



# Regresi

---

Ledy Elsera Astrianty, S.Kom., M.Kom

# Implementasi supervised learning

## Capaian:

Mahasiswa dapat menjelaskan, menyiapkan, dan mengimplementasikan model regresi dengan Algoritma Regresi Linier Sederhana dan Regresi Lilier Berganda

# Apa itu 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.

# Aplikasi Regresi

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

# Algoritma Regresi

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

Regresi Linier Sederhana:

Memprediksi co2emission vs EngineSize dari semua mobil

- variabel bebas (x): EngineSize
- variabel tak bebas (y): co2emission

Regresi Linier Variabel Jamak:

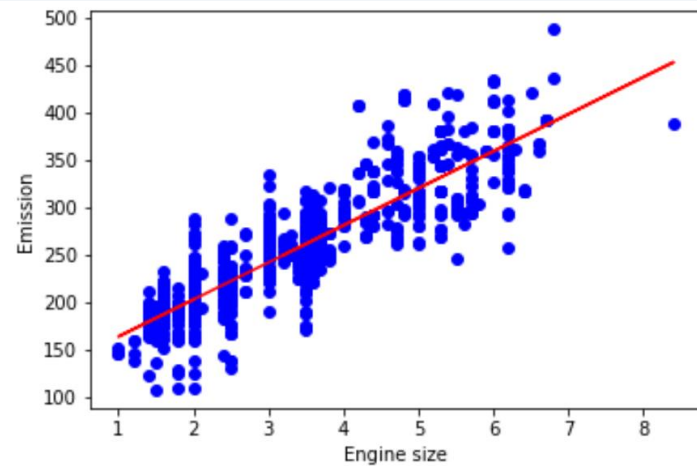
Memprediksi co2emission vs EngineSize dan Cylinders dari semua mobil

- variabel bebas (x): EngineSize, Cylinders, dst.
- variabel tak bebas (y): co2emission

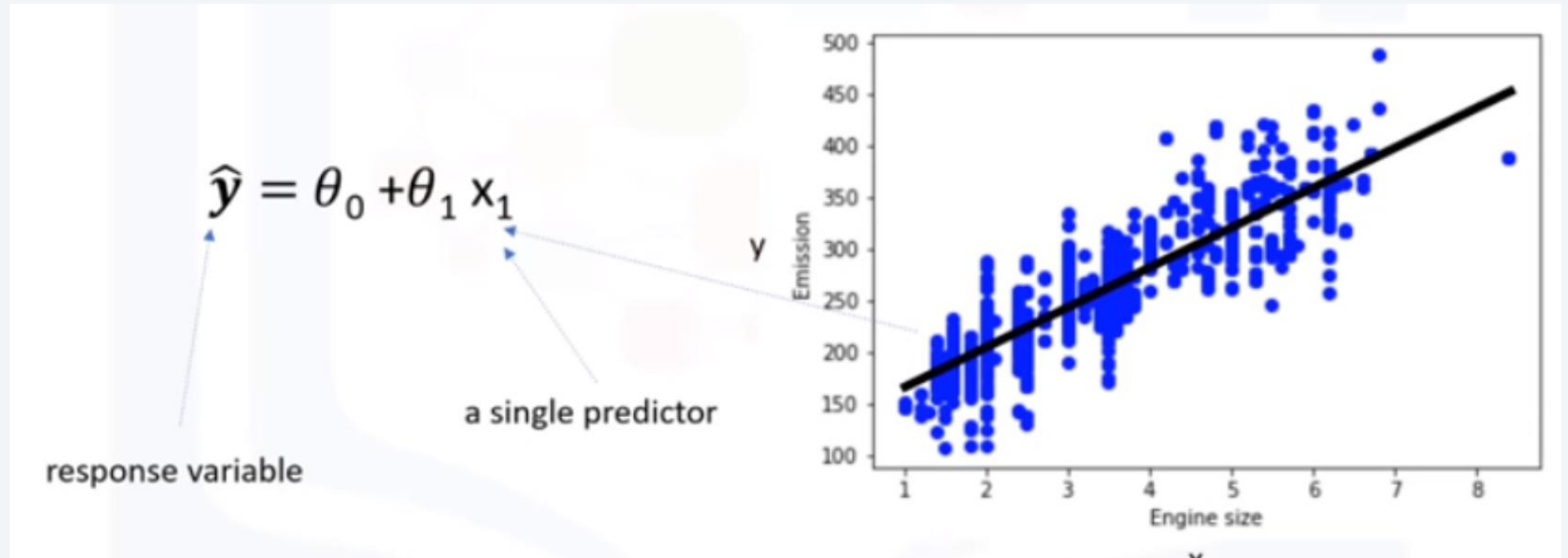
# Topologi Regresi Linier

# Cara Kerja Regresi Linier

	ENGINESIZE	CYLINDERS	FUELCONSUMPTION_COMB	CO2EMISSIONS
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	?







## Cara Kerja Regresi Linier

# Cara Mencari Parameter Model Terbaik

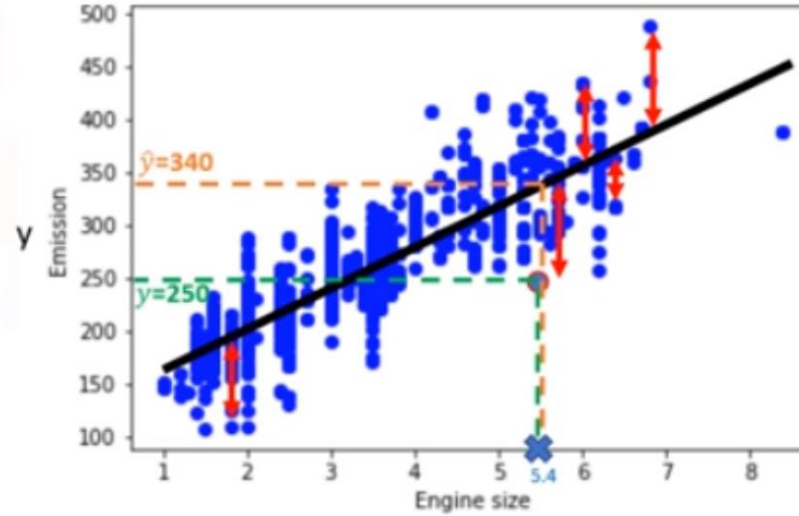
$x_1 = 5.4$  independent variable  
 $y = 250$  actual Co2 emission of  $x_1$

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

$\hat{y} = 340$  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

	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

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

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

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

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

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

$$\theta_1 = 39$$

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

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

$$\theta_0 = 125.74$$



# Prediksi dengan Model Regresi Linier

	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 (Berganda)

# Contoh Regresi Linier Berganda

## **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 Berganda

	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	?

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

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

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

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

$$\text{Theta} = (X^T X)^{-1} X^T Y$$

Figure 5: Co-efficient calculation using Normal Equation



# 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  $x_i$

$y_i - \hat{y}_i = 196 - 140 = 56$  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 Berganda

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 Berganda

	ENGINESIZE	CYLINDERS	FUELCONSUMPTION_COMB	CO2EMISSIONS
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	?

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

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

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

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

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

$$Co2Em = 214.1$$

**Thank You**

