

# Linear & Logistic Regression

Indonesia AI

Proprietary document of Indonesia AI 2023



# OBJECTIVE & OUTLINE

Proprietary document of Indonesia AI 2023



## Linear & Logistic Regression

Objektif: Memahami konsep dasar dari Machine learning dan mengenal 2 algoritma fundamental yaitu Regresi Linear dan Logistik.

Outline:

1. Apa itu Machine Learning?
2. Data Preprocessing
3. Algoritma Linear Regression
4. Algoritma Logistic Regression

# — Apa itu Machine Learning?

# AI, MACHINE LEARNING, DEEP LEARNING

Proprietary document of Indonesia AI 2023

## Expert System

Metode atau teknik AI dengan memanfaatkan **data** dalam jumlah besar (*Knowledge Base*)

## Artificial Intelligence

Konsep atau gagasan tentang mesin yang memiliki **kecerdasan** seperti manusia

## Machine Learning

Metode atau teknik membangun kecerdasan pada mesin dengan **belajar** dari **data** dalam jumlah besar

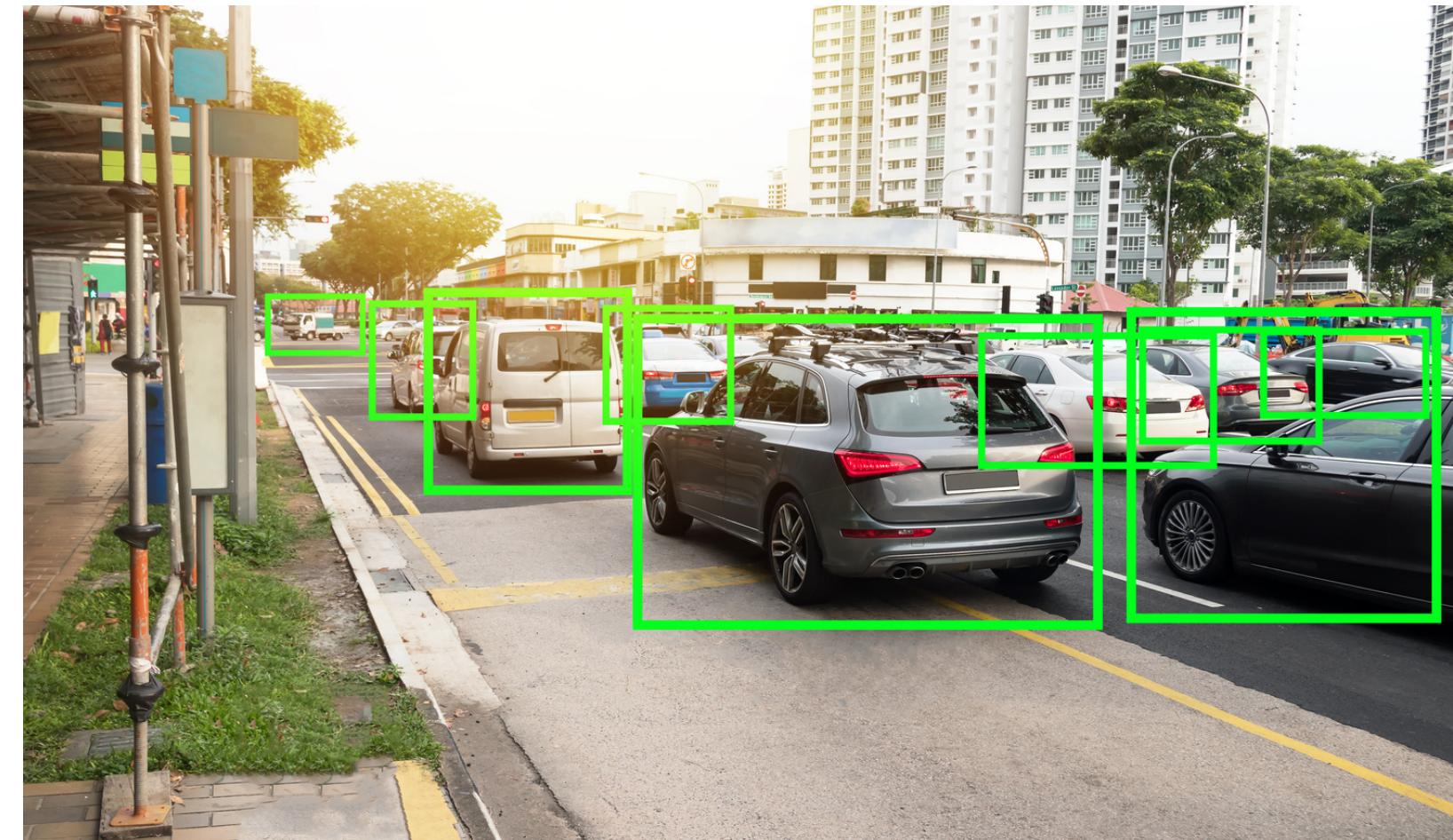
## Deep Learning

Metode atau teknik khusus untuk membangun kecerdasan pada mesin yang meniru **cara kerja otak** manusia

# DEFINISI

Proprietary document of Indonesia AI 2023

- Bidang studi yang bertujuan untuk memberikan komputer **kemampuan untuk belajar** tanpa diprogram secara eksplisit
- Machine Learning berkembang seiring dengan berkembangnya ilmu matematika dan statistika sehingga mesin dapat belajar dari data



# DEFINISI

Proprietary document of Indonesia AI 2023

## Traditional Programming

Menerima data untuk kemudian diprogram dan menghasilkan output



# DEFINISI

Proprietary document of Indonesia AI 2023

## Machine Learning

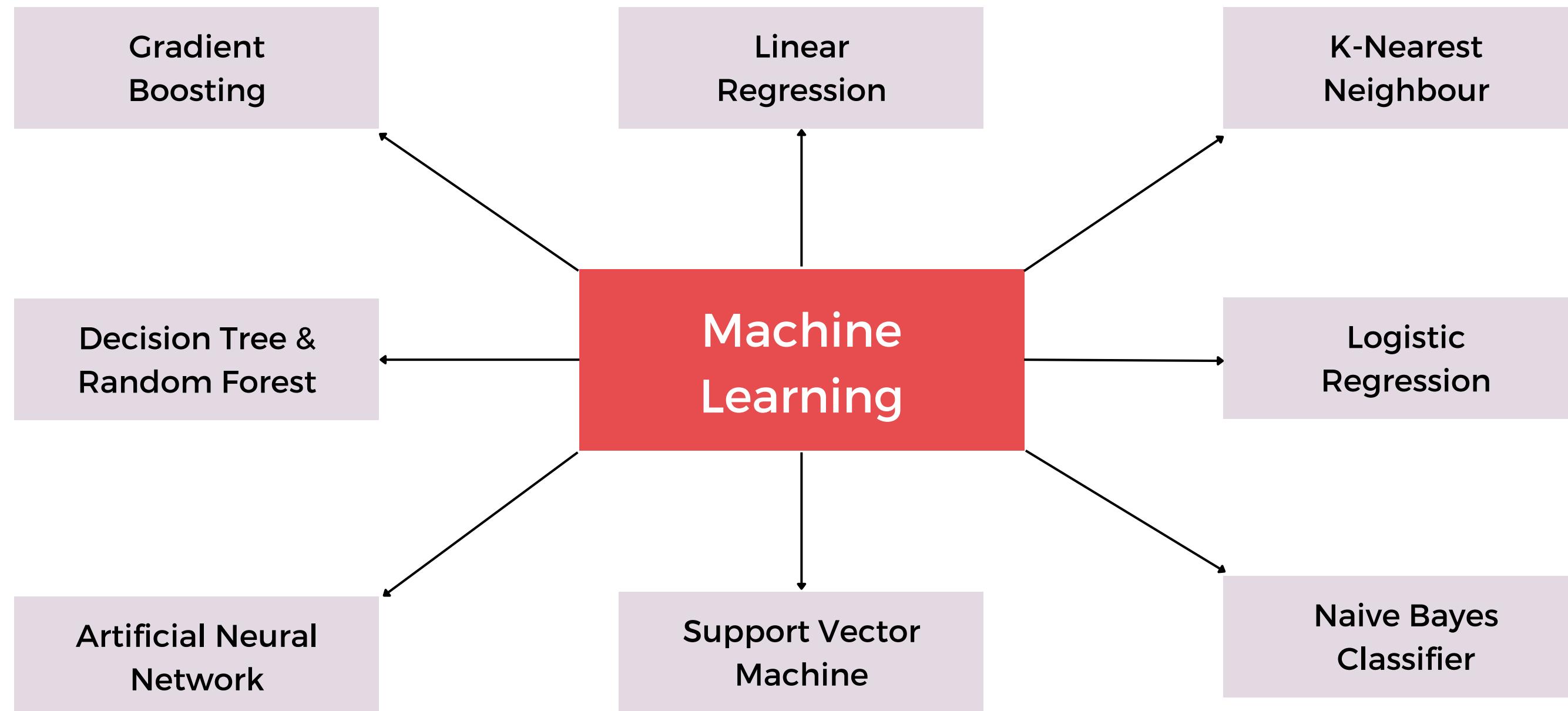
Menerima data dan output untuk menghasilkan program



