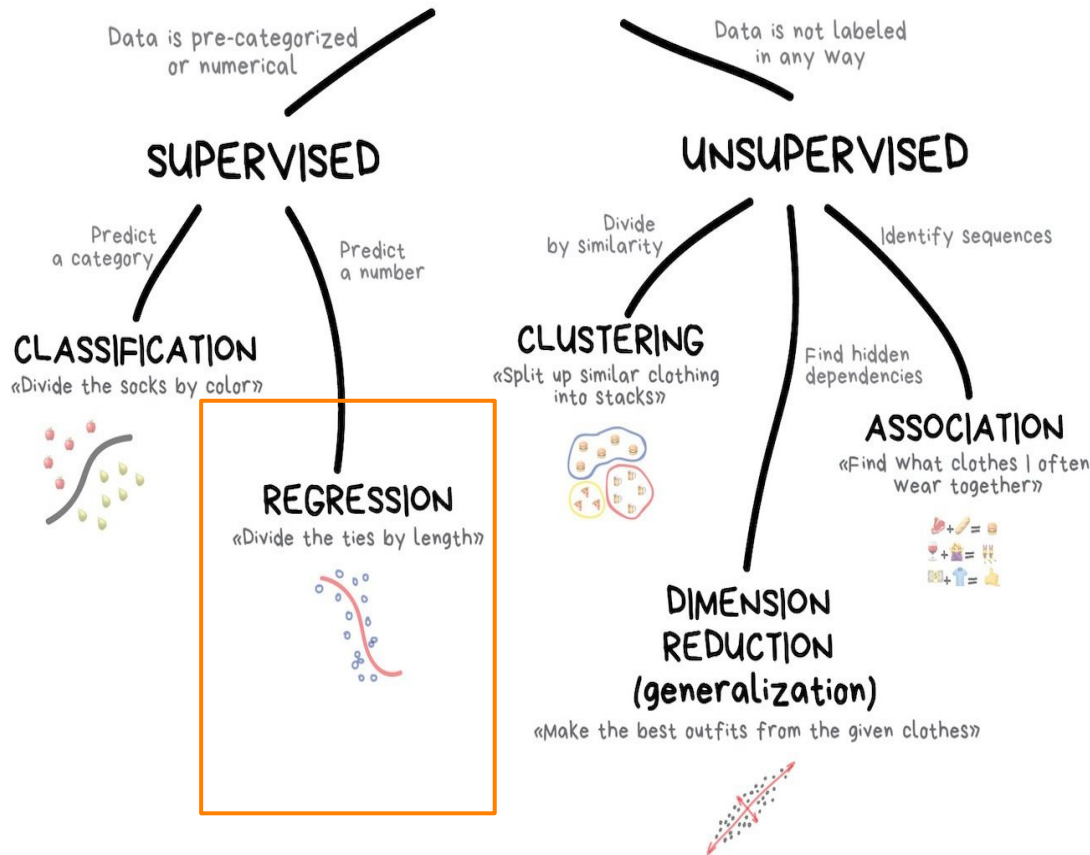
# Machine Learning

- Pembelajaran dari data digunakan pada situasi, saat tidak ada solusi analitis, tetapi kita dapat menggunakan data untuk membuat suatu solusi empiris
- Premis dasar pada proses ini adalah menggunakan suatu himpunan pengamatan untuk menguak proses yang ada di dalamnya (underlying process)
- Anggap kita mengobservasi suatu ruang ouput dan ruang input.
- Ada suatu hubungan antara  $Y$  dan paling tidak satu  $X$ . Maka dapat dimodelkan hubungan tersebut sebagai  $f$ , dimana  $f$  adalah suatu fungsi yang tidak diketahui dan  $\varepsilon$  adalah random error (noise), tak bergantung dari  $X$  dengan mean nol.

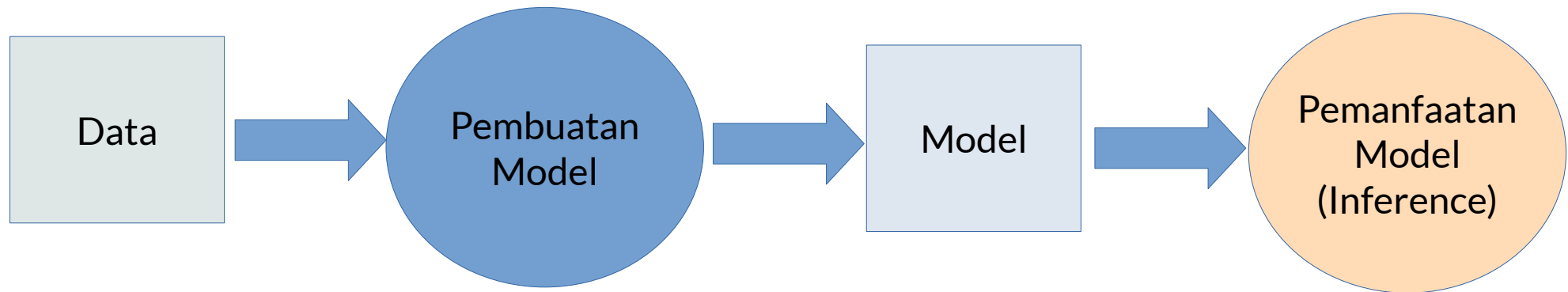


# Regresi dalam Machine Learning

## CLASSICAL MACHINE LEARNING

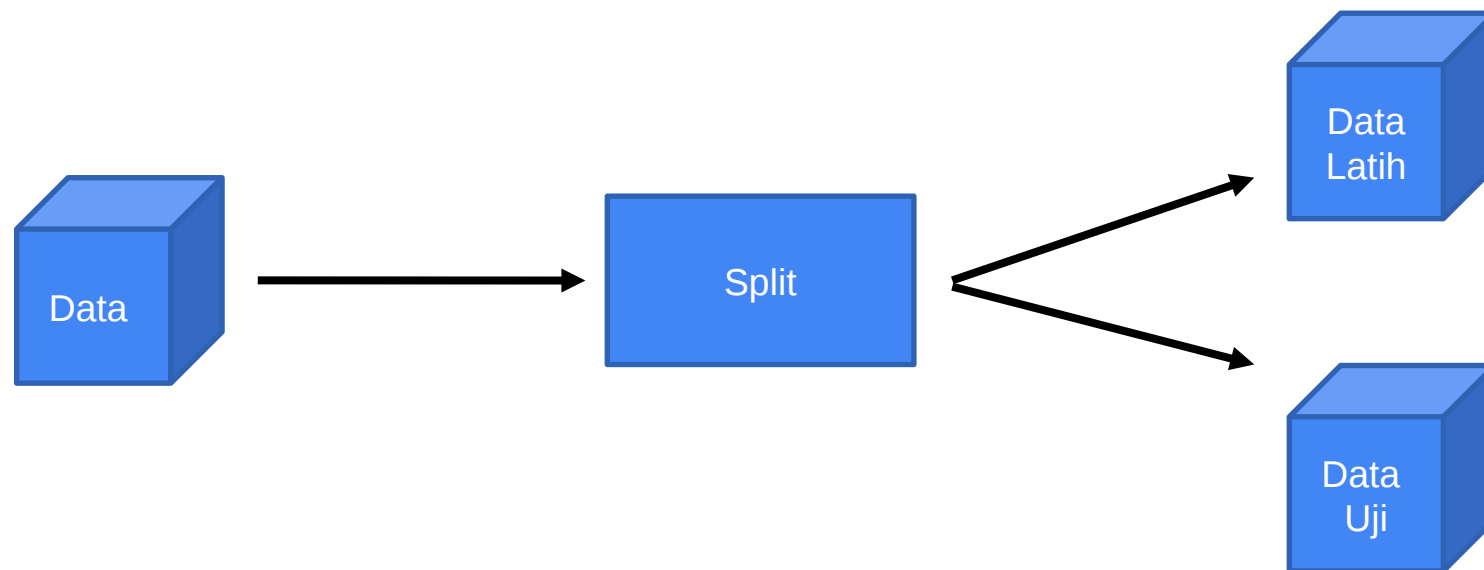


# Pembuatan Model

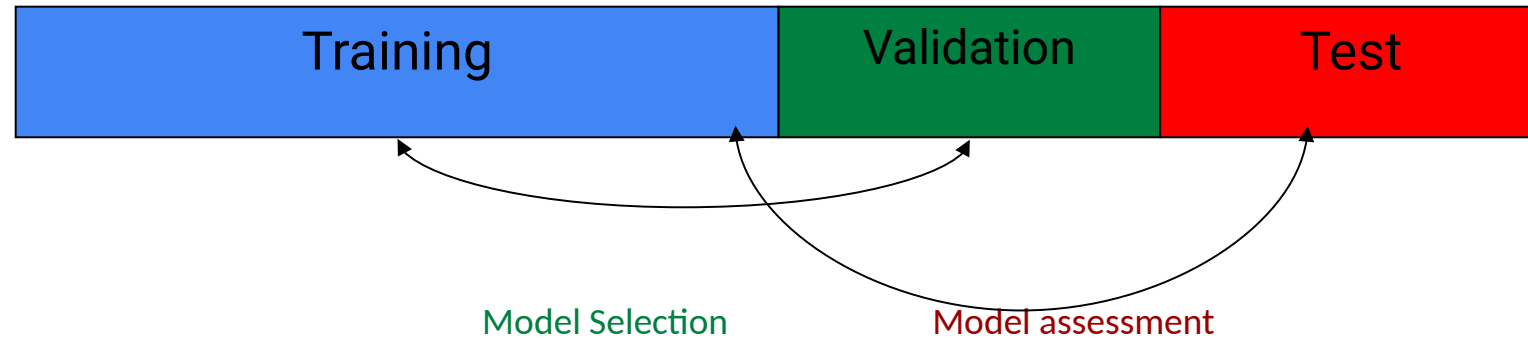


# Pembagian Data

- Data dibagi menjadi 2 bagian :
  - Data Latih (Training Data) : untuk mengembangkan model
  - Data Uji (Testing Data) : untuk Mengukur performansi model



# Training – Validation – Testing Data



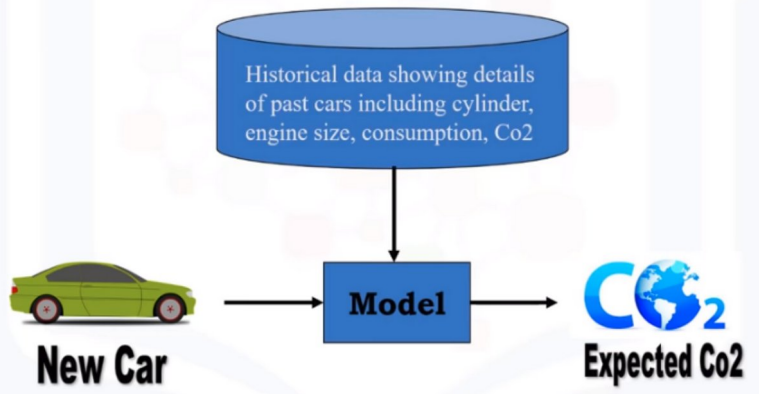
- Model Selection : Mengestimasi performa model - model yang berbeda untuk memilih model yang terbaik, yaitu model dengan minimum error
- Model Assessment : Dari model yang terpilih, mengestimasi error untuk data baru (data uji)



# Pengertian Regresi

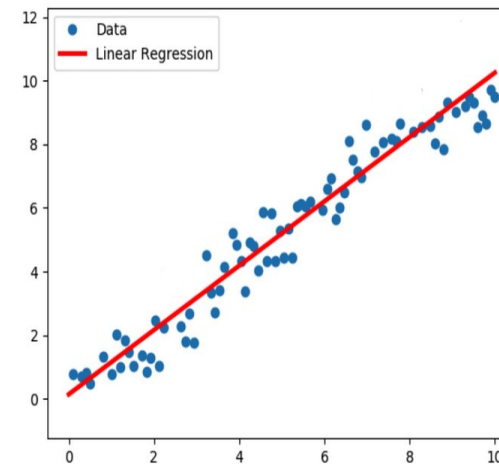
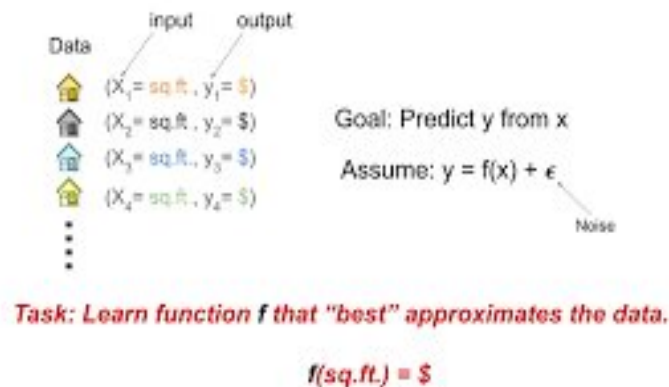
x : variabel bebas

y : variabel tak bebas



	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?

