

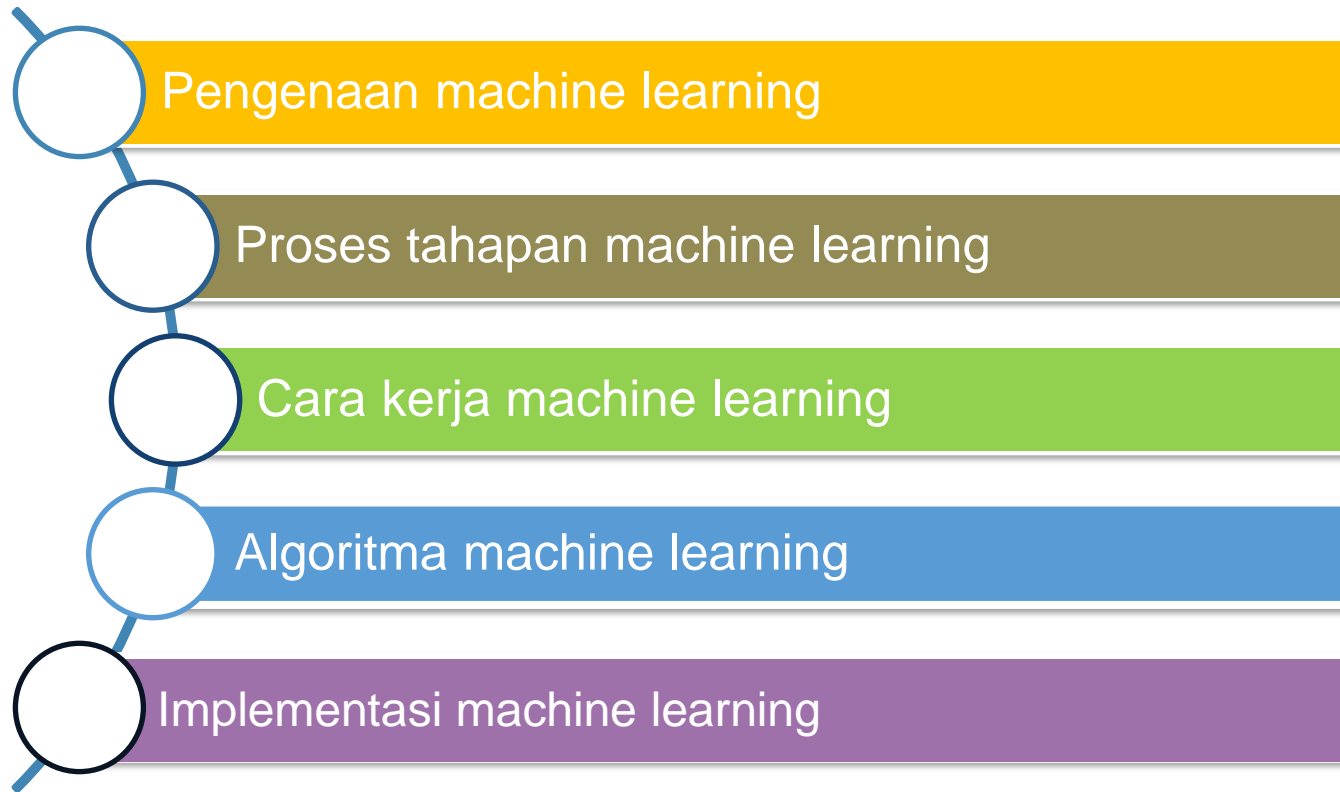


IBIK

Machine Learning

Pingky Dezar Zulkarnain, SE., M.Sc., Ph.D
Putri Cholidhazia., ST., Mkom

Overview

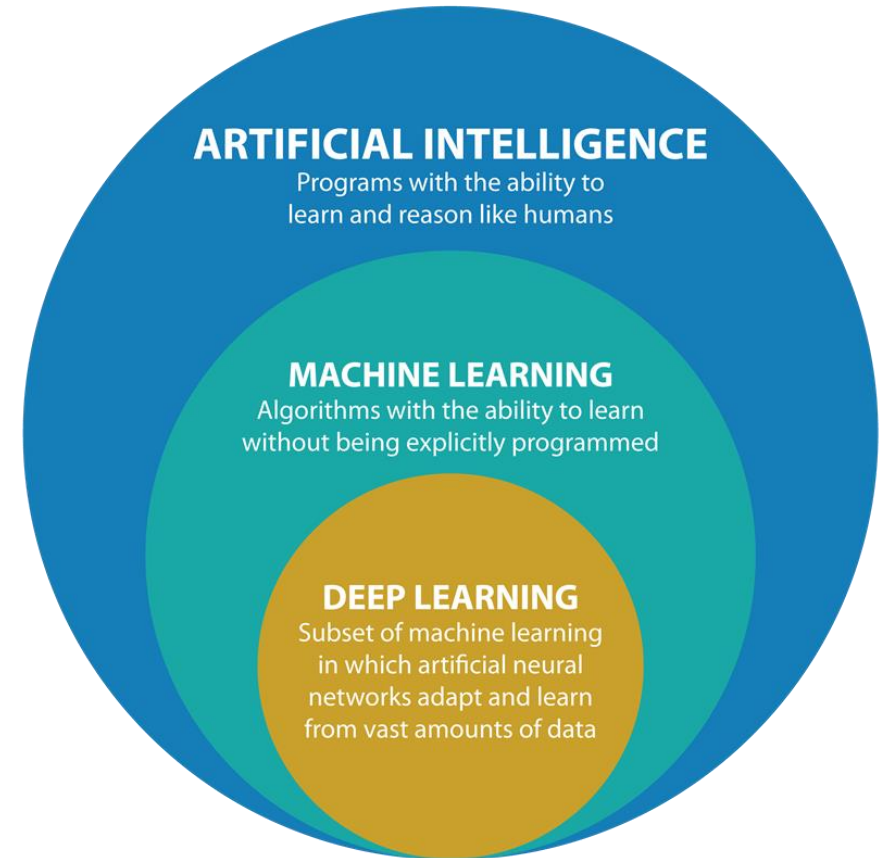


Perbedaan AI, Machine Learning?

Artificial Intelligence:

Adalah sebuah teknologi yang diciptakan untuk mengimplementasikan kecerdasan manusia. Dimana teknologi ini akan di implementasikan pada mesin computer

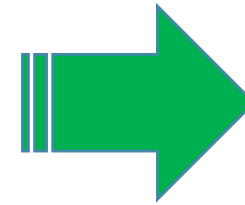
Contoh: Sistem Pakar



Contoh sederhana AI

Sistem Pakar

Membangun pengetahuan berupa aturan kompleks yang secara eksplisit kita terapkan dalam program

[illegible]

Bagaimana system kerja dari AI Rule based IF THEN ELSE

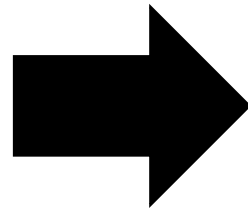


Sistem Kerjanya

- Kita berusaha memahami data / kondisi
- Kita merancang aturan berdasarkan aturan tersebut
- Kita implementasikan data tersebut
- Mesin menjalankan aturan tsb.

Bagaimana jika data kita kompleks?

Data yang kompleks.



```
1 # Program Artificial Intelligence
2 # Expert System v1.0
3
4 if (begini):
5     maka.begitu()
6
7     if (kalo gini):
8         maka.harus_gitu()
9     else (kalo gitu):
10         maka.jangan_gini_banget()
11
12 else (begitu):
13     maka.begono()
```

Bisa gak bikin program yang
kalau dikasih data mesin bisa
belajar sendiri ?