**Output** pada Machine Learning juga dikenal sebagai label atau anotasi  
**Machine Learning** merujuk kepada algoritma dengan kemampuan belajar

# ALGORITMA MACHINE LEARNING

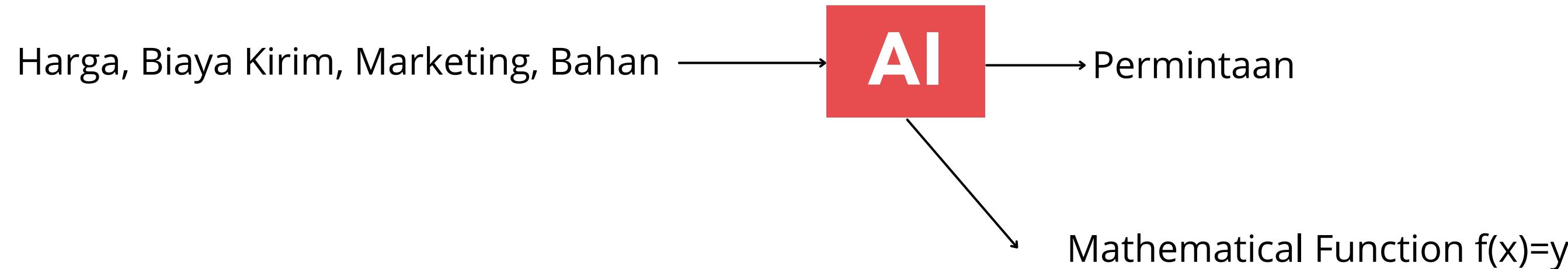
Proprietary document of Indonesia AI 2023



# CARA KERJA

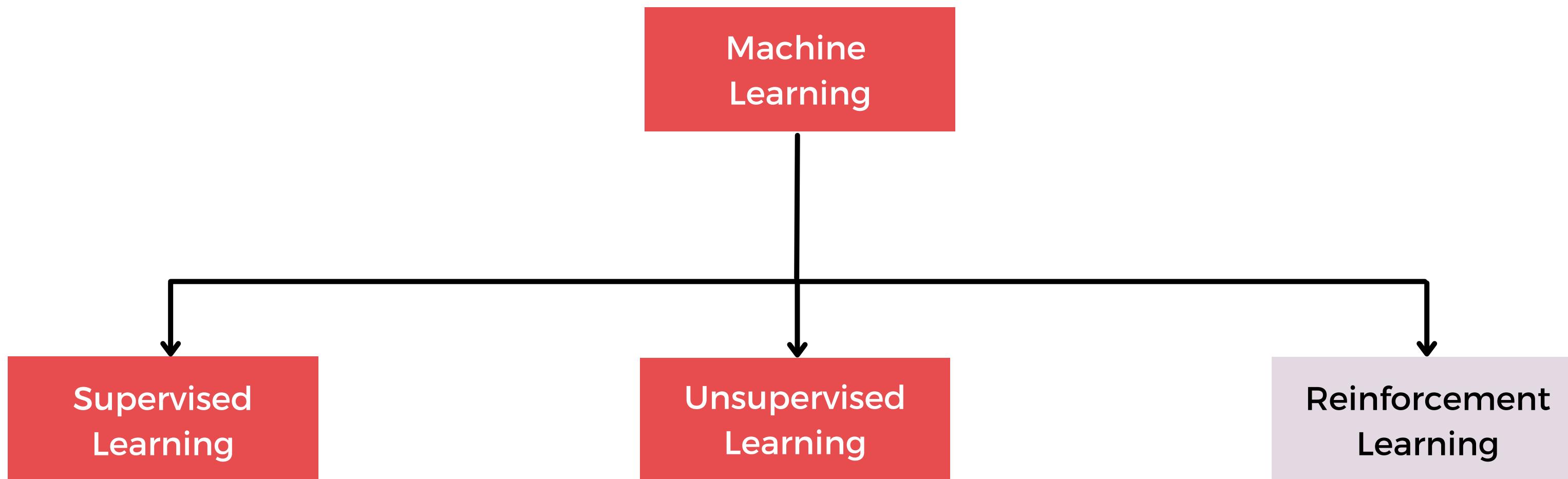
Proprietary document of Indonesia AI 2023

Salah satu cara kerja yang terkenal yaitu **Supervised Learning**. Merupakan pendekatan belajar pada machine learning yang melakukan proses pembelajaran melalui **pengawasan** atau **supervisi** dalam bentuk **labelisasi** oleh expert atau seorang ahli (misal business owner).



# CARA KERJA

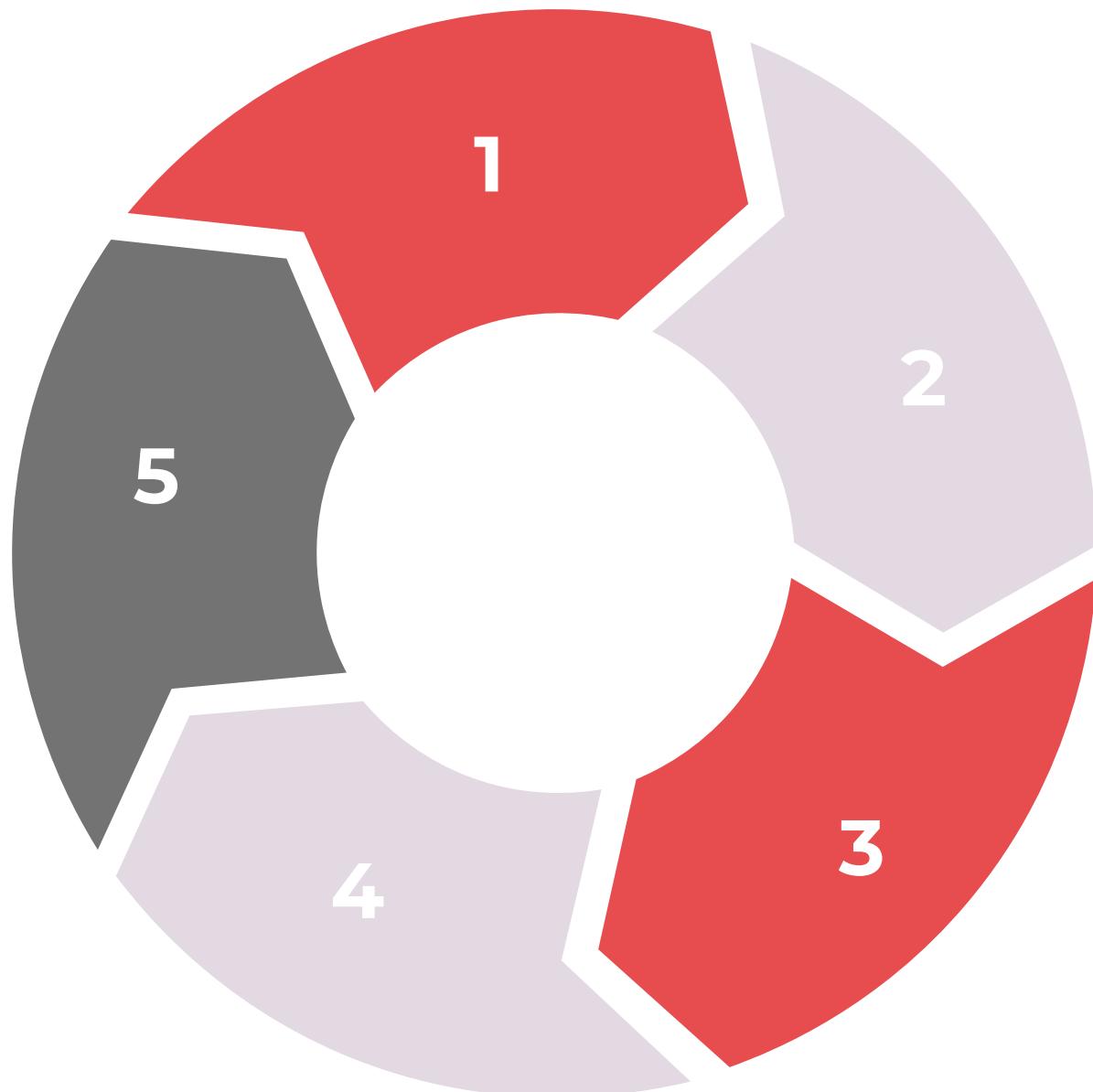
Proprietary document of Indonesia AI 2023



*3 metode belajar di atas adalah metode utama yang dipakai di industri. Namun di dunia riset, ada banyak metode lainnya seperti semi-unsupervised learning dan imitation learning.*

# MACHINE LEARNING LIFECYCLE

Proprietary document of Indonesia AI 2023



1. Data Collection
2. Data Preprocessing
3. Data Modeling
4. Model Evaluation
5. Model Deployment & Maintenance

Berlaku untuk projek AI/Machine Learning ataupun Data Science...