Regresi adalah proses  
Memprediksi nilai kontinu





# Tipe Model Regresi

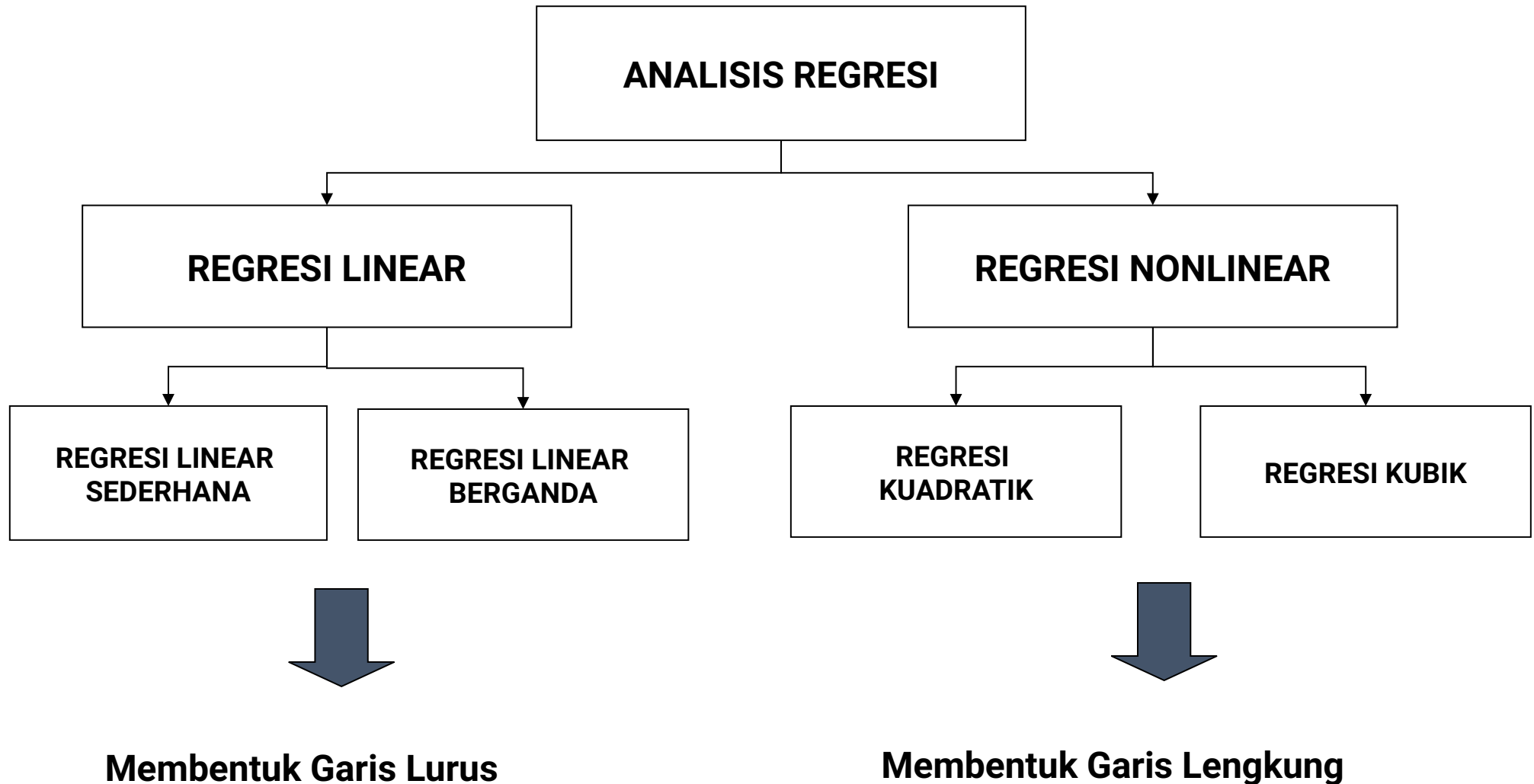
## Regresi Sederhana:

- Regresi sederhana linier
- Regresi sederhana non-linier
- Contoh: memprediksi co2emission vs EngineSize dari semua mobil.

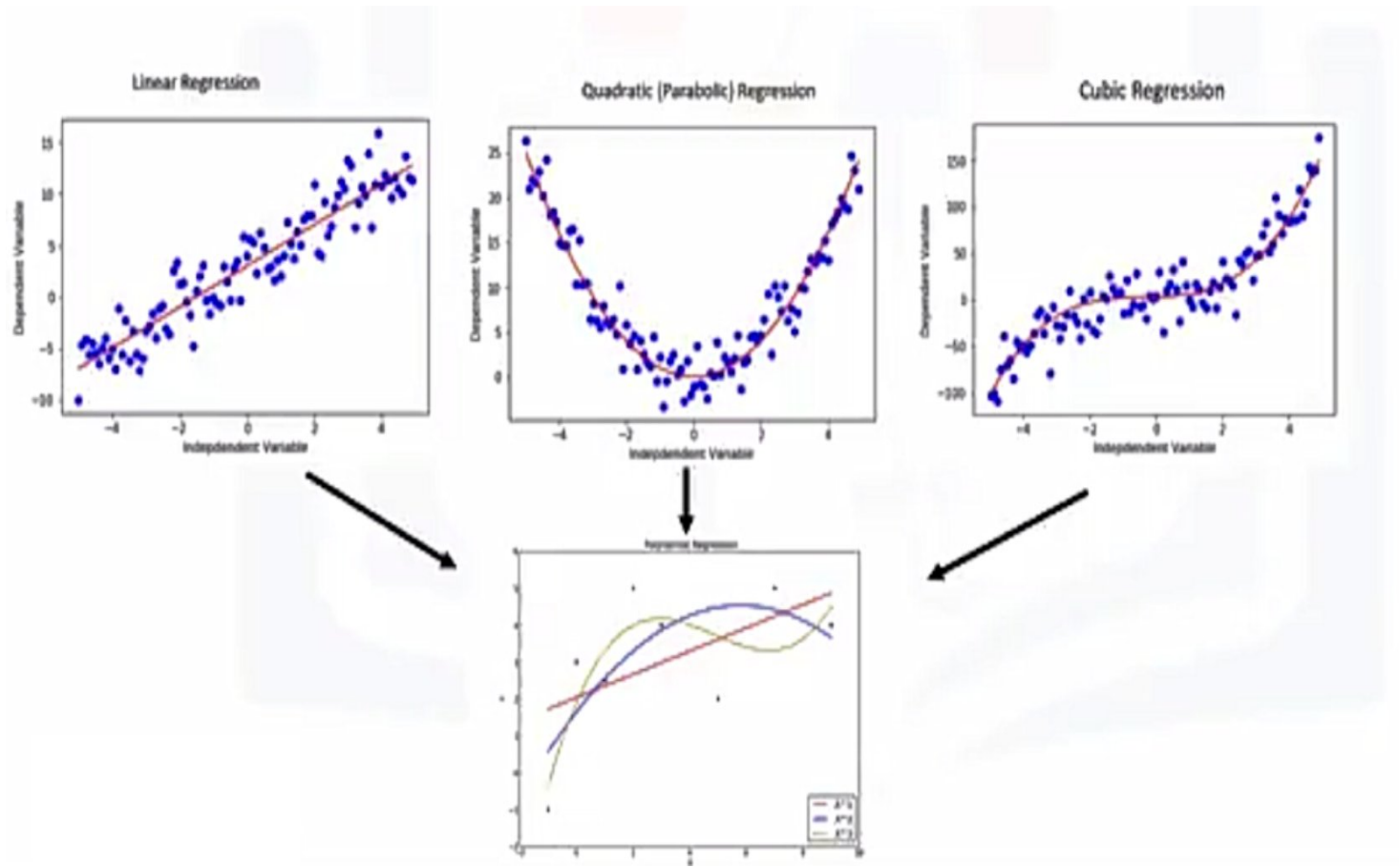
## Regresi Variabel Jamak:

- Regresi variabel jamak linier
- Regreasi variabel jamak non-linier
- Contoh: meprediksi co2emission vs EngineSize dan Cylinders dari semua mobil.

# Jenis Regresi



# Tipe Regresi



# Pendahuluan

- Penggunaan statistika dalam mengolah data penelitian berpengaruh terhadap tingkat analisis hasil penelitian.
- Penelitian-penelitian dalam bidang ilmu pengetahuan alam (science) yang menggunakan perhitungan-perhitungan statistika, akan menghasilkan data yang mendekati benar, jika memperhatikan tata cara analisis data yang digunakan.
- Dalam memprediksi dan mengukur nilai dari pengaruh satu variabel (bebas/independent/ predictor) terhadap variabel lain (tak bebas/ dependent/ response) dapat digunakan uji regresi

# Contoh Regresi Linier Variabel Jamak

Efektivitas variabel-variabel bebas terhadap prediksi

- Apakah kegelisahan, kehadiran dosen, dan jenis kelamin mempunyai efek pada kinerja ujian mahasiswa?

Prediksi dampak perubahan

- Seberapa besar kenaikan/penurunan tekanan darah terhadap kenaikan/penurunan BMI dari pasien?

# Prediksi Nilai Kontinu pada Regresi Linier Variabel Jamak

X: Independent variable      Y: Dependent variable

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?

$$Co2 Em = \theta_0 + \theta_1 Engine\ size + \theta_2 Cylinders + \dots$$

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$$

$$\hat{y} = \theta^T X$$

$$\theta^T = [\theta_0, \theta_1, \theta_2, \dots] \quad X = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \dots \end{bmatrix}$$

- Regresi linier variabel jamak menggunakan lebih dari satu variabel bebas antara lain ENGINE SIZE, CYLINDERS, FUEL CONSUMPTION\_COMB
- Untuk memprediksi nilai kontinu variabel tak bebas dalam hal ini CO2 EMISSION sebagaimana ditunjukkan pada Gambar 9.