Machine Learning

Machine Learning:

Sub-area dari computer science yang mampu memberikan computer kemampuan untuk belajar tanpa diprogram secara eksplisit

Rule – Based



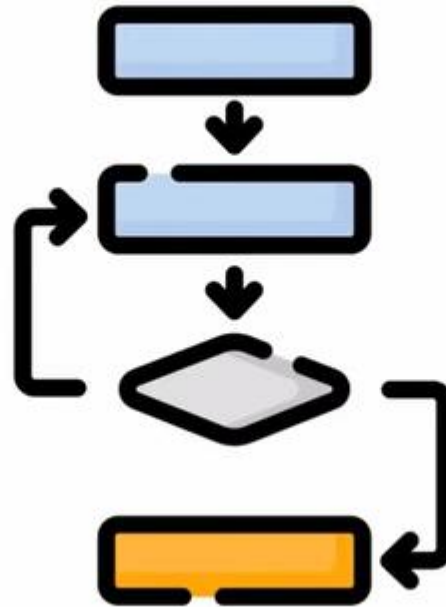
```
1 # Program Artificial Inteligence
2 # Expert System v1.0
3
4 if (begini):
5     maka.begitu()
6
7     if (kalo gini):
8         maka.harus_gitu()
9     else (kalo gitu):
10         maka.jangan_gini_banget()
11
12 else (begitu):
13     maka.begono()
```

Model based



```
1 # Program Machine Learning
2 # Algoritma 1 v2.0
3
4 data = data.ambil()
5
6 model = AlgoritmaML()
7 model.belajar(data)
8
9 print(model.prediksi())
```

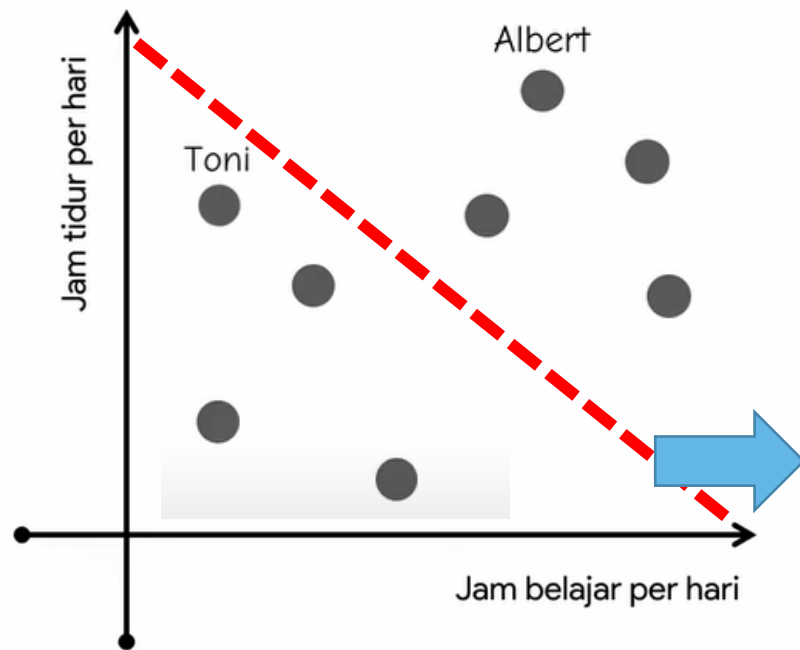

MENCIPTAKAN
ALGORITMA
BERDASARKAN
BAGAIMANA KITA
MEMAHAMI DATA



MENCIPTAKAN
ALGORITMA
YANG
MAMPU
MEMAHAMI DATA

Contoh I:

Kelompokan data ini menjadi 2 kelompok, kelompok lulus dan tidak lulus ?