# Any question guys ~

# — Data Preprocessing

# DEFINISI

Proprietary document of Indonesia AI 2023

**Data preprocessing** adalah satu aktivitas pada siklus Machine Learning atau Data Science dimana data akan diolah dan dipersiapkan sebelum kemudian dipahami dengan mudah oleh algoritma Machine Learning pada tahapan berikutnya.



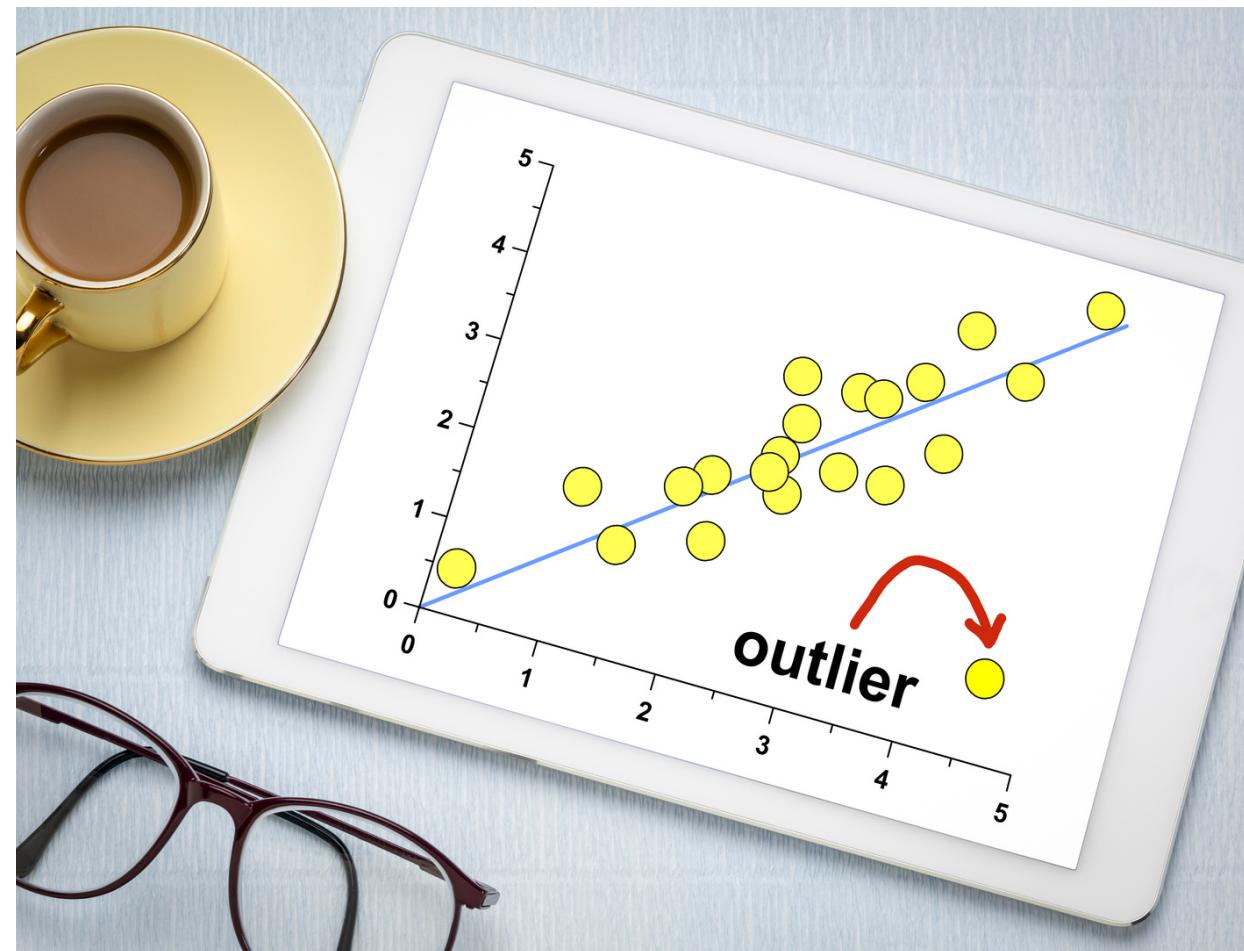


# JENIS DATA

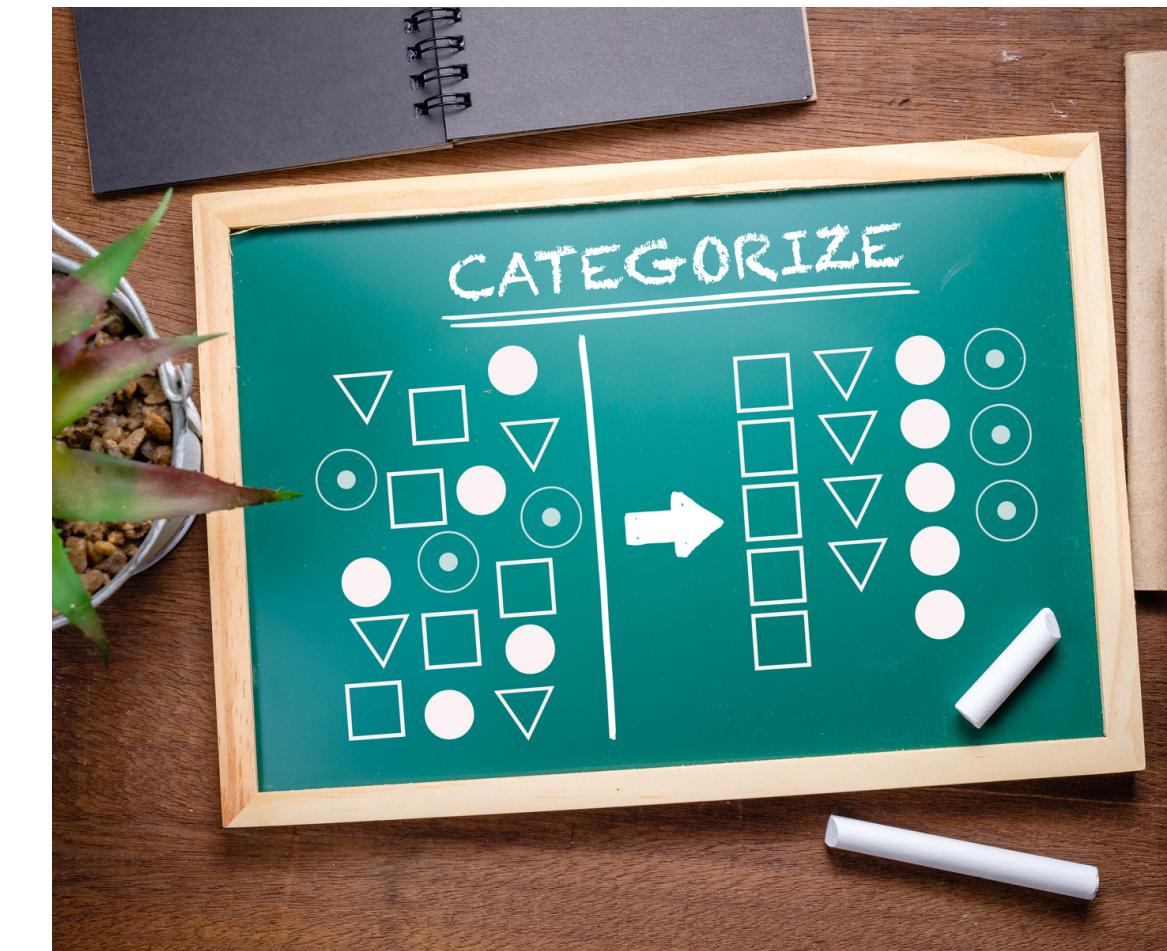
Proprietary document of Indonesia AI 2023