# Prediksi Menggunakan Regresi Linier Variabel Jamak

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?

$$\hat{y} = \theta^T X$$

$$\theta^T = [125, 6.2, 14, \dots]$$

$$\hat{y} = 125 + 6.2x_1 + 14x_2 + \dots$$

$$Co2Em = 125 + 6.2EngSize + 14 Cylinders + \dots$$

$$Co2Em = 125 + 6.2 \times 2.4 + 14 \times 4 + \dots$$

$$Co2Em = 214.1$$

Gambar 11 menunjukkan bagaimana prediksi nilai numerik CO2EMISSIONS berdasarkan variabel bebas jamak yaitu ENGINE SIZE = 2,4 serta CYLINDERS = 4 dan FUEL CONSUMPTION = 9,2 dengan hasil prediksi = 214,1 menggunakan parameter terbaik yang sudah didapatkan dari data latih.



# Apa itu Regresi?

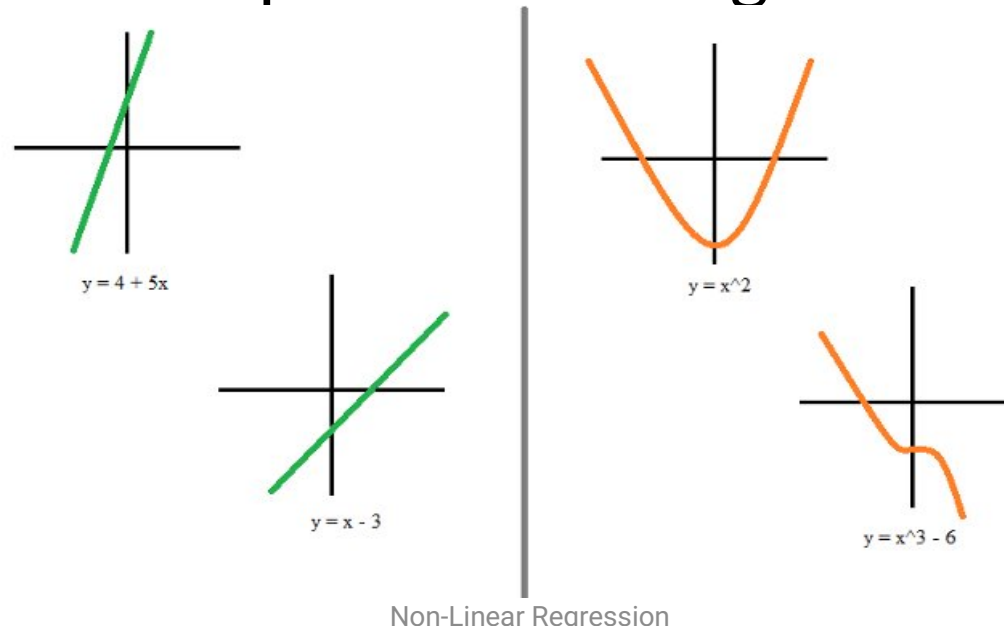
- Analisis/uji regresi merupakan suatu kajian dari hubungan antara satu variabel, yaitu variabel yang diterangkan (the explained variabel) dengan satu atau lebih variabel, yaitu variabel yang menerangkan (the explanatory).
- Apabila variabel bebasnya hanya satu, maka analisis regresinya disebut dengan regresi linear sederhana.
- Apabila variabel bebasnya lebih dari satu, maka analisis regresinya dikenal dengan regresi linear berganda.
  - Dikatakan berganda karena terdapat beberapa variabel bebas yang mempengaruhi variabel tak bebas

## Apa itu Regresi? (2)

- Analisis/uji regresi banyak digunakan dalam perhitungan hasil akhir untuk penulisan karya ilmiah/penelitian.
- Hasil perhitungan analisis/uji regresi akan dimuat dalam kesimpulan penelitian dan akan menentukan apakah penelitian yang sedang dilakukan berhasil atau tidak.
- Analisis perhitungan pada uji regresi menyangkut beberapa perhitungan statistika seperti uji signifikansi (uji-t, uji-F), anova dan penentuan hipotesis.
- Hasil dari analisis/ uji regresi berupa suatu persamaan regresi.
  - Persamaan regresi ini merupakan suatu fungsi prediksi variable yang mempengaruhi variabel lain

# Linear VS Non-Linear

- Bagaimana Anda bisa menyatakan bahwa dataset yang ada, secara alamiah bersifat linear atau non-linear?
- Model yang dipilih untuk regresi, sangat bergantung kepada dataset itu sendiri.
- Mari tinjau kembali perbedaan fungsi linear dan non linear:



## Linear VS Non-Linear (2)

- Fungsi linear: Secara sederhana didefinisikan sebagai fungsi yang mengikuti prinsip:
  - Input/output = konstan
- Persamaan linear selalu merupakan polinomial derajat 1 (misalnya  $x + 2y + 3 = 0$ ).
- Dalam kasus dua dimensi, fungsi linear selalu membentuk garis
- Pada dimensi lain, fungsi linear mungkin juga membentuk bidang atau titik. "Bentuk" fungsi linear selalu benar-benar lurus, tanpa kurva apa pun.

# Algoritma Regresi

- Linier Regression
- Polynomial Regression
- Support Vector Regression
- Decision Tree Regression
- Random Forest Regression
- LASSO Regression
- ANN Regression
- K-NN Regression
- dst.



# Regresi Linear

# Regresi Linear Sederhana

- Regresi Linear Sederhana adalah Metode Statistik yang berfungsi untuk menguji sejauh mana hubungan sebab akibat antara Variabel Faktor Penyebab (X) terhadap Variabel Akibatnya.
- Faktor Penyebab pada umumnya dilambangkan dengan X atau disebut juga dengan Predictor sedangkan Variabel Akibat dilambangkan dengan Y atau disebut juga dengan Response.
- Regresi Linear Sederhana atau sering disingkat dengan SLR (Simple Linear Regression) juga merupakan salah satu Metode Statistik yang dipergunakan dalam produksi untuk melakukan peramalan ataupun prediksi tentang karakteristik kualitas maupun kuantitas.



# Contoh Penggunaan

- Contoh penggunaan analisis Regresi Linear Sederhana dalam kegiatan produksi, antara lain:
  - Hubungan antara lamanya kerusakan mesin dengan kualitas produk yang dihasilkan
  - Hubungan jumlah pekerja dengan output yang diproduksi
  - Hubungan antara suhu ruangan dengan cacat produksi yang dihasilkan.

# Regresi Linier Untuk Memprediksi Nilai Kontinu

x : variabel bebas

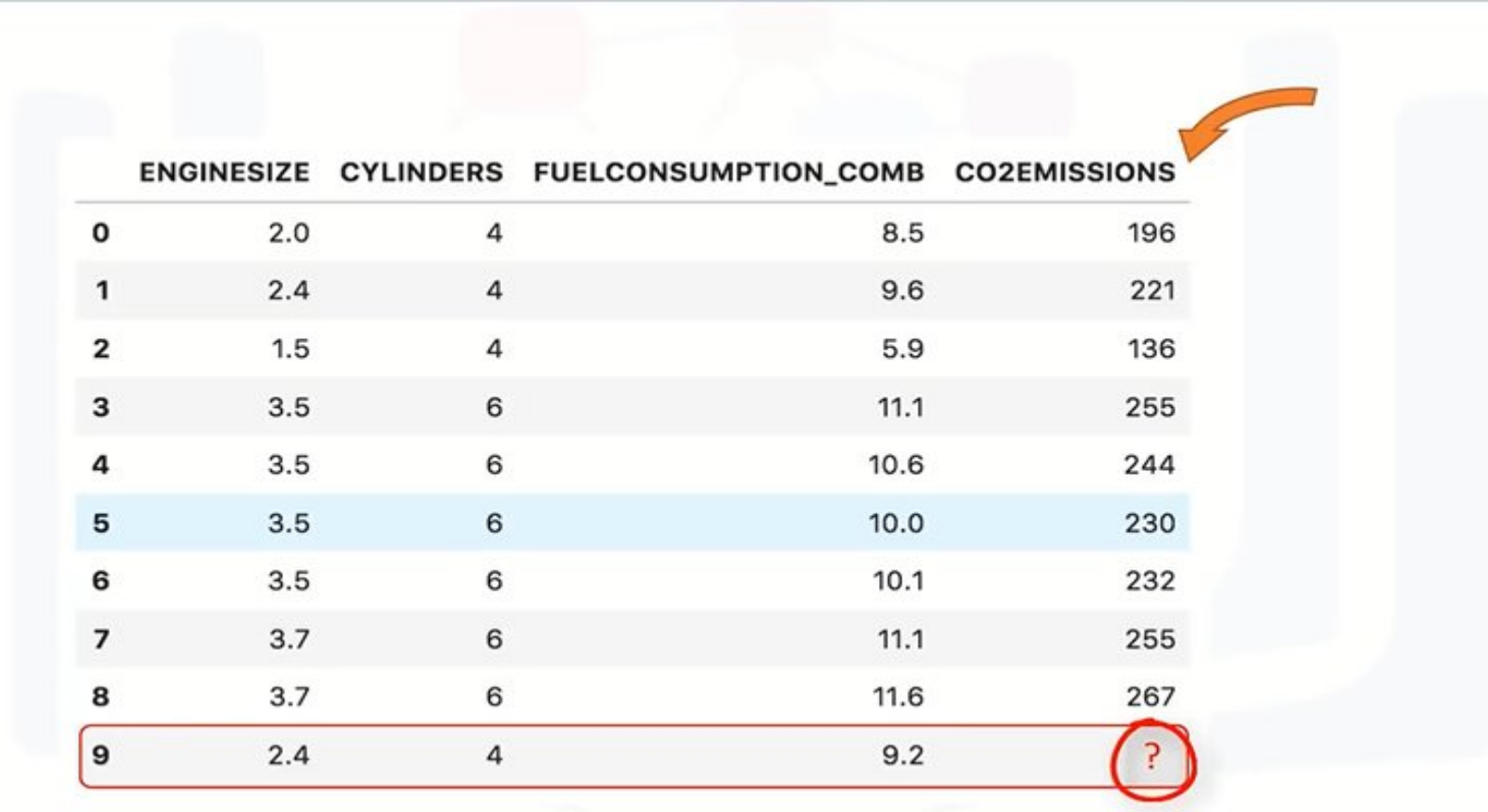
y : variabel tak bebas

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?



Nilai kontinyu / numerik

# Menggunakan Regresi Untuk Melakukan Prediksi Pada Data yang Kontinyu



	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?

# Menggunakan Regresi Untuk Melakukan Prediksi Pada Data yang Kontinyu

X: Independent variable

Y: Dependent variable

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?