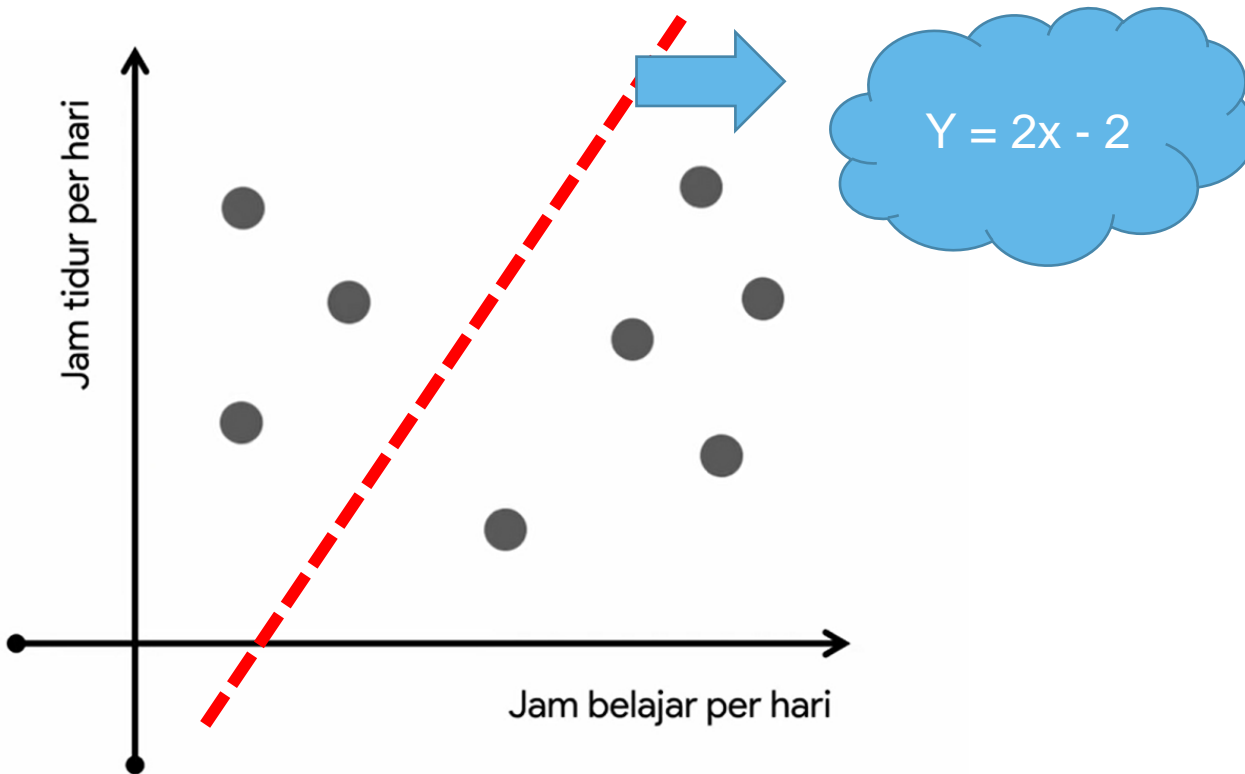


$$Y = -3x + 5$$

```
1 # ambil data, masukan ke x
2 x = ambil.data()
3
4
5 # pokoknya w tau aja disini persamaan gua tadi
6 y = -3*(x) + 5
7
8 # kalo diatas garis
9 if (y>0):
10 # kelompokkin lulus
11 print("lulus")
12
13 # kalo dibawah garis
14 else:
15 # kelompokkin ga lulus
16 print("ga lulus")
```

Contoh I :

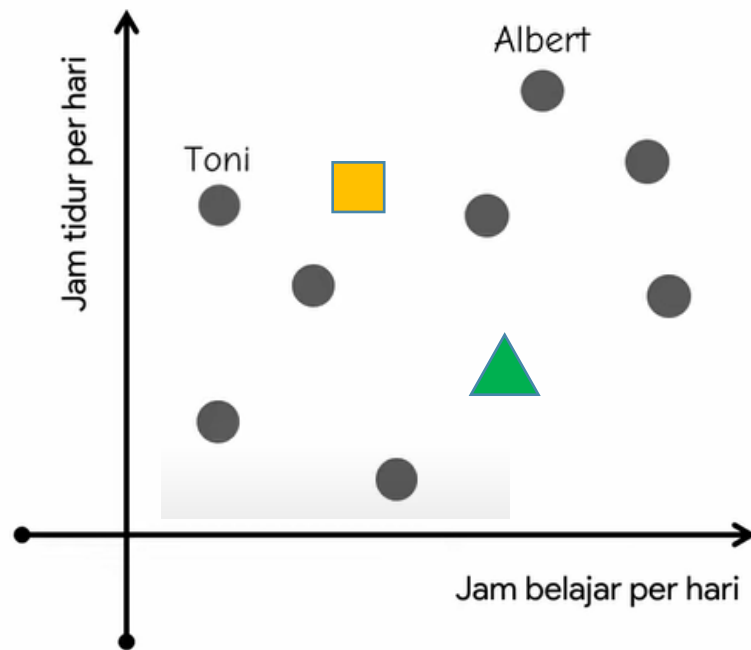
Kelompokan data ini menjadi 2 kelompok, kelompok lulus dan tidak lulus ?



```
1 # ambil data, masukan ke x
2 x = ambil.data()
3
4
5
6 y = 2*(x) - 2
7
8 # kalo diatas garis
9 if (y>0):
10     # kelompokkin lulus
11     print("lulus")
12
13 # kalo dibawah garis
14 else:
15     # kelompokkin ga lulus
16     print("ga lulus")
```