Incomplete/Missing Data



Noise/Outlier Data



Categorical/Continuous Data



## UNGKAPAN TERKENAL:

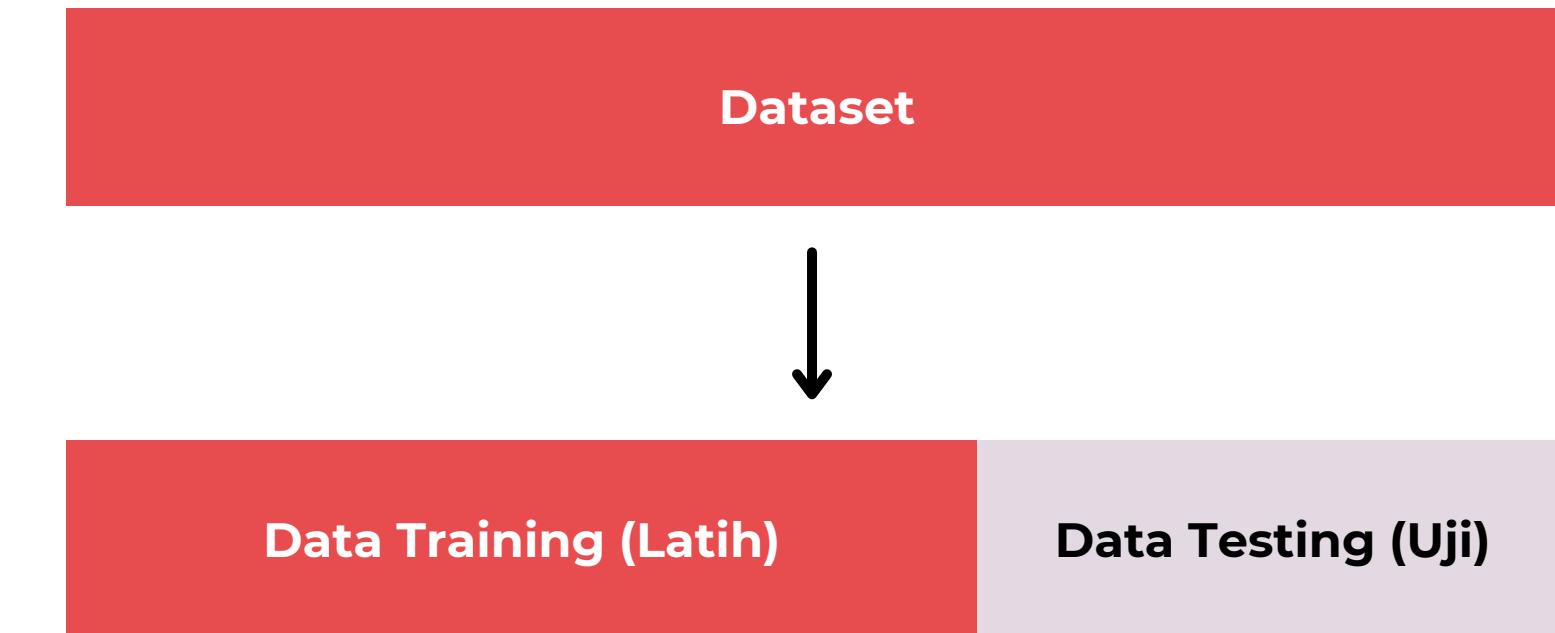
*"Garbage in, Garbage out!"*

# DATA SPLIT

Proprietary document of Indonesia AI 2023

Ukuran (m <sup>2</sup> )	Jumlah Kamar Tidur	Harga (Juta)
30	1	230
45	1	219
66	2	500
72	3	510
73	3	690

↓  
Label



Ratio Pembagian  
80 Training : 30 Testing  
70 Training : 40 Testing

# WELCOME TO MACHINE LEARNING!

Proprietary document of Indonesia AI 2023



Sudah siap masuk tahapan dimana kita akan  
mengajarkan mesin agar bisa **belajar** dan  
menjadi **pintar**?

# Algoritma Linear Regression

# PENGENALAN

Proprietary document of Indonesia AI 2023

**Linear Regression** atau regresi linear adalah algoritma Machine Learning yang popular digunakan untuk membuat model sederhana yang mampu **memprediksi nilai-nilai baru**.

Linear Regression bekerja pada **persoalan regresi** yang bersifat numerik berdasarkan satu atau lebih variabel/**fitur** penentu (independen).



# PERSOALAN REGRESI

Proprietary document of Indonesia AI 2023



Harga Rumah



Gaji Karyawan



Kecepatan Mobil

# PERSOALAN REGRESI

Proprietary document of Indonesia AI 2023



**Dependent Variable:** Harga rumah?

**Independent Variable:** Jumlah kamar,  
jumlah lantai, luas tanah, tahun  
didirikan dan lain sebagainya

# PERSOALAN REGRESI

Proprietary document of Indonesia AI 2023

## Fun Fact

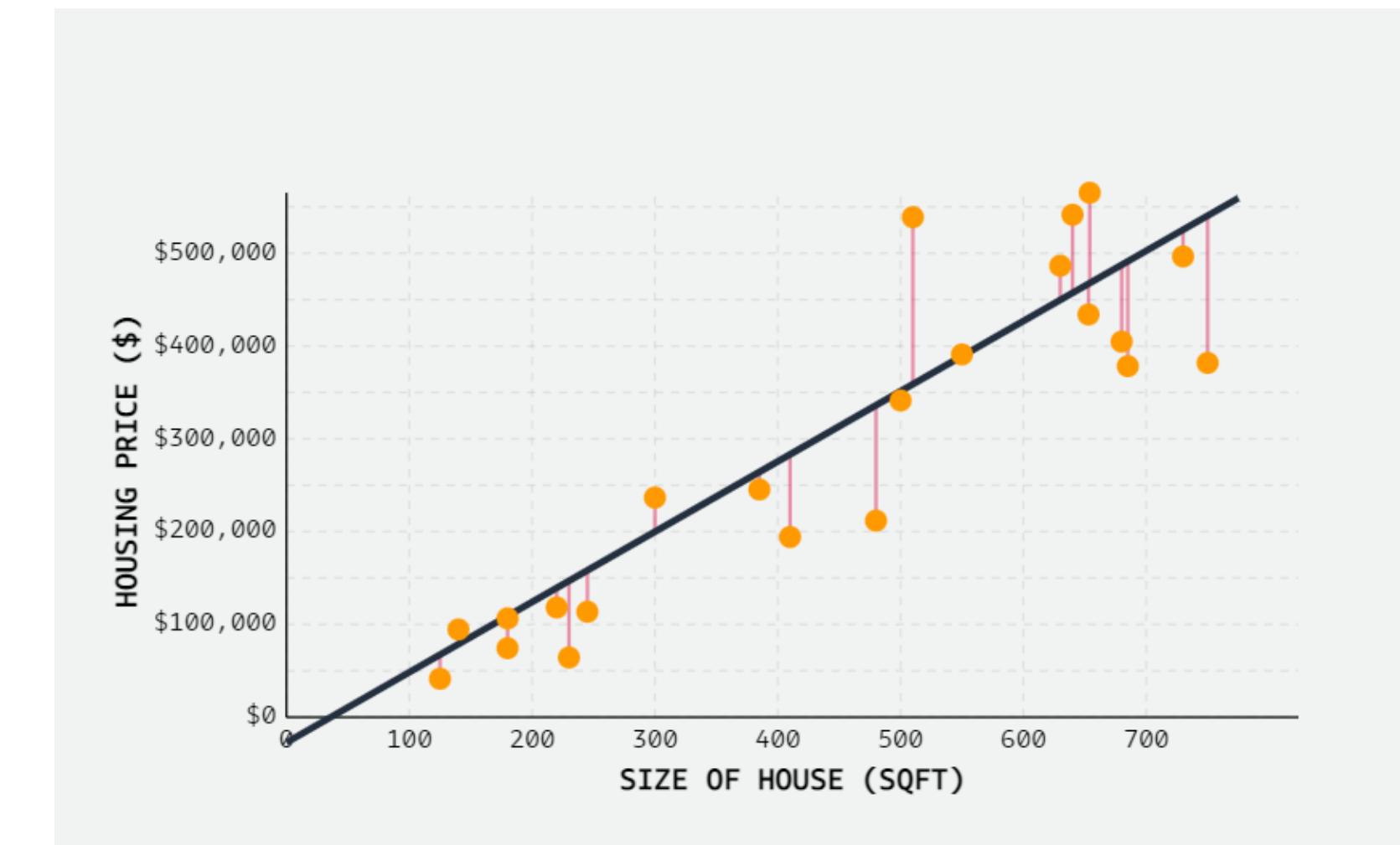
Menyelesaikan persoalan regresi menggunakan algoritma Linear Regression dianggap sebagai **Hello World** di bidang Machine Learning!

# LINEAR REGRESI

Proprietary document of Indonesia AI 2023

**Linear Regression** merupakan algoritma Machine Learning kategori Supervised Learning yang belajar dengan memodelkan variable dependen (y) sebagai sebuah fungsi dari beberapa variable independen (x).

Pemodelan terjadi dengan menemukan garis/line (atau permukaan/surface) yang "fits" terhadap data.