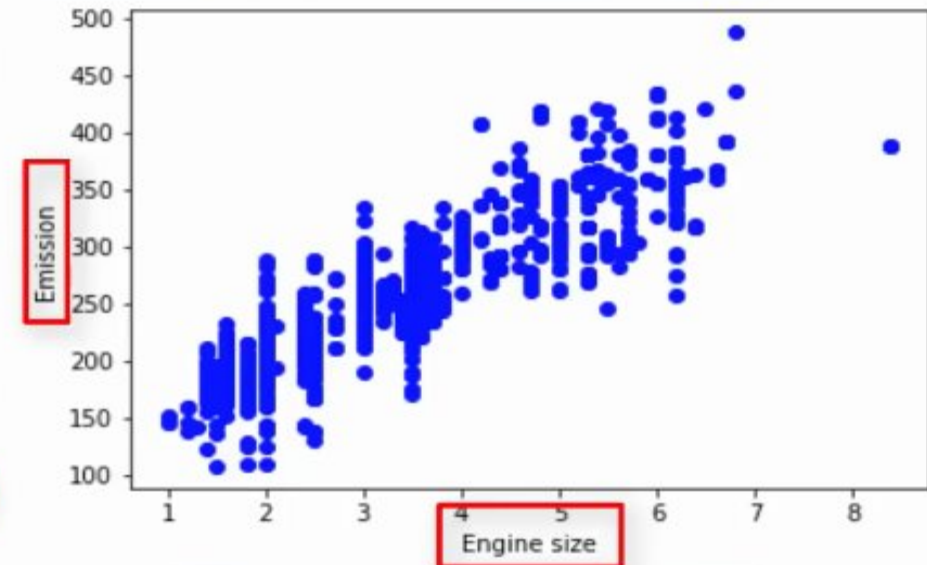
Continuous Values

# Topologi Regresi Linear

- Regresi Linear Sederhana :
  - Prediksi emisi Co2 VS Ukuran mesin (engine size)
    - Variabel Independen (X): Ukuran mesin (engine size)
    - Variabel Dependen (Y): Emisi Co2
- Regresi Linear Berganda:
  - Prediksi emisi Co2 VS Ukuran mesin (engine size) dan Silinder
    - Variabel Independen (X): Ukuran mesin (engine size), Silinder
    - Variabel Dependen (Y): Emisi Co2

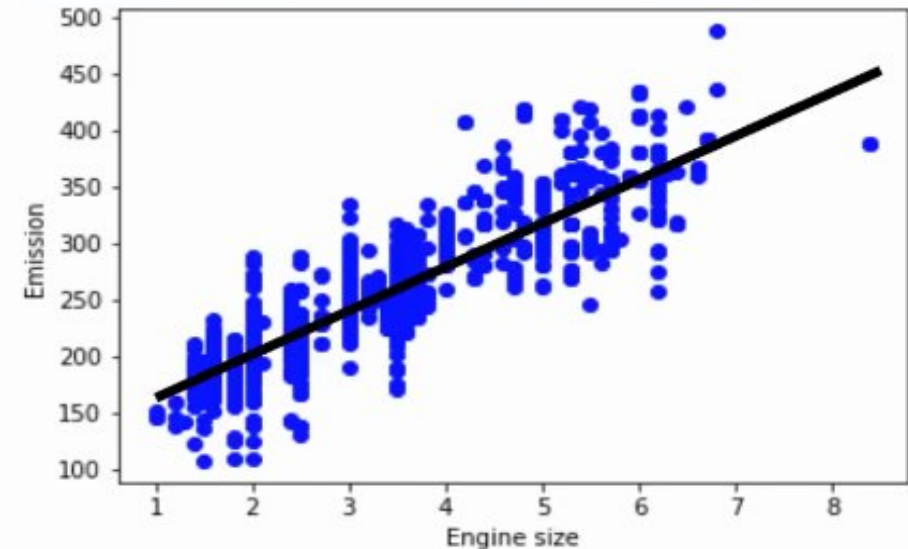
# Bagaimana Regresi Linear Bekerja?

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?



# Bagaimana Regresi Linear Bekerja?

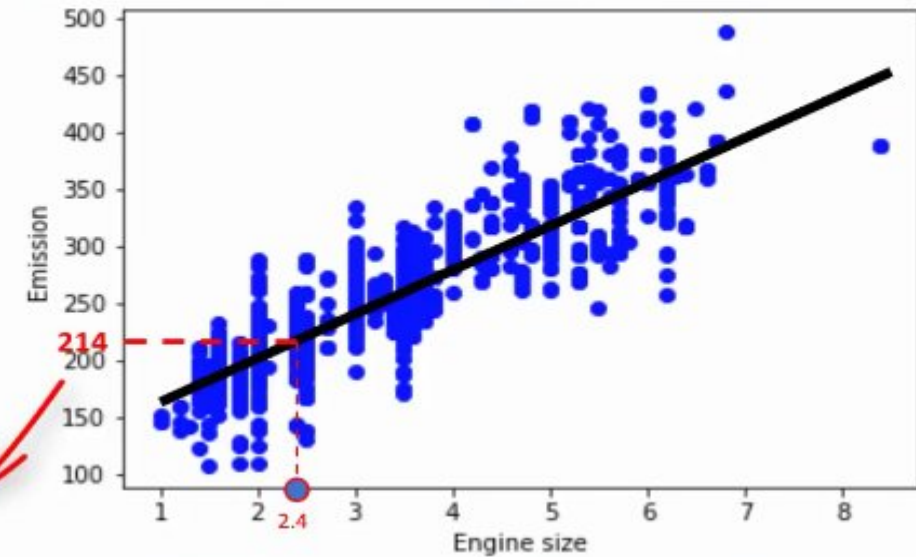
	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?



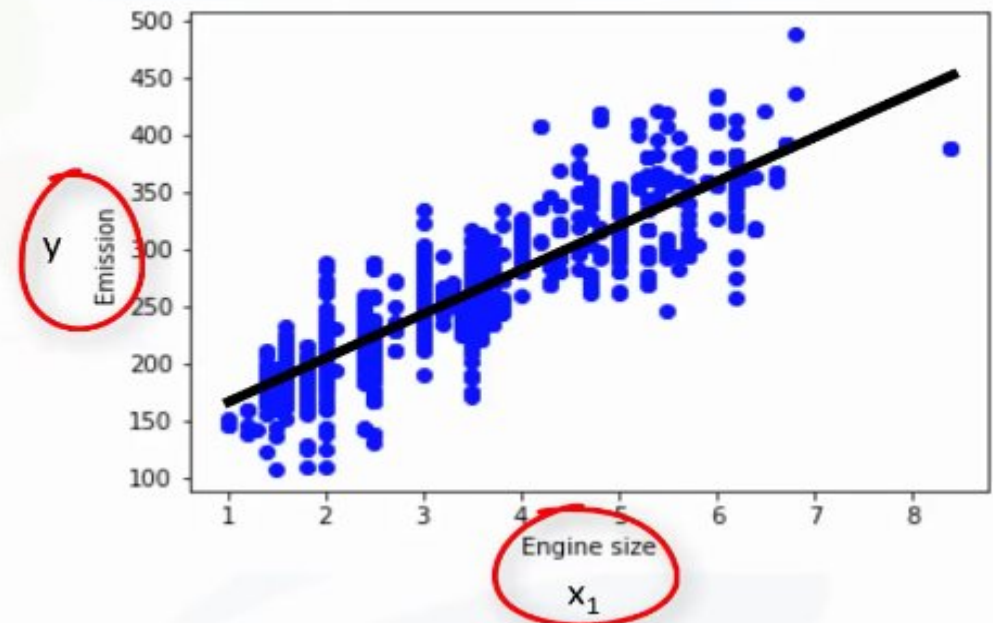


# Bagaimana Regresi Linear Bekerja?

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?