Contoh II:

Kelompokan data ini menjadi 2 kelompok, kelompok lulus dan tidak lulus ?

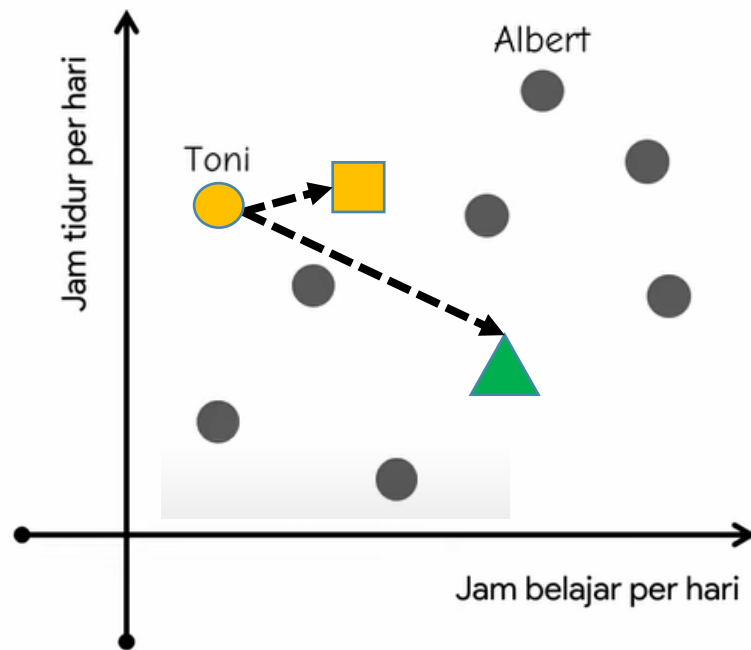


Sistem Kerjanya

1. Tempat dua titik centroid secara sembarang

Contoh II:

Kelompokan data ini menjadi 2 kelompok, kelompok lulus dan tidak lulus ?