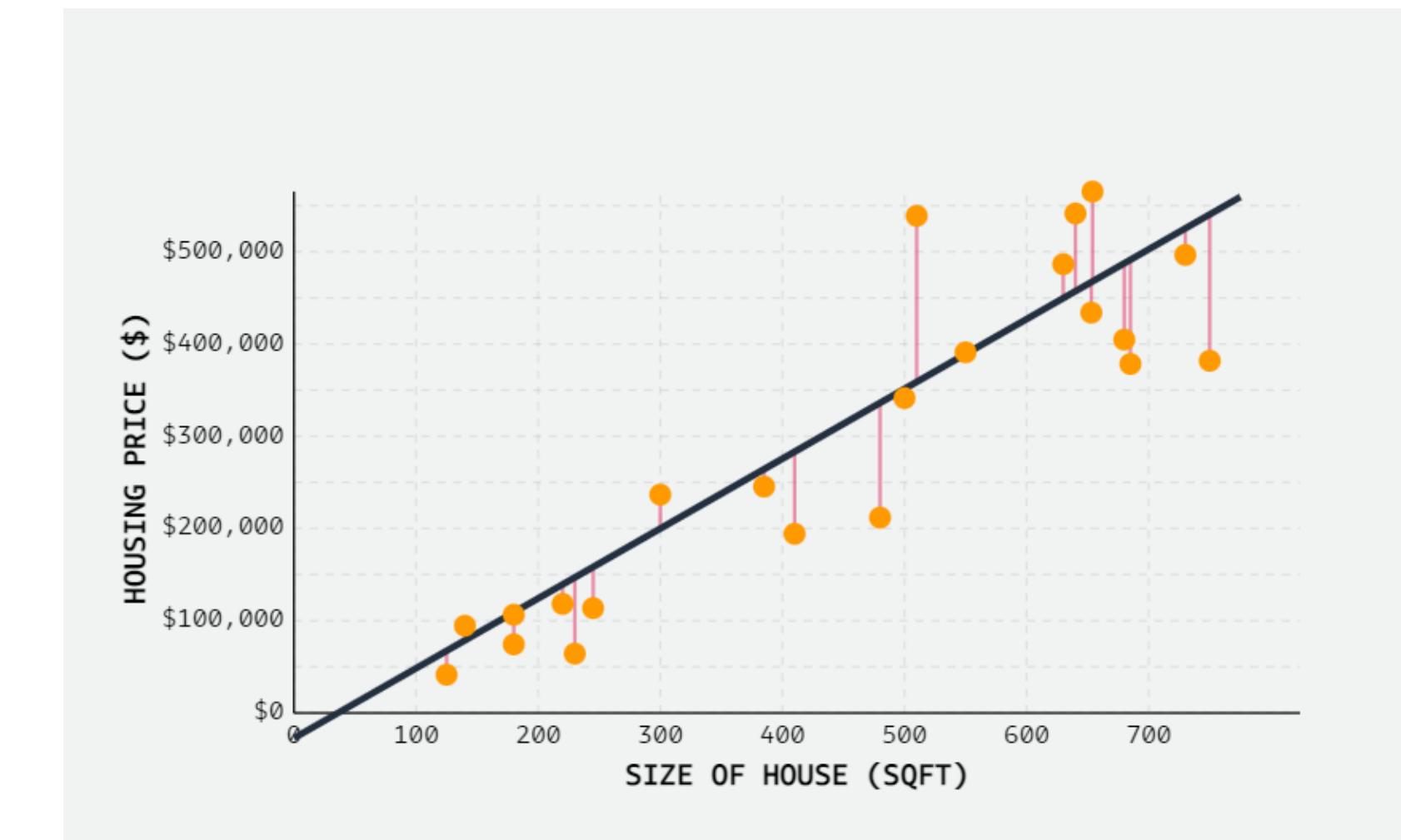


# LINEAR REGRESI

Proprietary document of Indonesia AI 2023

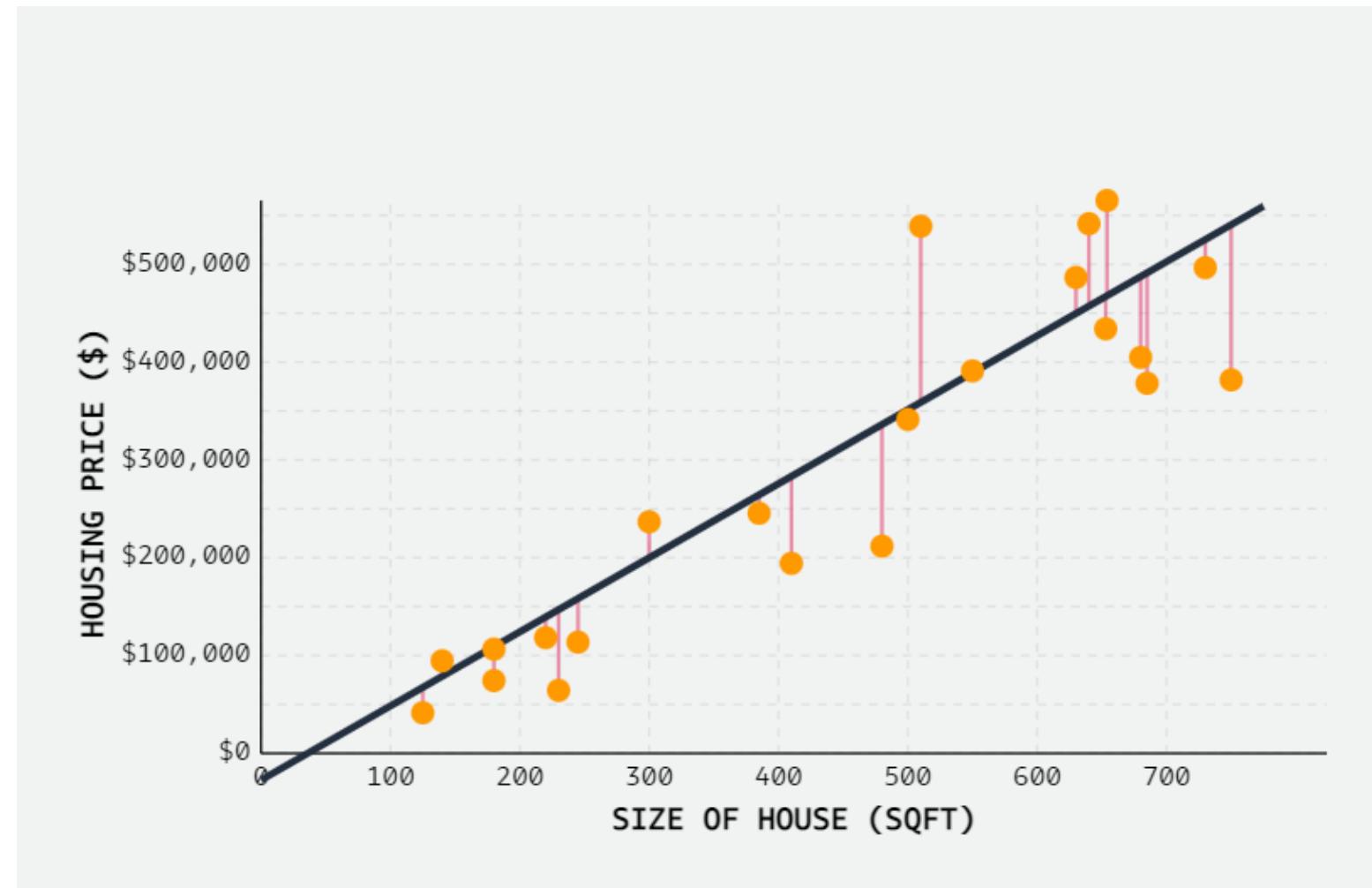
Dalam **Linear Regression**, Nilai y merupakan nilai numerik yang diprediksi berdasarkan satu atau lebih nilai x yang merupakan fitur.

Contohnya, memprediksi harga rumah (x) berdasarkan jumlah kamar di rumah tersebut atau gaji karyawan berdasarkan pengalaman kerja (x1) dan jabatan (x2).