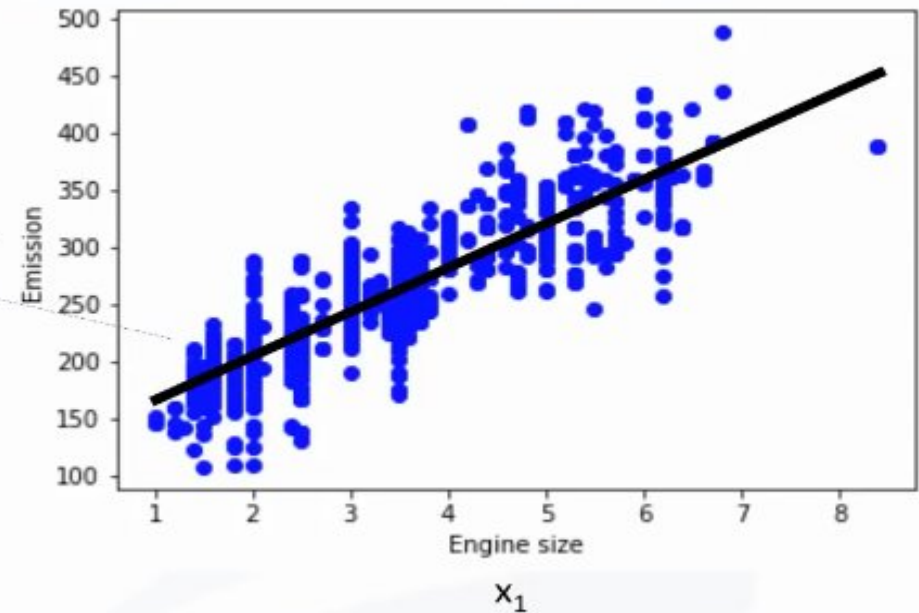
# Representasi Model Regresi Linear



# Representasi Model Regresi Linear

$$\hat{y} = \theta_0 + \theta_1 x_1$$

y

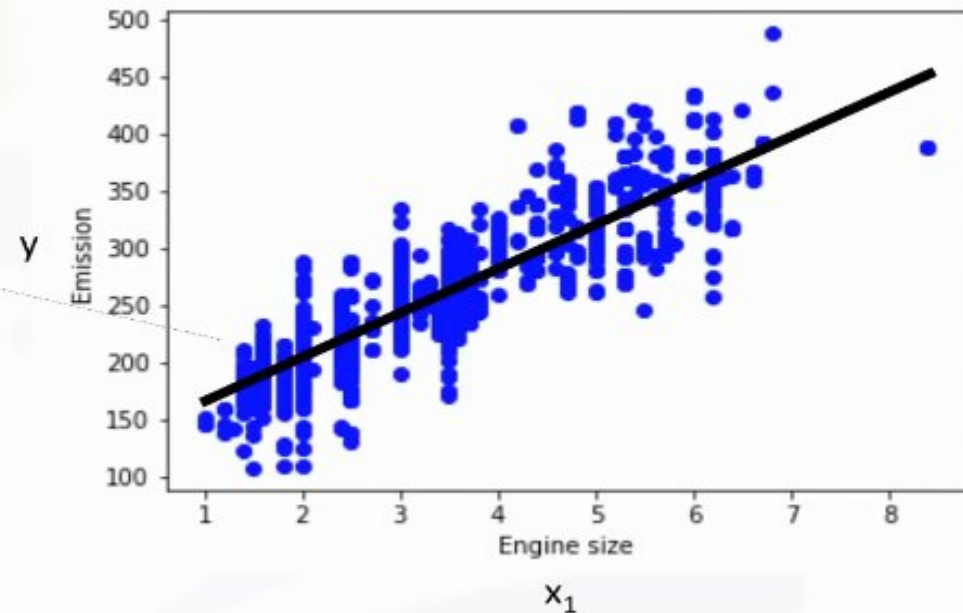


# Representasi Model Regresi Linear

$$\hat{y} = \theta_0 + \theta_1 x_1$$

response variable

a single predictor

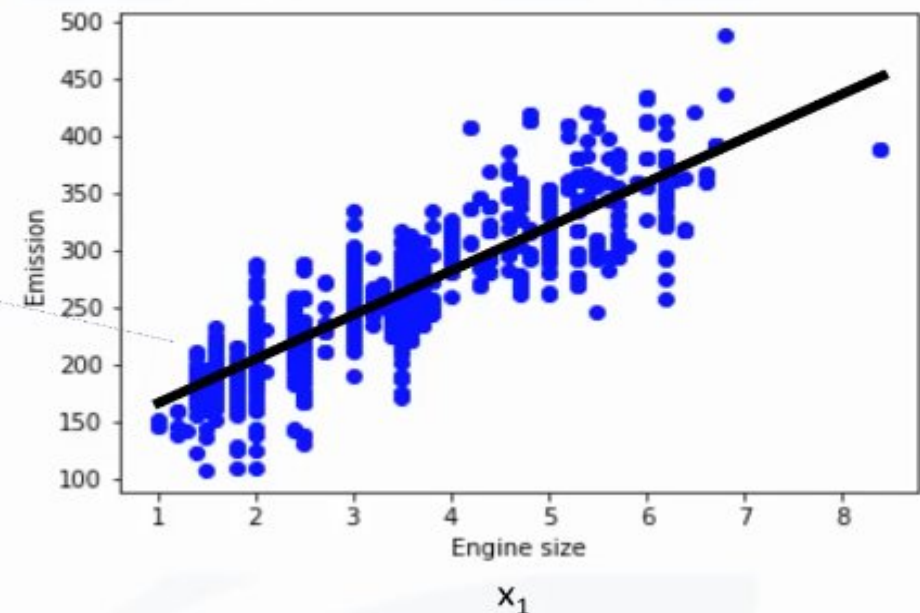


# Representasi Model Regresi Linear

$$\hat{y} = \theta_0 + \theta_1 x_1$$

response variable

a single predictor

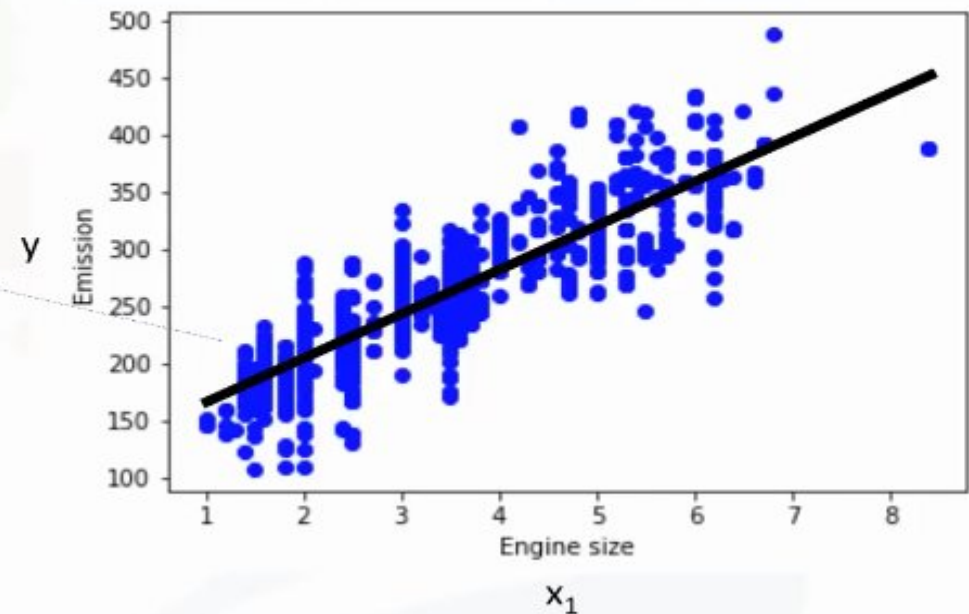


# Representasi Model Regresi Linear

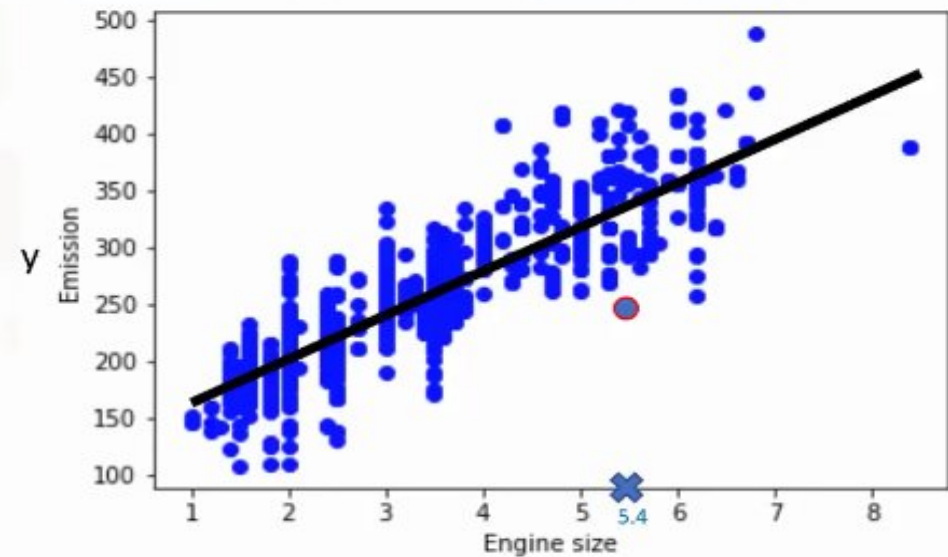
$$\hat{y} = \theta_0 + \theta_1 x_1$$

response variable

a single predictor

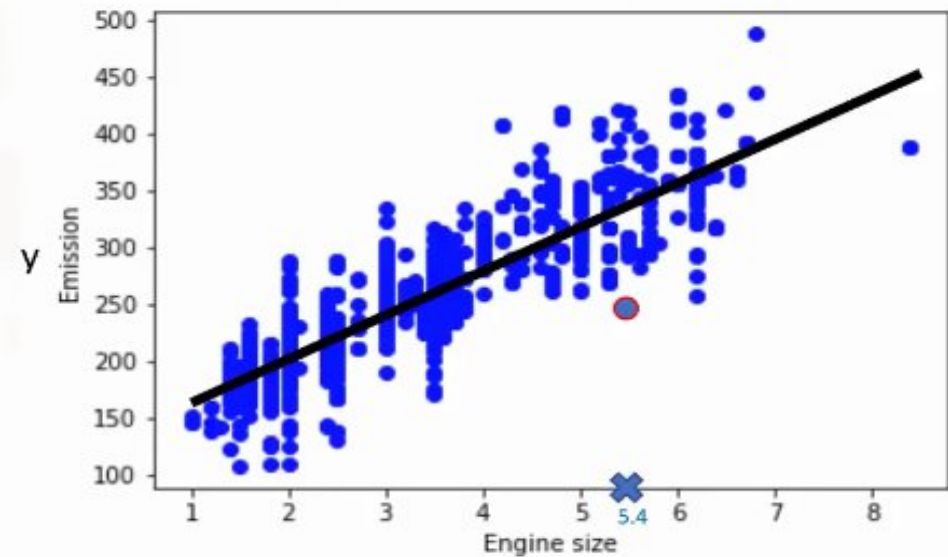


# Bagaimana Mencari Model yang Fit





# Bagaimana Mencari Model yang Fit



# Bagaimana Mencari Model yang Fit

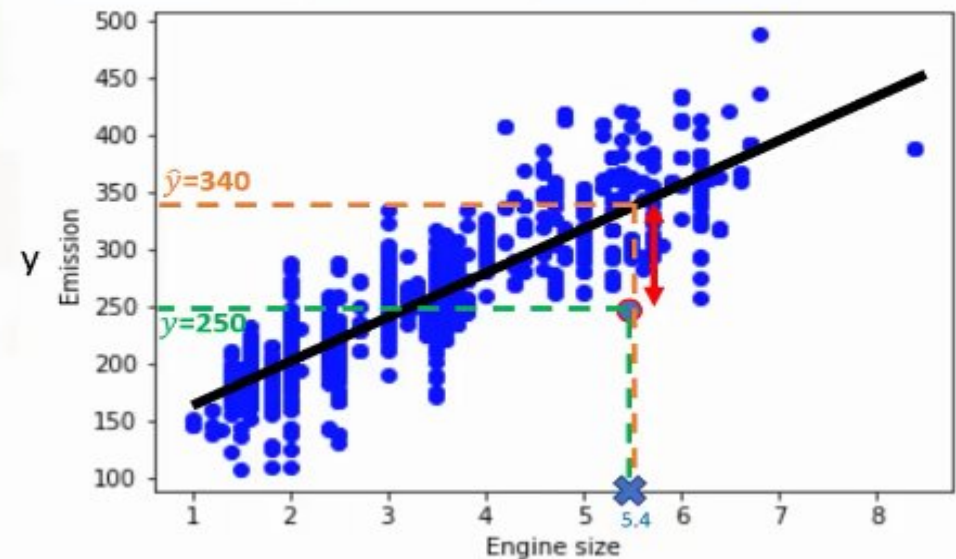
$x_1 = 2.4$  independent variable

$y = 250$  actual Co2 emission of  $x_1$

$$\hat{y} = \theta_0 + \theta_1 x_1$$

$\hat{y} = 340$  the predicted emission of  $x_1$

$$\begin{aligned} \text{Error} &= y - \hat{y} \\ &= 250 - 340 \\ &= -90 \end{aligned}$$



# Bagaimana Mencari Model yang Fit

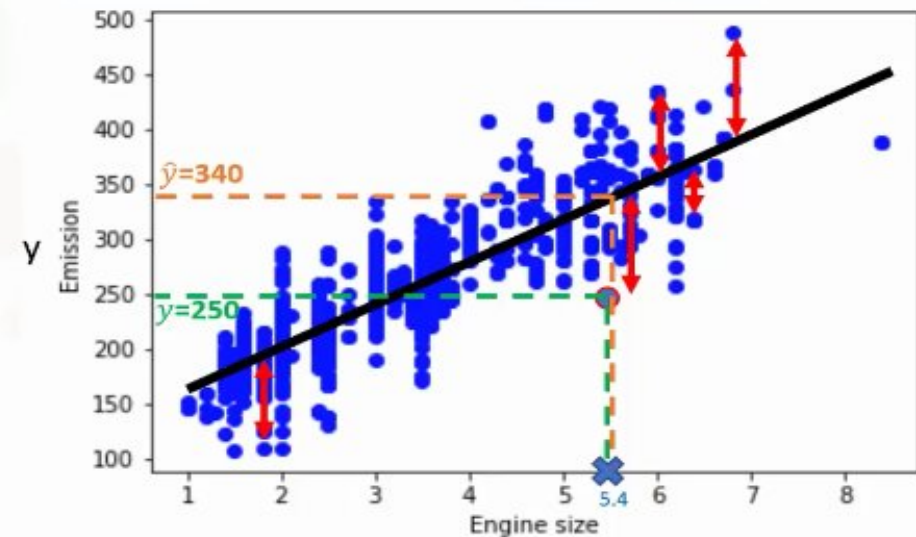
$x_1 = 2.4$  independent variable  
 $y = 250$  actual Co2 emission of  $x_1$

$$\hat{y} = \theta_0 + \theta_1 x_1$$

$\hat{y} = 340$  the predicted emission of  $x_1$

$$\begin{aligned} \text{Error} &= y - \hat{y} \\ &= 250 - 340 \\ &= -90 \end{aligned}$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$