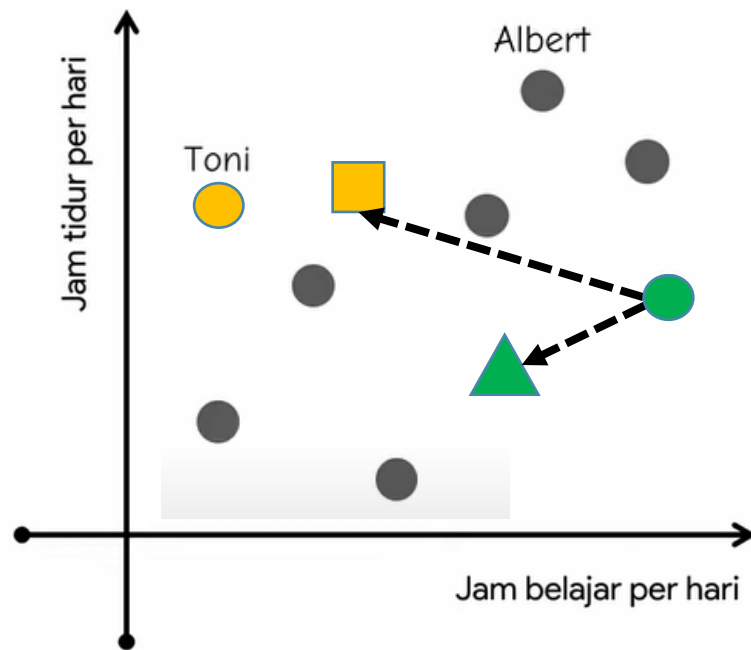


Sistem Kerjanya

1. Tempat dua titik centroid secara sembarang
2. Untuk setiap data tentukan mana centroid yang terdekat

Contoh II:

Kelompokan data ini menjadi 2 kelompok, kelompok lulus dan tidak lulus ?

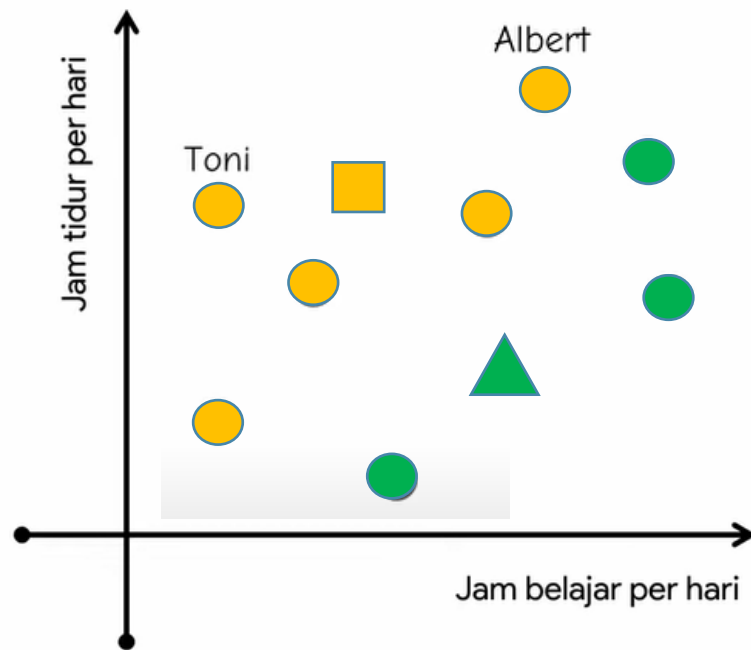


Sistem Kerjanya

1. Tempat dua titik centroid secara sembarang
2. Untuk setiap data tentukan mana centroid yang terdekat

Contoh II:

Kelompokan data ini menjadi 2 kelompok, kelompok lulus dan tidak lulus ?

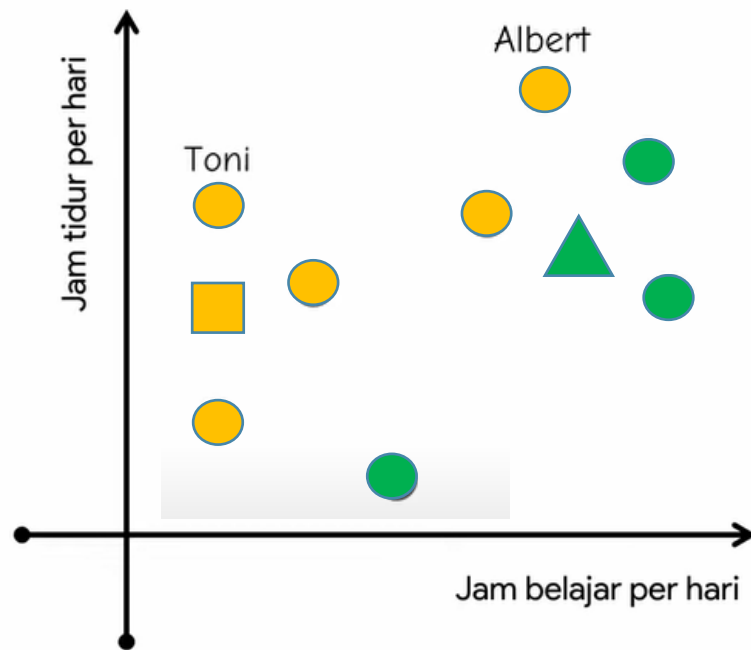


Sistem Kerjanya

1. Tempatkan dua titik centroid secara sembarangan
2. Untuk setiap data tentukan mana centroid yang terdekat
3. Pindahkan centroid ke tengah kelompoknya

Contoh II:

Kelompokan data ini menjadi 2 kelompok, kelompok lulus dan tidak lulus ?