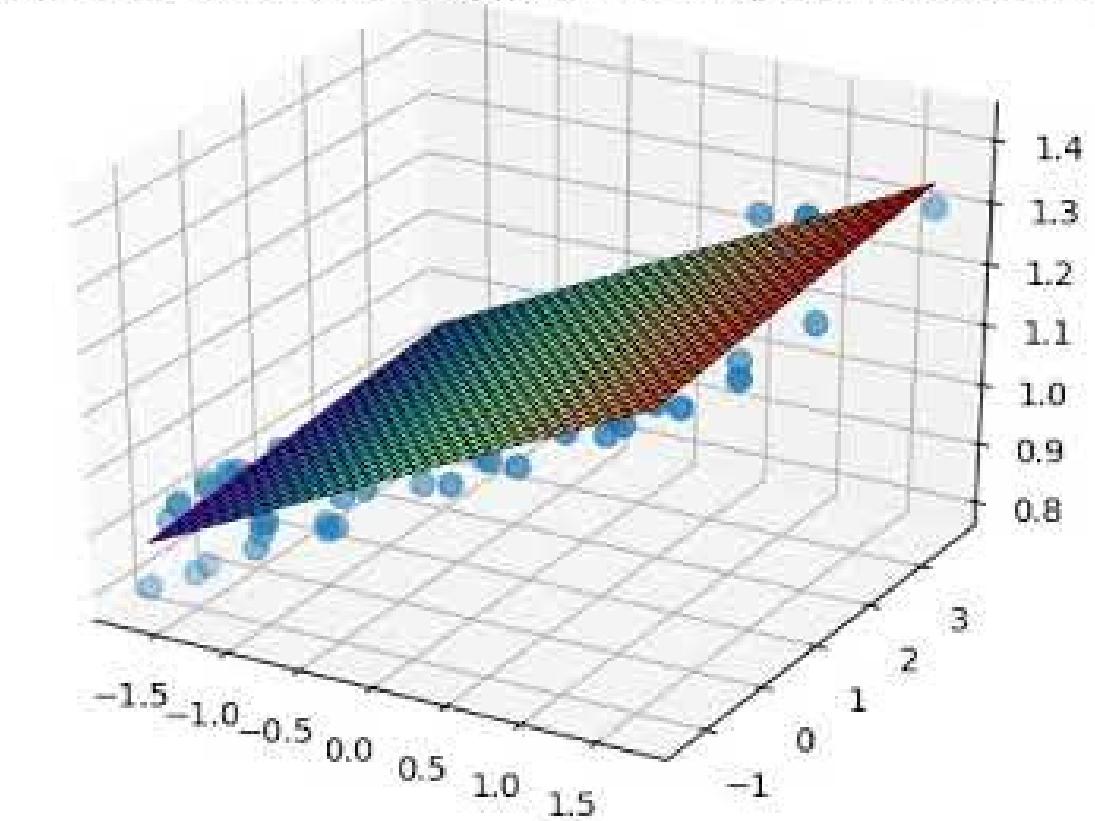


# LINEAR REGRESI

Proprietary document of Indonesia AI 2023

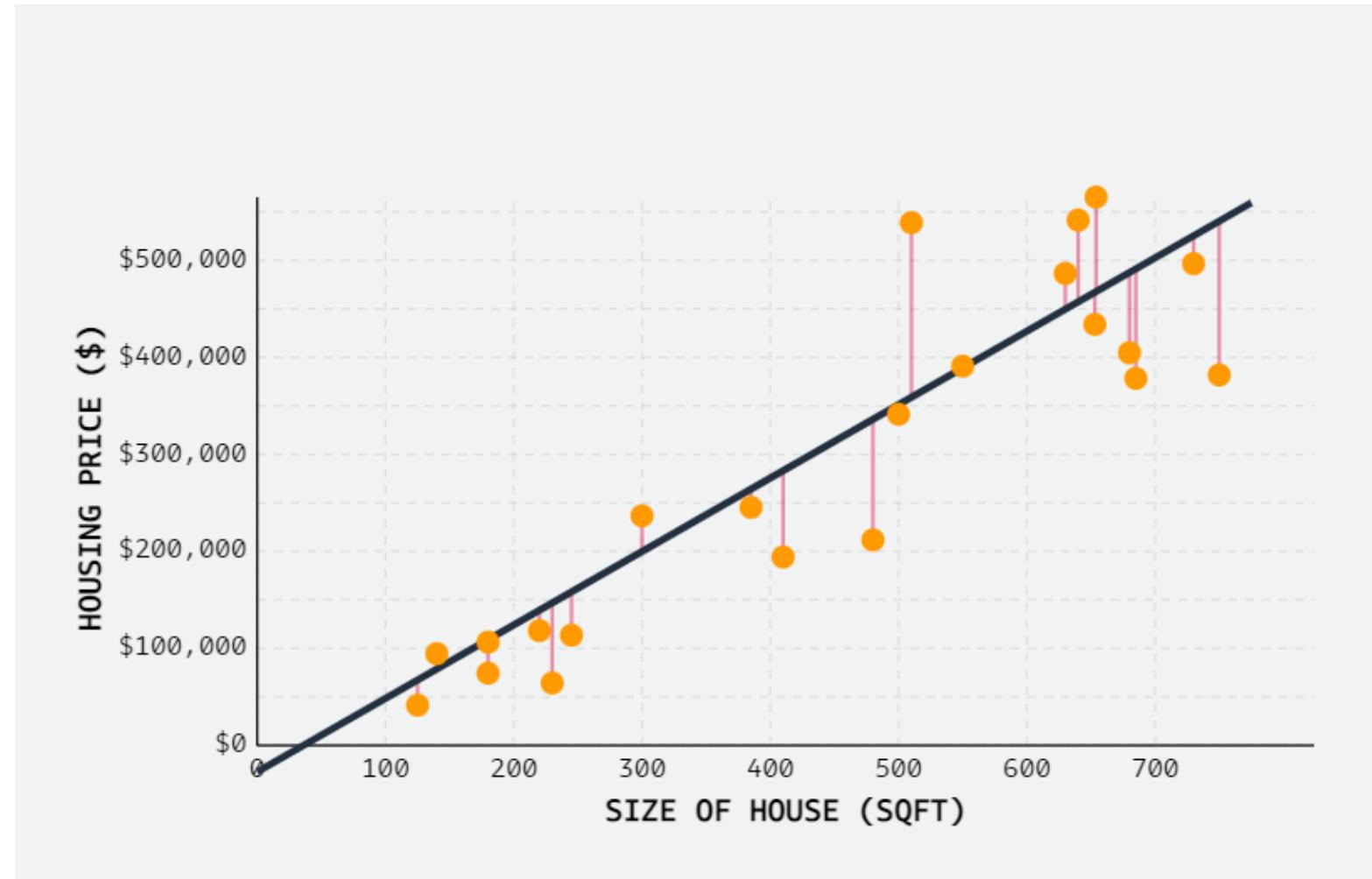


Gradient Descent Iteration: 25 hyperplane weights: 1.096,0.118,0.012



# LINEAR REGRESI

Proprietary document of Indonesia AI 2023



Secara umum, formula dari **Linear Regression** adalah sebagai berikut:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

y: variable yang akan diprediksi

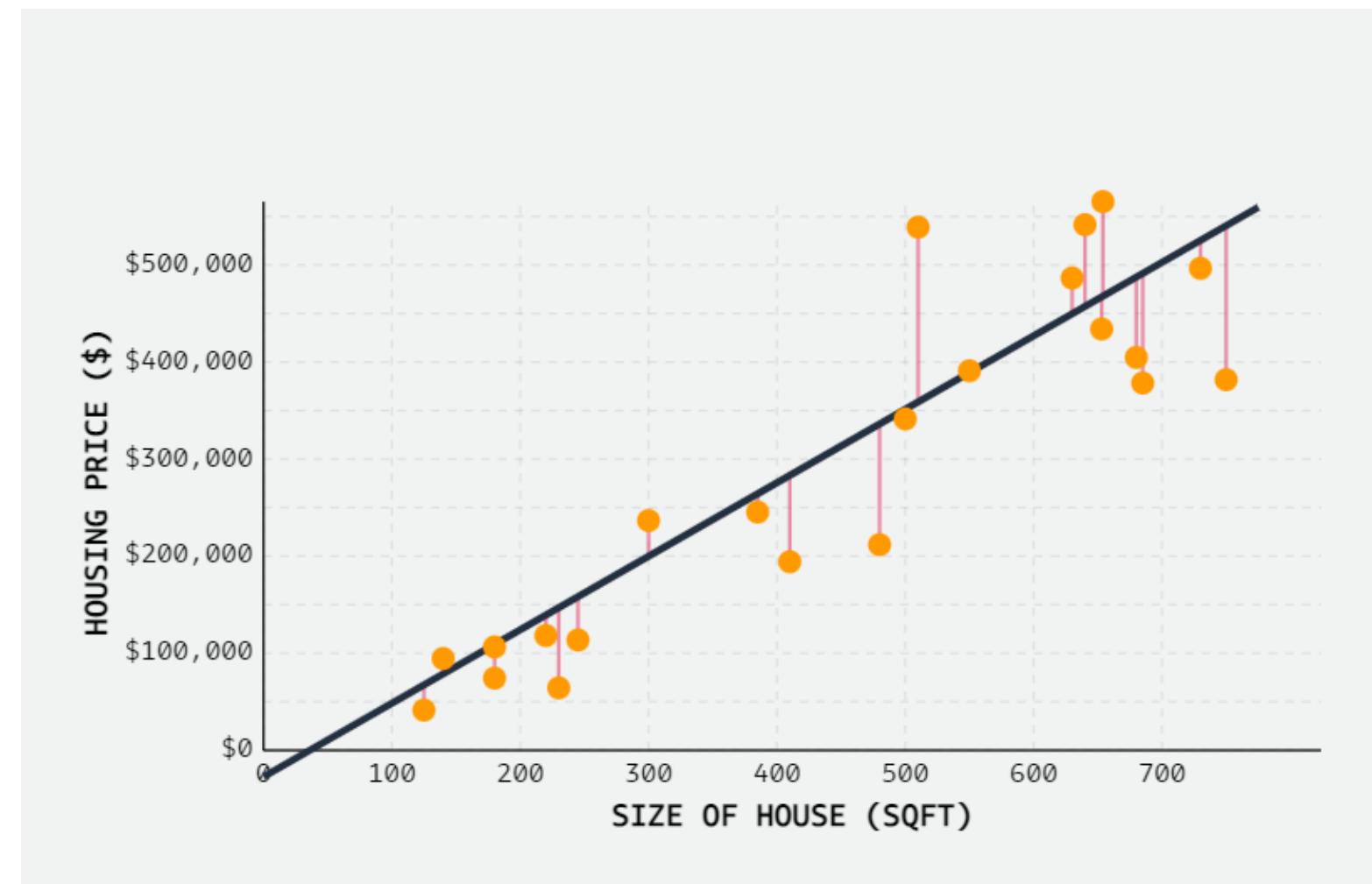
x: variable yang digunakan sebagai fitur

$\beta$ : nilai koefisien dari model regresi

$\beta_0$ : nilai intersepsi

# LINEAR REGRESI

Proprietary document of Indonesia AI 2023



Esensi dari proses belajar **Linear Regression** adalah dengan mendapatkan (mengestimasi) sekumpulan koefisien ( $\beta$ ) yang paling optimal terhadap nilai  $y$ .

Setelah kita mengestimasi nilai koefisien ( $\beta$ ), maka kita bisa melakukan prediksi terhadap nilai-nilai  $y$  di masa depan.