# Bagaimana Mencari Model yang Fit

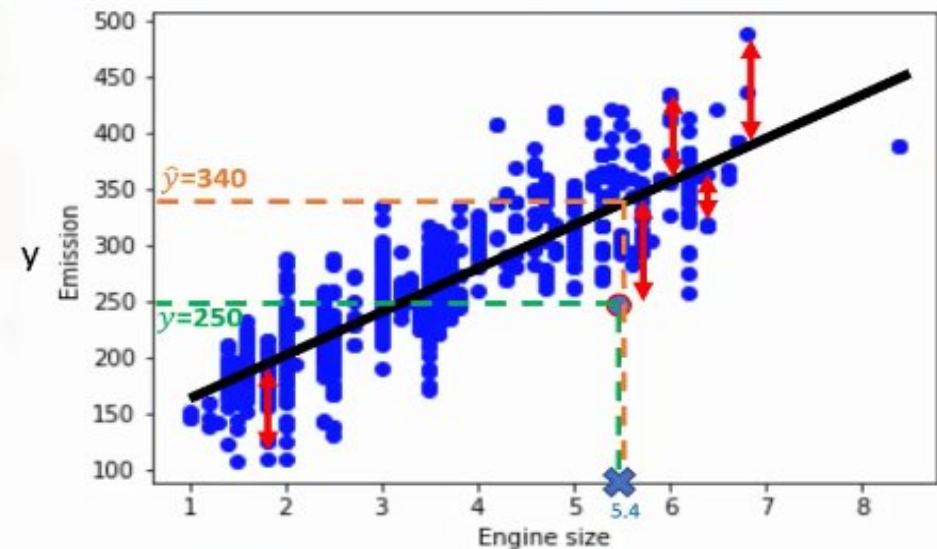
$x_1 = 2.4$  independent variable  
 $y = 250$  actual Co2 emission of  $x_1$

$$\hat{y} = \theta_0 + \theta_1 x_1$$

$\hat{y} = 340$  the predicted emission of  $x_1$

$$\begin{aligned} \text{Error} &= y - \hat{y} \\ &= 250 - 340 \\ &= -90 \end{aligned}$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$



# Estimasi Parameter-Parameter

$$\hat{y} = \theta_0 + \theta_1 x_1$$

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267

$x_1$

$y$

# Estimasi Parameter-Parameter

$$\hat{y} = \theta_0 + \theta_1 x_1$$

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267

$x_1$

$y$

# Estimasi Parameter-Parameter

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267

$$\hat{y} = \theta_0 + \theta_1 x_1$$

$$\theta_1 = \frac{\sum_{i=1}^s (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^s (x_i - \bar{x})^2}$$

$$\theta_0 = \bar{y} - \theta_1 \bar{x}$$

# Estimasi Parameter-Parameter

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267

Diagram illustrating the data structure for linear regression. The input variable  $X_1$  (Engine Size) is grouped by a blue bracket on the left. The output variable  $y$  (CO2 Emissions) is grouped by a blue bracket on the right. The data rows are indexed 0 through 8.

$$\hat{y} = \theta_0 + \theta_1 x_1$$

$$\theta_1 = \frac{\sum_{i=1}^S (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^S (x_i - \bar{x})^2}$$

$$\bar{x} = (2.0 + 2.4 + 1.5 + \dots) / 9 = 3.34$$

$$\bar{y} = (196 + 221 + 136 + \dots) / 9 = 256$$

$$\theta_0 = \bar{y} - \theta_1 \bar{x}$$



# Estimasi Parameter-Parameter

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267

$$\hat{y} = \theta_0 + \theta_1 x_1$$

$$\theta_1 = \frac{\sum_{i=1}^s (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^s (x_i - \bar{x})^2}$$

$$\bar{x} = (2.0 + 2.4 + 1.5 + \dots) / 9 = 3.34$$

$$\bar{y} = (196 + 221 + 136 + \dots) / 9 = 256$$

$$\theta_0 = \bar{y} - \theta_1 \bar{x}$$

# Estimasi Parameter-Parameter

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267

$$\hat{y} = \theta_0 + \theta_1 x_1$$

$$\theta_1 = \frac{\sum_{i=1}^s (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^s (x_i - \bar{x})^2}$$

$$\bar{x} = (2.0 + 2.4 + 1.5 + \dots) / 9 = 3.34$$

$$\bar{y} = (196 + 221 + 136 + \dots) / 9 = 256$$

$$\theta_0 = \bar{y} - \theta_1 \bar{x}$$

# Estimasi Parameter-Parameter

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267

$$\hat{y} = \theta_0 + \theta_1 x_1$$

$$\theta_1 = \frac{\sum_{i=1}^s (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^s (x_i - \bar{x})^2}$$

$$\bar{x} = (2.0 + 2.4 + 1.5 + \dots) / 9 = 3.34$$

$$\bar{y} = (196 + 221 + 136 + \dots) / 9 = 256$$

$$\theta_1 = \frac{(2.0 - 3.34)(196 - 256) + (2.4 - 3.34)(221 - 256) + \dots}{(2.0 - 3.34)^2 + (2.4 - 3.34)^2 + \dots}$$

$$\theta_0 = \bar{y} - \theta_1 \bar{x}$$

# Estimasi Parameter-Parameter

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267

$$\hat{y} = \theta_0 + \theta_1 x_1$$

$$\theta_1 = \frac{\sum_{i=1}^S (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^S (x_i - \bar{x})^2}$$

$$\bar{x} = (2.0 + 2.4 + 1.5 + \dots) / 9 = 3.34$$

$$\bar{y} = (196 + 221 + 136 + \dots) / 9 = 256$$

$$\theta_1 = \frac{(2.0 - 3.34)(196 - 256) + (2.4 - 3.34)(221 - 256) + \dots}{(2.0 - 3.34)^2 + (2.4 - 3.34)^2 + \dots}$$

$$\theta_0 = \bar{y} - \theta_1 \bar{x}$$

# Estimasi Parameter-Parameter

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267

$$\hat{y} = \theta_0 + \theta_1 x_1$$

$$\theta_1 = \frac{\sum_{i=1}^s (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^s (x_i - \bar{x})^2}$$

$$\bar{x} = (2.0 + 2.4 + 1.5 + \dots) / 9 = 3.34$$

$$\bar{y} = (196 + 221 + 136 + \dots) / 9 = 256$$

$$\theta_1 = \frac{(2.0 - 3.34)(196 - 256) + (2.4 - 3.34)(221 - 256) + \dots}{(2.0 - 3.34)^2 + (2.4 - 3.34)^2 + \dots}$$

$$\theta_1 = 39$$

$$\theta_0 = \bar{y} - \theta_1 \bar{x}$$

# Estimasi Parameter-Parameter

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267

$$\hat{y} = \theta_0 + \theta_1 x_1$$

$$\theta_1 = \frac{\sum_{i=1}^s (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^s (x_i - \bar{x})^2}$$

$$\bar{x} = (2.0 + 2.4 + 1.5 + \dots) / 9 = 3.34$$

$$\bar{y} = (196 + 221 + 136 + \dots) / 9 = 256$$

$$\theta_1 = \frac{(2.0 - 3.34)(196 - 256) + (2.4 - 3.34)(221 - 256) + \dots}{(2.0 - 3.34)^2 + (2.4 - 3.34)^2 + \dots}$$

$$\theta_1 = 39$$

$$\theta_0 = \bar{y} - \theta_1 \bar{x}$$

$$\theta_0 = 256 - 39 * 3.34$$

$$\theta_0 = 125.74$$



# Estimasi Parameter-Parameter

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267

$$\hat{y} = \theta_0 + \theta_1 x_1$$

$$\theta_1 = \frac{\sum_{i=1}^s (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^s (x_i - \bar{x})^2}$$

$$\bar{x} = (2.0 + 2.4 + 1.5 + \dots) / 9 = 3.34$$

$$\bar{y} = (196 + 221 + 136 + \dots) / 9 = 256$$

$$\theta_1 = \frac{(2.0 - 3.34)(196 - 256) + (2.4 - 3.34)(221 - 256) + \dots}{(2.0 - 3.34)^2 + (2.4 - 3.34)^2 + \dots}$$

$$\theta_1 = 39$$

$$\theta_0 = \bar{y} - \theta_1 \bar{x}$$

$$\theta_0 = 256 - 39 * 3.34$$

$$\theta_0 = 125.74$$

$$\hat{y} = 125.74 + 39x_1$$

# Prediksi dengan Garis Model

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?



# Prediksi dengan Garis Model

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?

$$\hat{y} = \theta_0 + \theta_1 x_1$$

$$Co2Emission = \theta_0 + \theta_1 EngineSize$$

$$Co2Emission = 125 + 39 EngineSize$$

# Prediksi dengan Garis Model

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?

$$\hat{y} = \theta_0 + \theta_1 x_1$$

$$Co2Emission = \theta_0 + \theta_1 EngineSize$$

$$Co2Emission = 125 + 39 EngineSize$$

$$Co2Emission = 125 + 39 \times 2.4$$

$$Co2Emission = 218.6$$

# Kelebihan Regresi Linier

- Ringan
- Tidak perlu tuning parameter
- Mudah dipahami dan diinterpretasikan

# Regresi Linier Variabel Jamak

Efektivitas variabel-variabel bebas terhadap prediksi

- Apakah kegelisahan, kehadiran dosen, dan jenis kelamin mempunyai efek pada kinerja ujian mahasiswa?

Prediksi dampak perubahan

- Seberapa besar kenaikan/penurunan tekanan darah terhadap kenaikan/penurunan BMI dari pasien?

# Prediksi Nilai Kontinu pada Regresi Linier Variabel Jamak

X: Independent variable      Y: Dependent variable

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?

$$\text{Co2 Em} = \theta_0 + \theta_1 \text{Engine size} + \theta_2 \text{Cylinders} + \dots$$

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$$

$$\hat{y} = \theta^T X$$

$$\theta^T = [\theta_0, \theta_1, \theta_2, \dots] \quad X = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \dots \end{bmatrix}$$

# MSE Untuk Menunjukkan Error Pada Model

$$\hat{y} = \theta^T X$$

$$\hat{y}_i = 140$$

the predicted emission of  $x_i$

$$y_i = 196$$

actual value of  $x_i$

$$y_i - \hat{y}_i = 196 - 140 = 56 \quad \text{residual error}$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267

# Estimasi Parameter Regresi Linier Variabel Jamak

Cara-cara mengestimasi parameter  $\theta$

## Least Squares

- Operasi aljabar linier
- Perlu waktu yang lama untuk dataset yang besar (lebih dari 10000 baris)

## Algoritma optimisasi

- Gradient Descent
- Metode yang sesuai apabila dataset sangat besar

# Prediksi Menggunakan Regresi Linier Variabel Jamak

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?

$$\hat{y} = \theta^T X$$

$$\theta^T = [125, 6.2, 14, \dots]$$

$$\hat{y} = 125 + 6.2x_1 + 14x_2 + \dots$$

$$Co2Em = 125 + 6.2EngSize + 14 Cylinders + \dots$$

$$Co2Em = 125 + 6.2 \times 2.4 + 14 \times 4 + \dots$$

$$Co2Em = 214.1$$





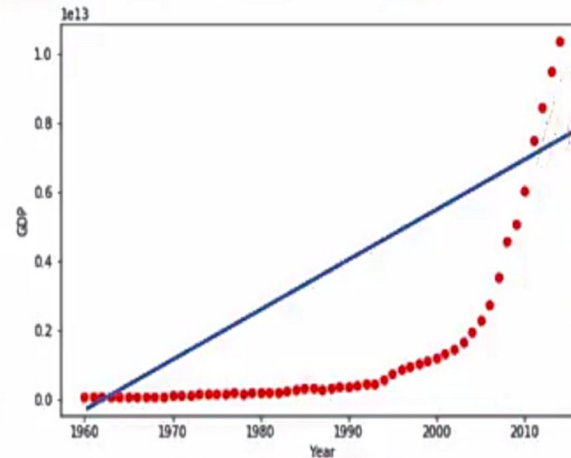
# Regresi Non Linear

# Pendahuluan

- Regresi non linier adalah suatu metode untuk mendapatkan model non linier yang menyatakan hubungan variabel dependen dan variabel independen
- Regresi nonlinier dapat mengestimasi model hubungan variabel dependen dan independen dalam bentuk non linier dengan keakuratan yang lebih baik daripada regresi linier, karena dalam mengestimasi model dipakai iterasi algoritma.