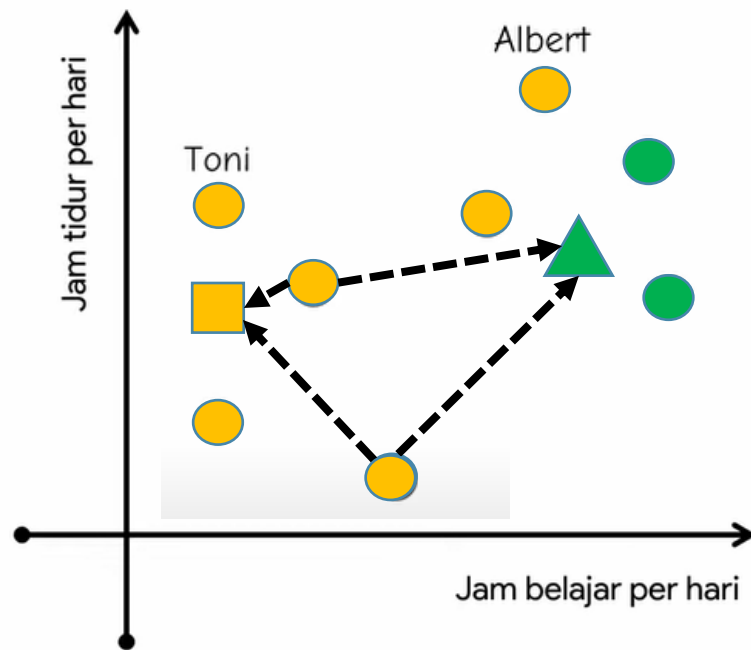


Sistem Kerjanya

1. Tempatkan dua titik centroid secara sembarangan
2. Untuk setiap data tentukan mana centroid yang terdekat
3. Pindahkan centroid ke tengah kelompoknya
4. Ulangi Langkah 2 dan 3

Contoh II:

Kelompokan data ini menjadi 2 kelompok, kelompok lulus dan tidak lulus ?

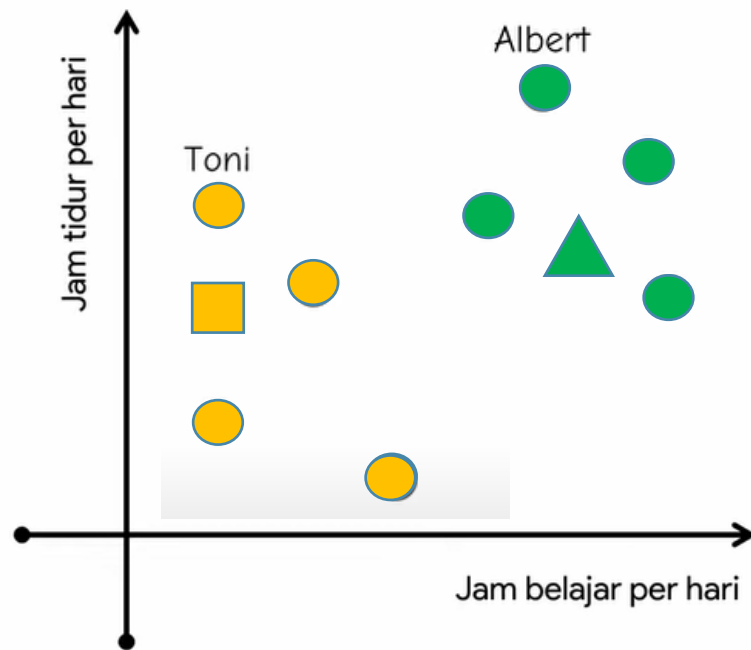


Sistem Kerjanya

1. Tempatkan dua titik centroid secara sembarangan
2. Untuk setiap data tentukan mana centroid yang terdekat
3. Pindahkan centroid ke tengah kelompoknya
4. Ulangi Langkah 2 dan 3