# LINEAR REGRESI

Proprietary document of Indonesia AI 2023

Bagaimana algoritma **Linear Regression** mendapatkan garis/line (atau permukaan/surface) yang "fits" terhadap semua sebaran data?

# Any question guys ~

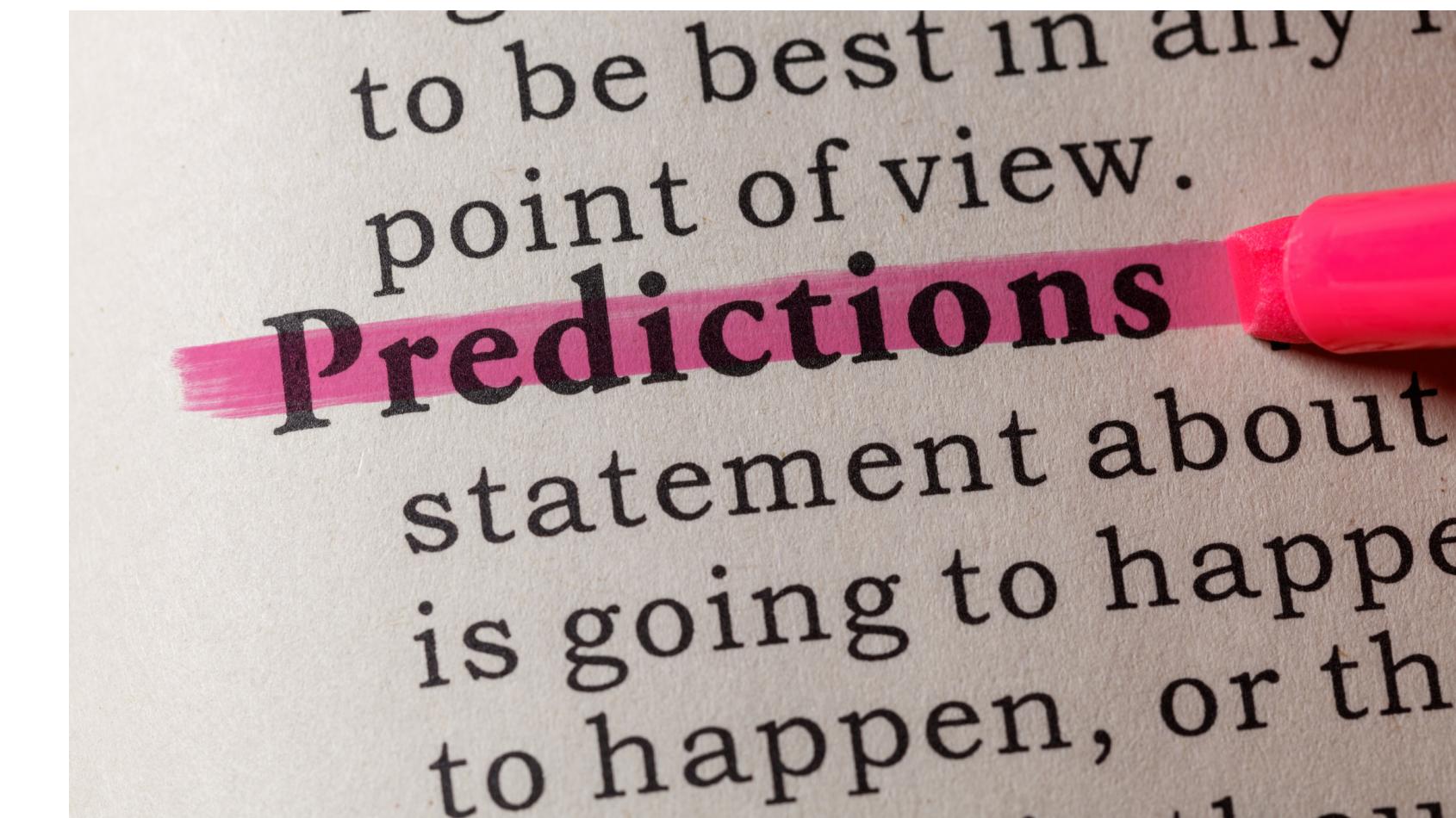
# Algoritma Logistic Regression

# PENGENALAN

Proprietary document of Indonesia AI 2023

**Logistic Regression** adalah salah satu algoritma yang populer untuk menyelesaikan **persoalan Machine Learning**. Dinamakan demikian, karena teknik dasarnya sama seperti apa yang ada di Linear Regression.

Namun, Logistic Regression tidak untuk menyelesaikan persoalan Regresi, **melainkan Klasifikasi**.



# PERSOALAN KLASIFIKASI

Proprietary document of Indonesia AI 2023



## REGRESI

*Berapa Gaji Karyawan X?*

## KLASIFIKASI

*Apakah Karyawan X  
dipromosikan?*

# PERSOALAN KLASIFIKASI

Proprietary document of Indonesia AI 2023



Status Karyawan



Transaksi Fraud



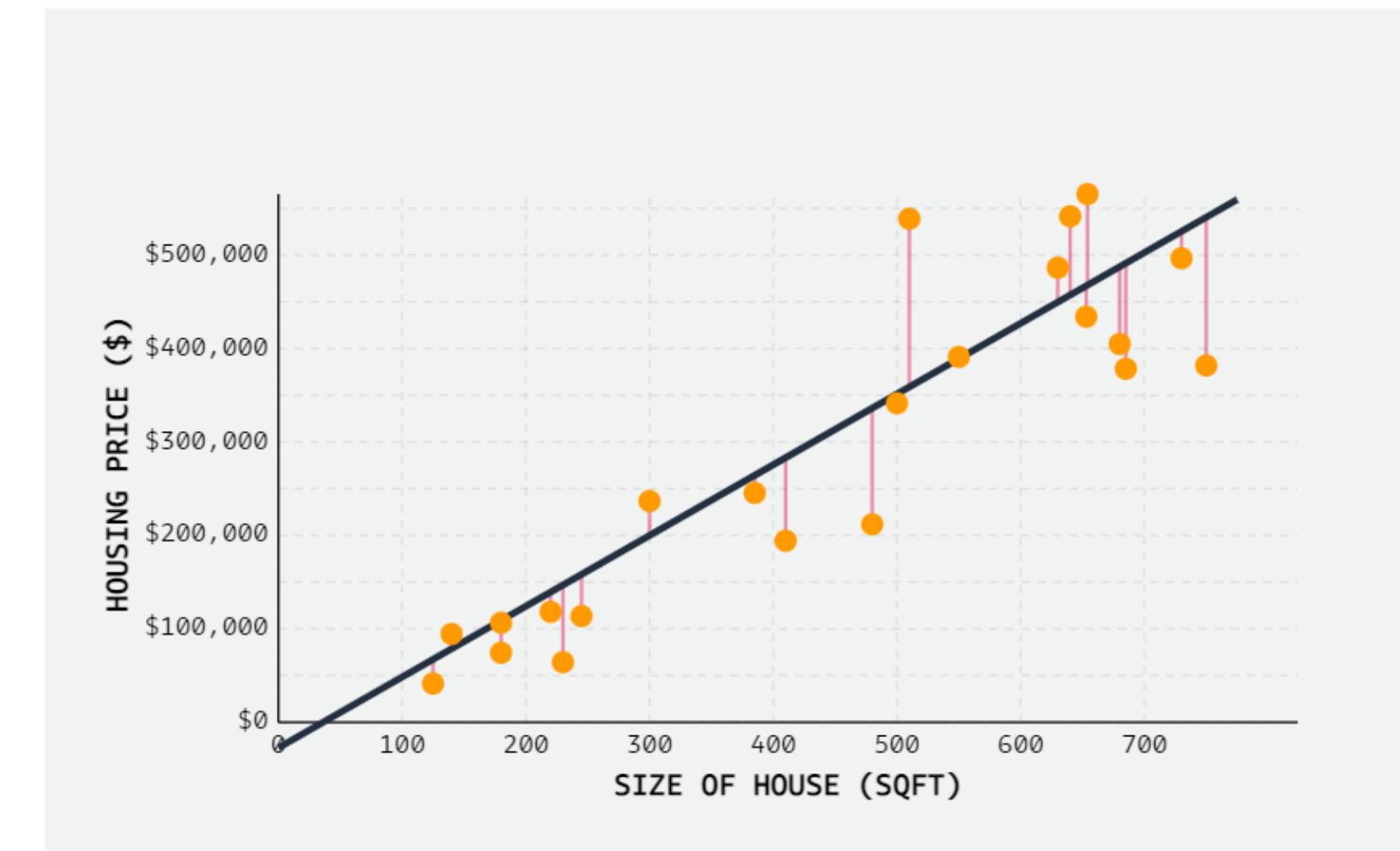
Penyakit Kanker

# GARIS REGRESI

Proprietary document of Indonesia AI 2023

Pada algoritma **Linear Regression**, kita sudah mengenali Garis Regresi yang berada pada posisi optimal di antara sebaran data yang ada dan dengan garis regresi tersebut kita bisa memprediksi data-data baru.