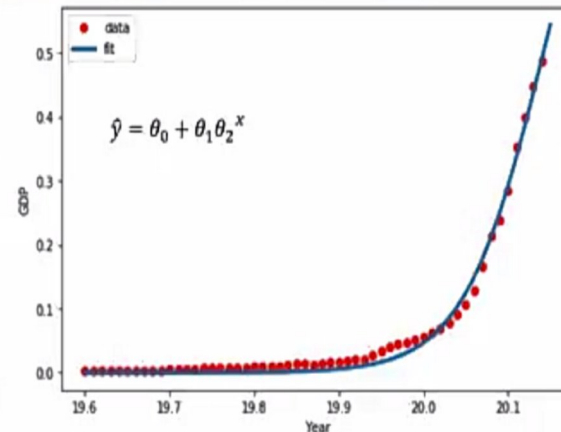
# Mengapa Regresi Non-Linier Diperlukan?

	Year	Value
0	1960	5.918412e+10
1	1961	4.955705e+10
2	1962	4.668518e+10
3	1963	5.009730e+10
4	1964	5.906225e+10
5	1965	6.970915e+10
6	1966	7.587943e+10
7	1967	7.205703e+10
8	1968	6.999350e+10
9	1969	7.871882e+10
...	...	...



- Tidak setiap data menunjukkan hubungan linier
- Sehingga error akan besar ketika dipaksakan menggunakan model linier

	Year	Value
0	1960	5.918412e+10
1	1961	4.955705e+10
2	1962	4.668518e+10
3	1963	5.009730e+10
4	1964	5.906225e+10
5	1965	6.970915e+10
6	1966	7.587943e+10
7	1967	7.205703e+10
8	1968	6.999350e+10
9	1969	7.871882e+10
...	...	...



- Penggunaan model non-linier tampak mempunyai error yang lebih kecil

# Regresi Polinomial

- Beberapa data yang berbentuk kurva dapat dimodelkan dengan regresi linier
- Contoh:

$$\hat{y} = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$$

- Model regresi polinomial dapat ditransformasikan menjadi model regresi linier.

$$x_1 = x$$

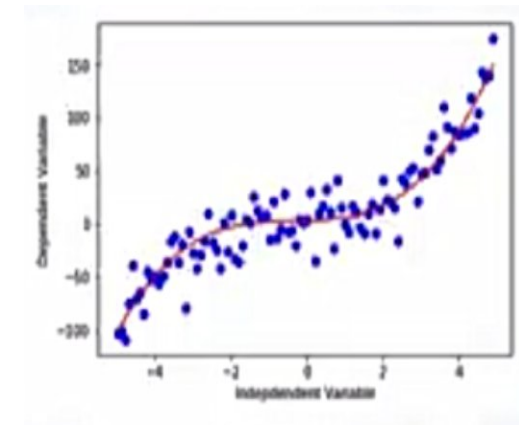
$$x_2 = x^2$$

$$x_3 = x^3$$

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$$

Regresi linier variabel jamak

-> dapat diselesaikan dengan least squares



# Regresi Non-Linier

- Memodelkan hubungan tidak linier antara variabel tak bebas dengan himpunan variabel bebas
- $\hat{y}$  berupa fungsi non-linier dari parameter  $\theta$  dan fitur  $x$ .

$$\hat{y} = \theta_0 + \theta_2^2 x$$

$$\hat{y} = \theta_0 + \theta_1 \theta_2^x$$

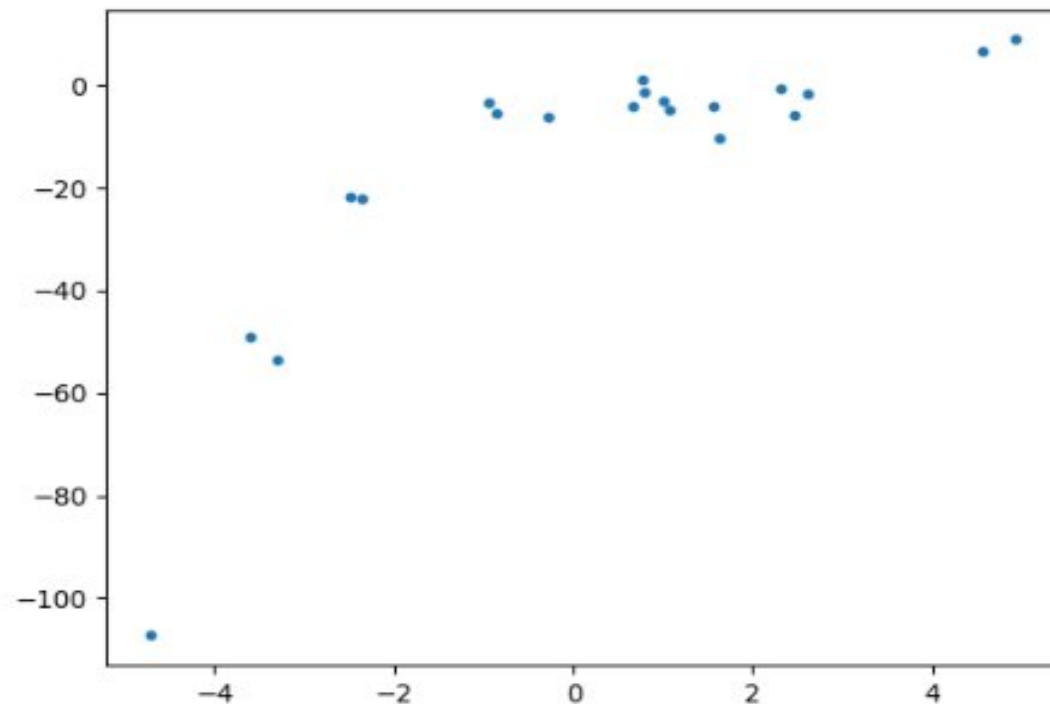
$$\hat{y} = \log(\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3)$$

$$\hat{y} = \frac{\theta_0}{1 + \theta_1^{(x-\theta_2)}}$$



# Mengapa Regresi Non-Linear

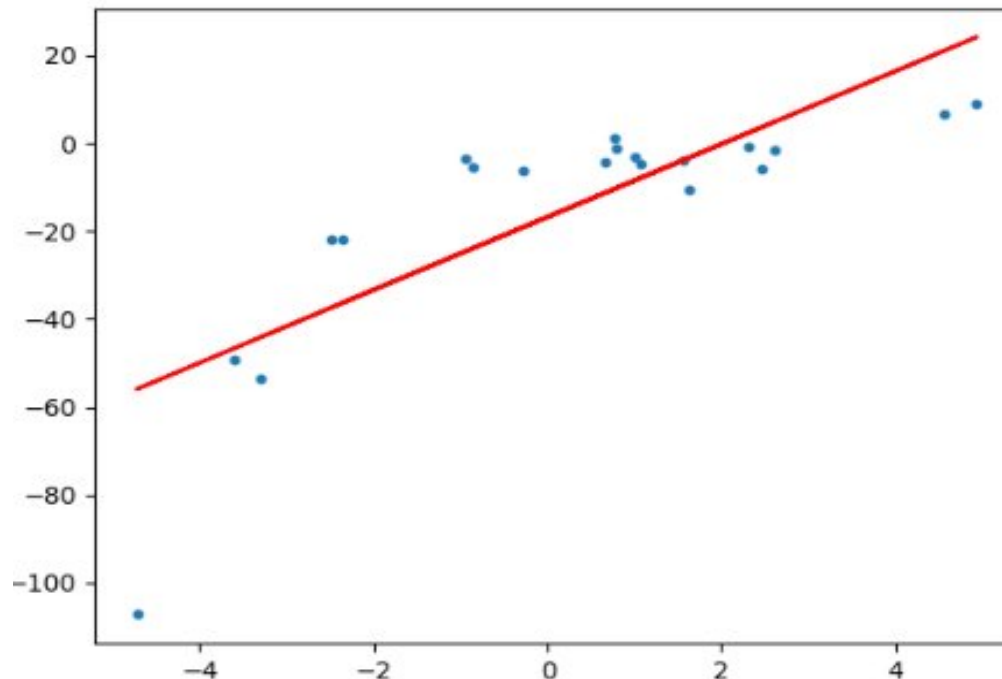
- Misalkan diberikan data seperti ini:



- Gunakan regresi linear atau non-linear?

# Mengapa Regresi Non-Linear (2)

- Digunakan model linear/ derajat 1:



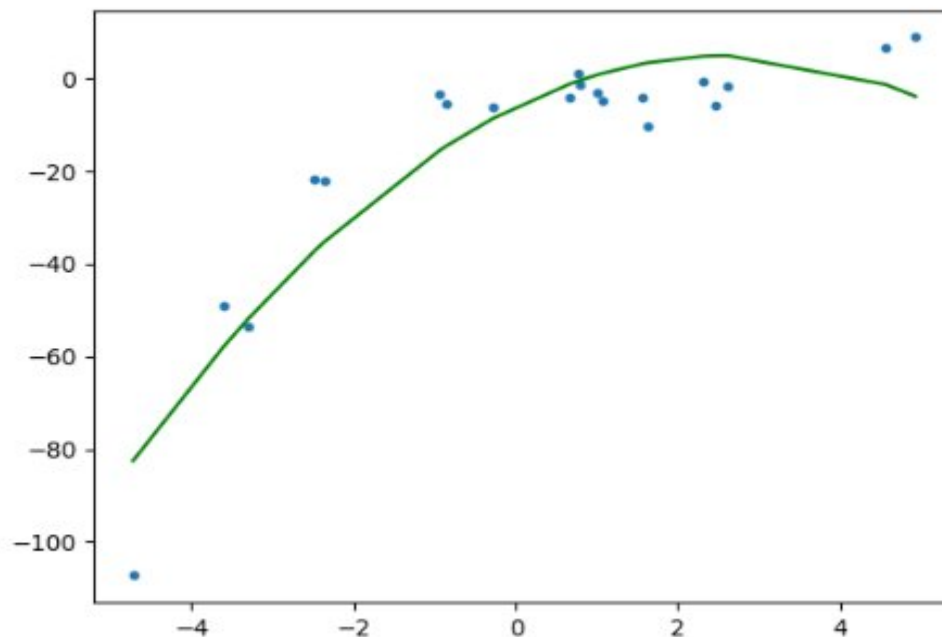
RMSE of linear  
regression is  
15.908242501429998.

R2 score of linear  
regression is  
**0.6386750054827146**

- Banyak data yang “jauh” dari garis model.
- Kondisi: under-fitting

# Mengapa Regresi Non-Linear (3)

- Jika menggunakan model non-linear (kuadratik) / derajat 2:



RMSE of polynomial  
regression is  
10.120437473614711.