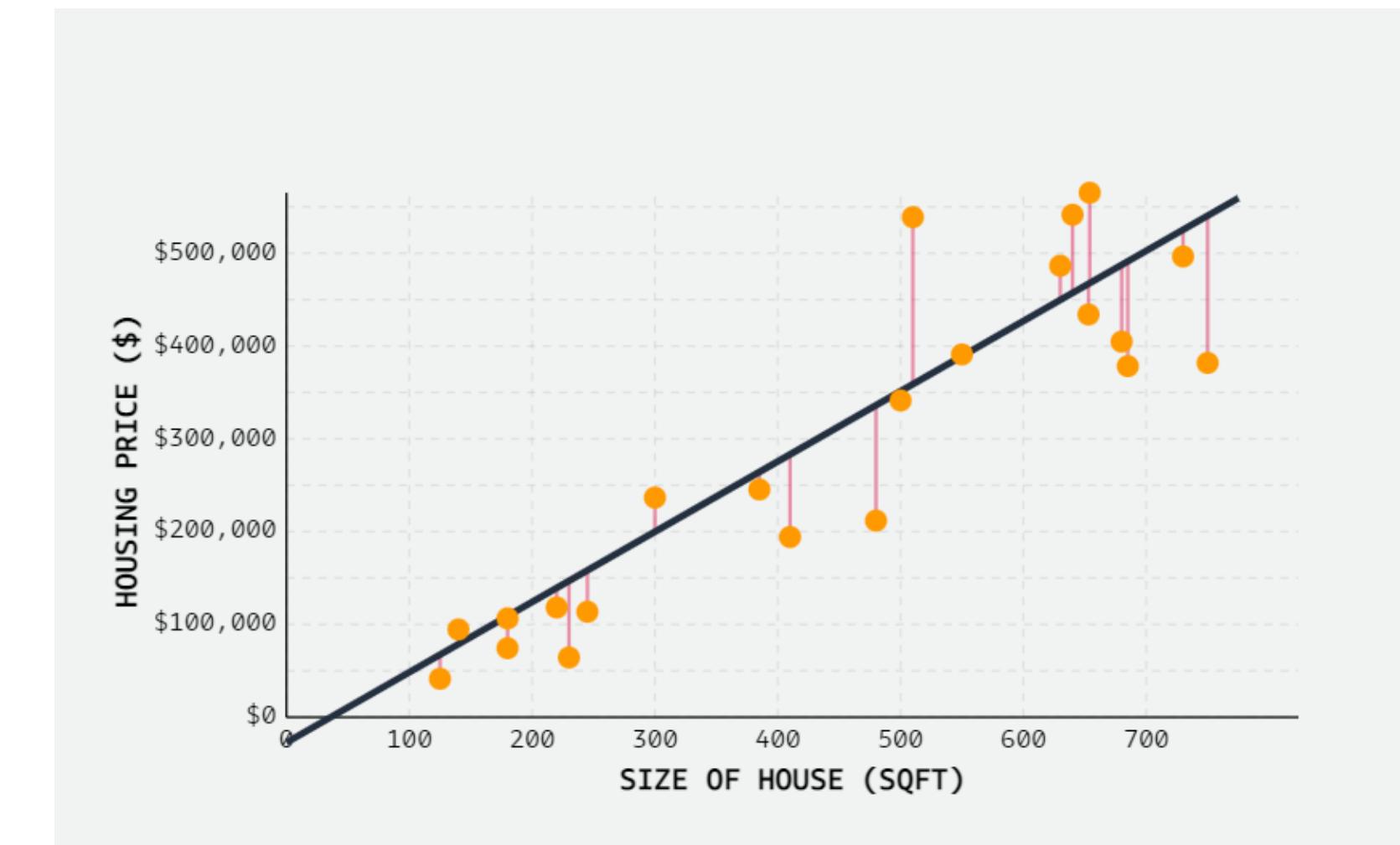
Garis tersebut bisa dimanfaatkan untuk persoalan klasifikasi!



# GARIS REGRESI

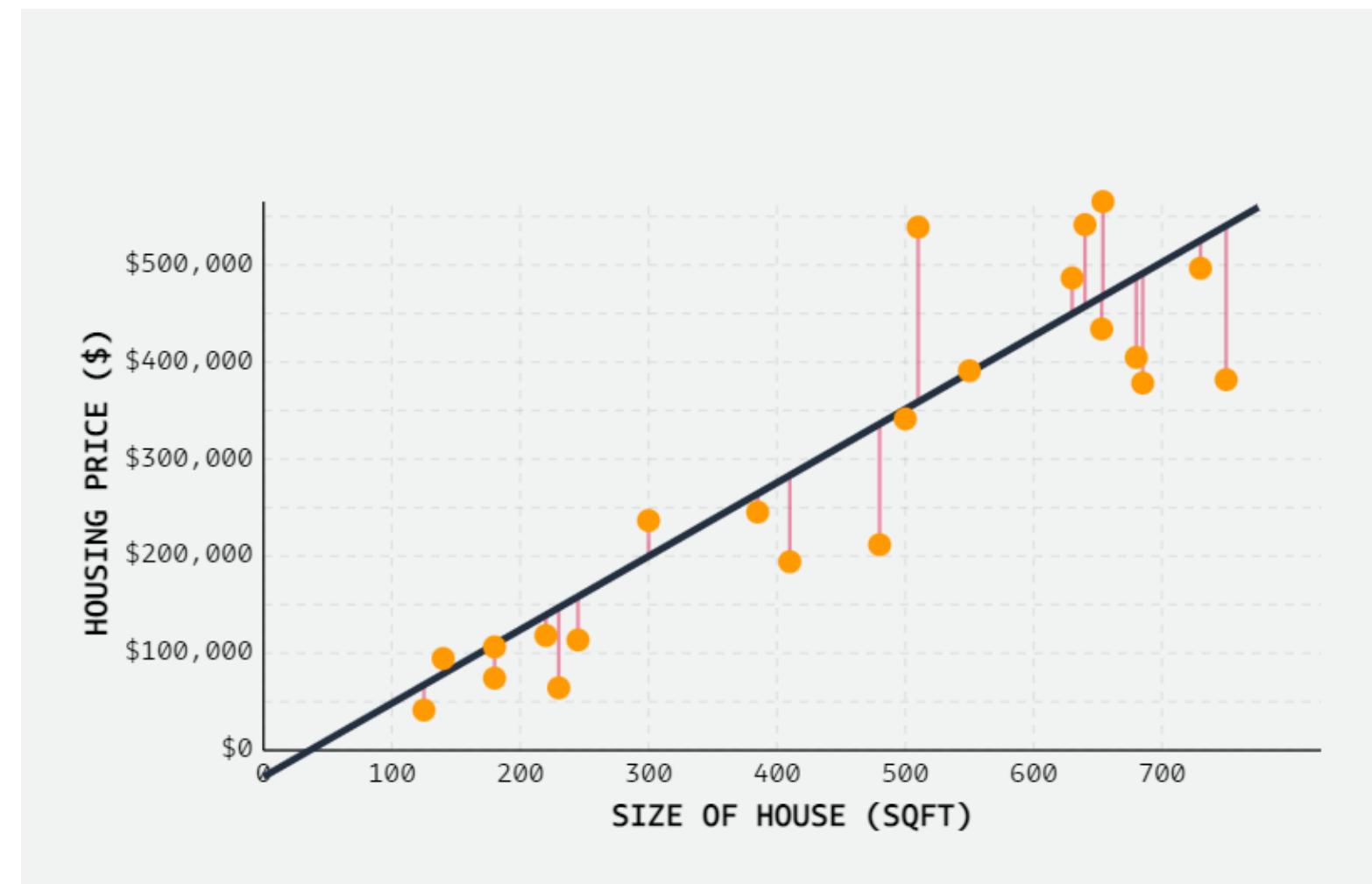
Proprietary document of Indonesia AI 2023

Plot sebaran data pada persoalan **Promosi Karyawan**, maka sebaran data akan berada pada nilai 0 atau 1. Dan jika garis regresi sudah berada di lokasi yang optimal, maka asumsikan **threshold pada nilai 0.5** untuk menentukan apakah karyawan dipromosikan atau tidak. Mudah bukan?



# GARIS REGRESI

Proprietary document of Indonesia AI 2023



## OUTLIER PADA DATA

**Garis Regresi** yang optimal tidak akan mampu melakukan proses klasifikasi dengan baik ketika ada bahkan 1 saja outlier pada data. Akan banyak data yang pada akhirnya terjadi misklasifikasi (banyak karyawan layak yang tidak dipromosikan).

# SIGMOID FUNCTION

Proprietary document of Indonesia AI 2023

Untuk menghadapi masalah pada outlier, algoritma Logistic Regression pada akhirnya dilengkapi dengan fungsi kontrol yang disebut **Sigmoid Function!**

# SIGMOID FUNCTION

Proprietary document of Indonesia AI 2023

Algoritma **Logistic Regression** menggunakan **Sigmoid Function** untuk mengatasi data outlier. Sigmoid Function, yang disebut juga sebagai Logistic Function, akan memproses semua nilai input menjadi nilai pada rentang 0 sd 1 menggunakan formula berikut.

$$S(x) = \frac{1}{1 + e^{-x}}$$

# Any question guys ~

# Terimakasih!