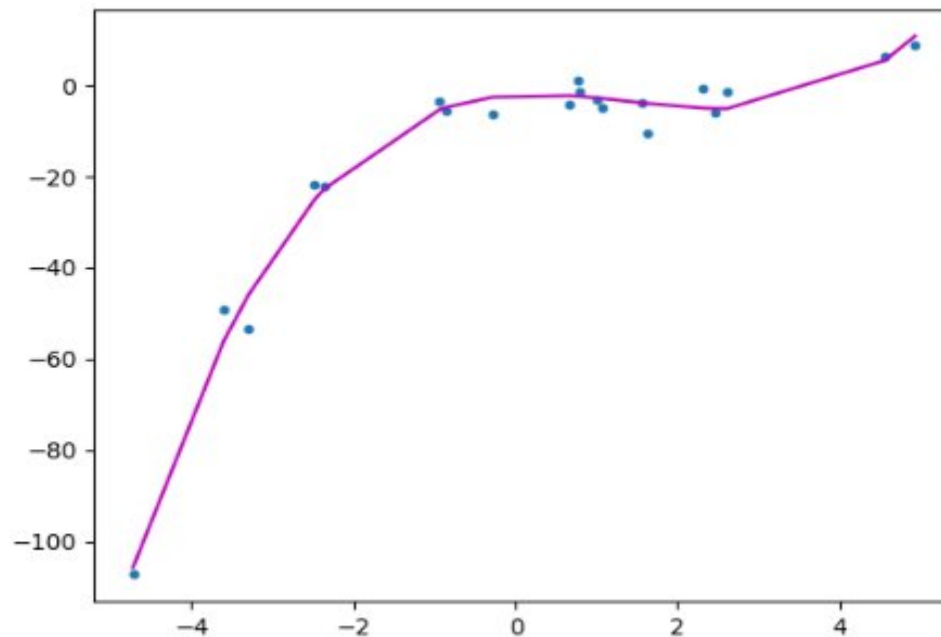
R2 of polynomial  
regression is  
**0.8537647164420812.**

- Garis model terlihat sudah mendekati persebaran data.



# Mengapa Regresi Non-Linear (4)

- Jika menggunakan model non-linear (kubik) / derajat 3:



RMSE is

3.449895507408725

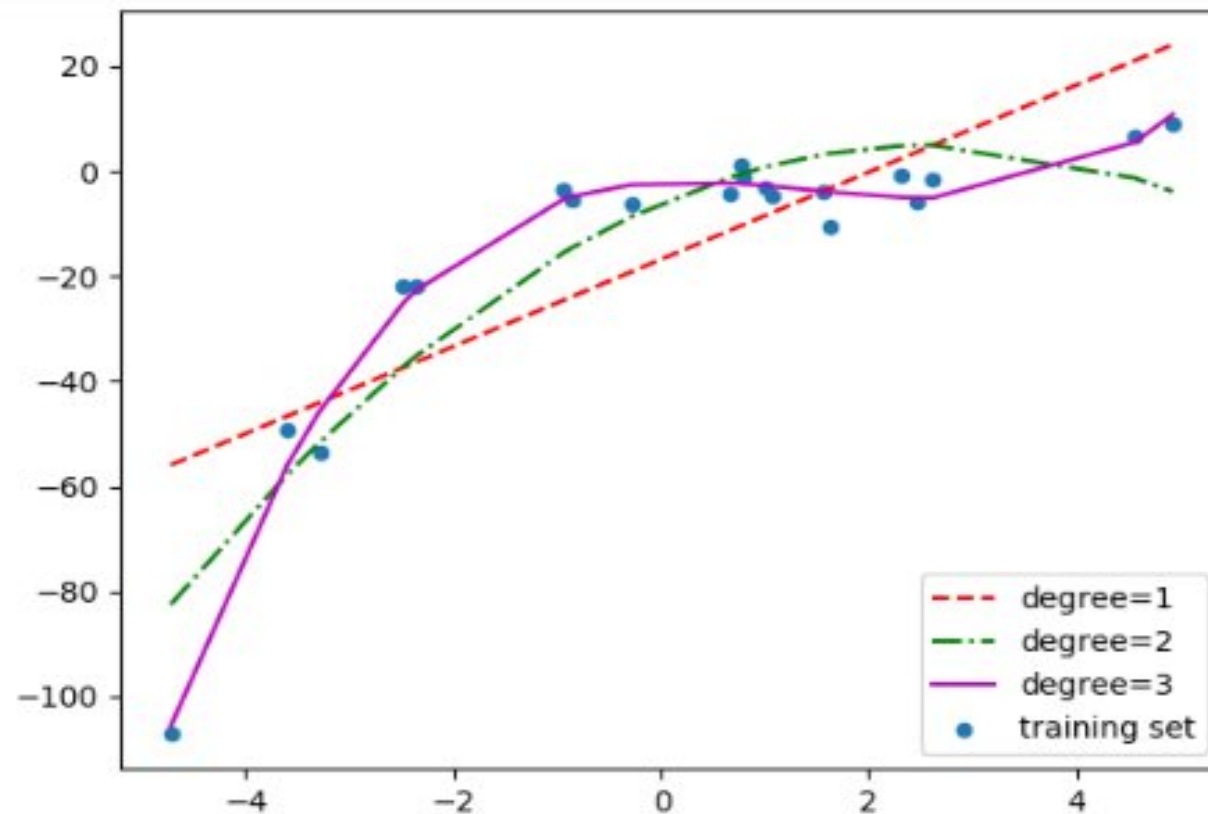
R2 score is

**0.9830071790386679**

- Garis model terlihat semakin mendekati persebaran data.
- Kondisi: appropriate-fitting / correct-fit

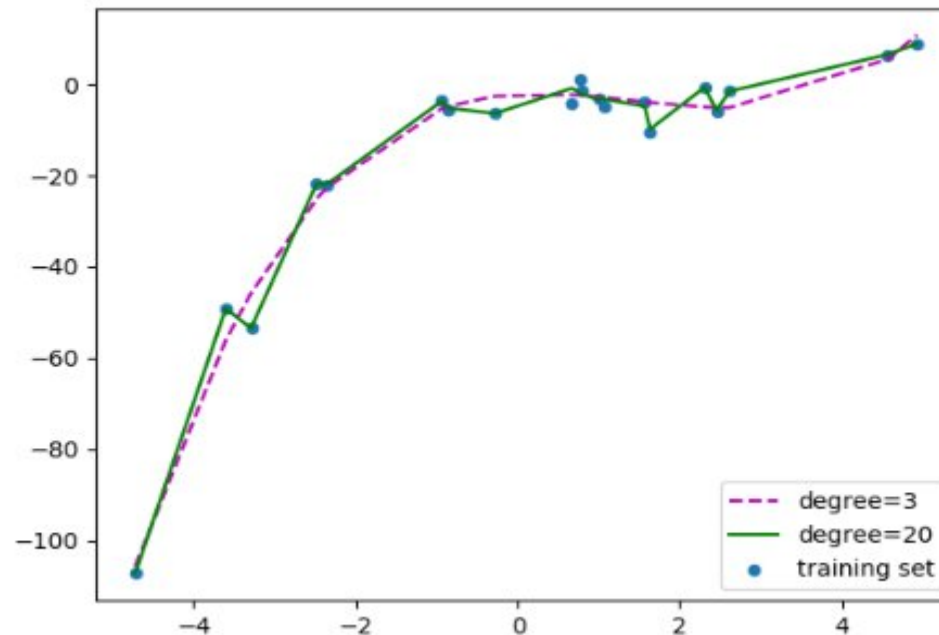
# Perbandingan 3 Model

- Berikut adalah perbandingan model linear, kuadratik dan kubik:



# Bagaimana Dengan Derajat Yang Lain

- Misalkan dimodelkan dengan derajat “20”:

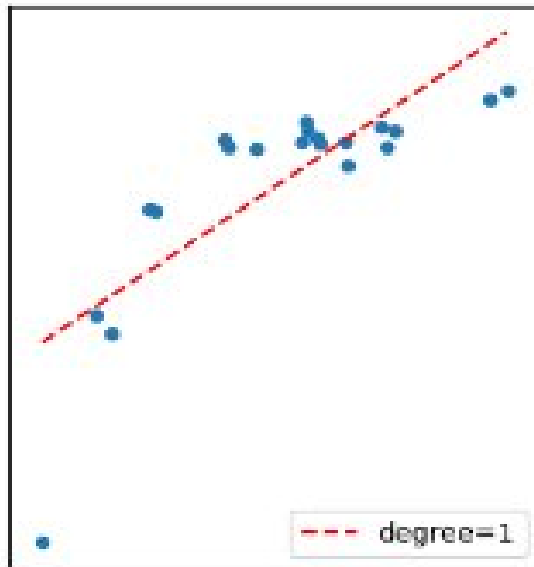


- Garis model juga melewati “noise” data.
- Kondisi: over-fitting

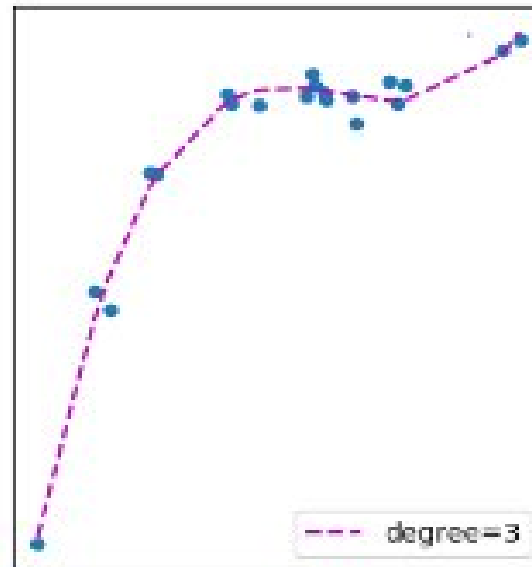
# Bias vs Varians

- Bias merujuk pada kesalahan, karena asumsi model yang sederhana dalam menyesuaikan data.
- Bias yang tinggi artinya bahwa, model tidak dapat menangkap pola di dalam data dan ini mengakibatkan kondisi under-fitting.
- Varians mengacu pada kesalahan, karena model yang kompleks mencoba menyesuaikan data.
- Varians tinggi artinya model melewati sebagian besar titik pada data dan menghasilkan kondisi over-fitting

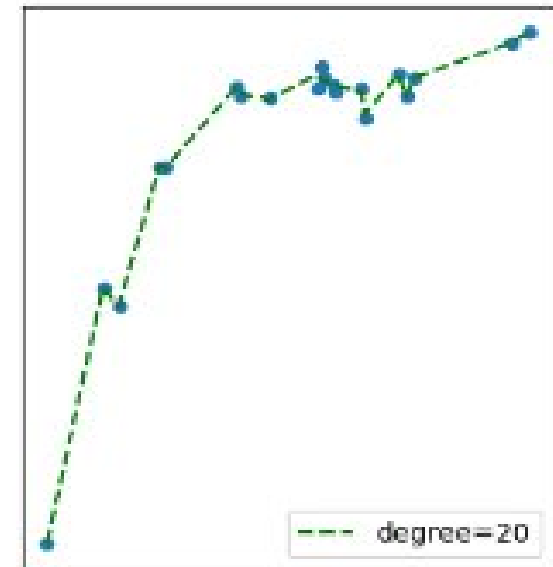
# Bias vs Varians (2)



Underfit  
High Bias  
Low Variance



Correct Fit  
Low Bias  
Low Variance



Overfit  
Low Bias  
High Variance

# Regresi Linier atau Non-Linier?

Cara untuk mengetahui apakah permasalahan cocok diselesaikan dengan regresi linier atau non linier

- Pengamatan visual atas data (visualisasi)
- Pengamatan akurasi hasil pemodelan

Cara untuk memodelkan data apabila visualisasi mengindikasikan non-linier

- Regresi polinomial
- Regresi non-linier
- Transformasi data non-linier menjadi linier



# Contoh Regresi

# Aplikasi Regresi

- Prakiraan penjualan produk
- Analisis kepuasan
- Estimasi harga
- Pendapatan pekerjaan
- dst.



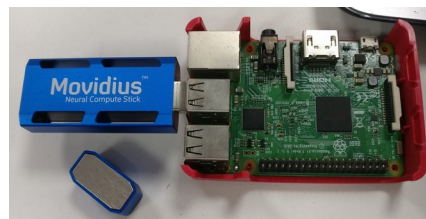


Colab Lab



Super Computer DGX-1/A100

# TERIMA KASIH



EdgeAI