Contoh II:

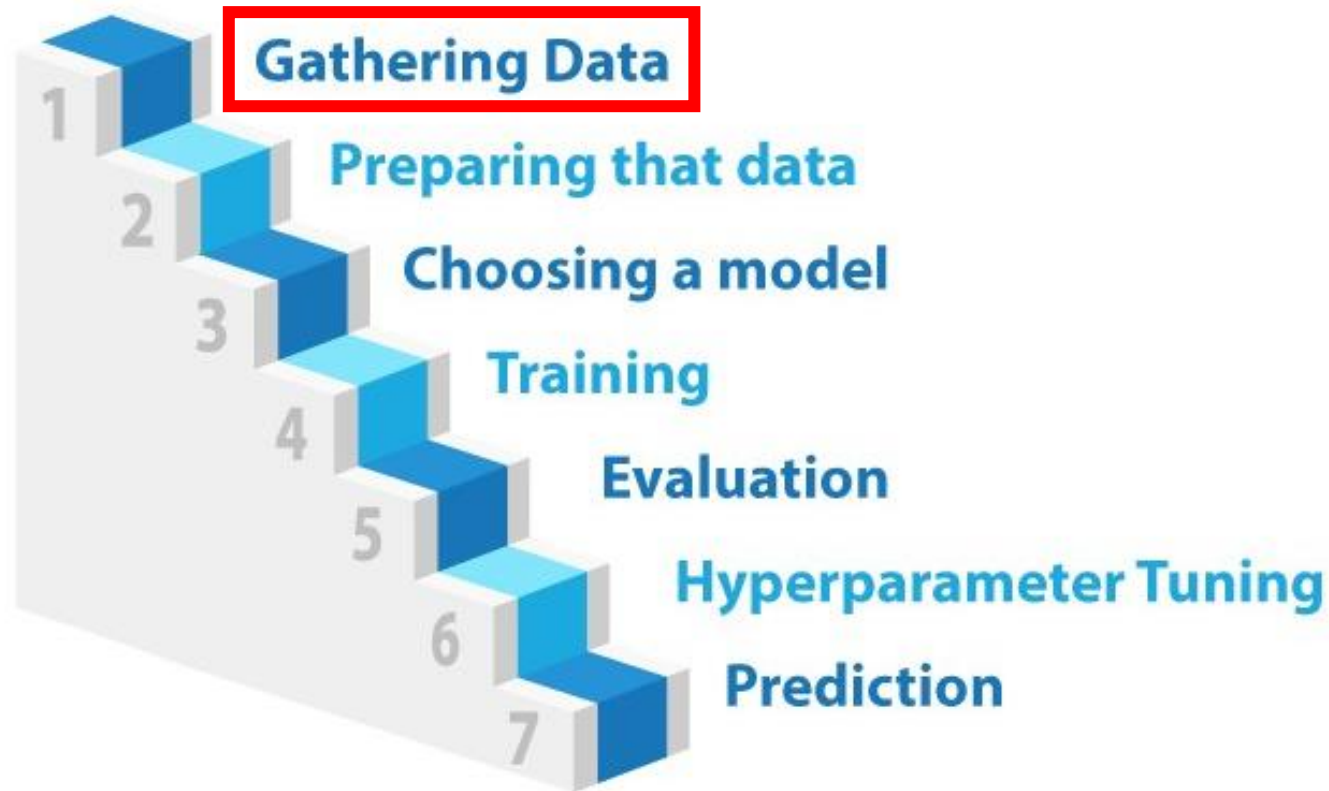
Kelompokan data ini menjadi 2 kelompok, kelompok lulus dan tidak lulus ?



Sistem Kerjanya

1. Tempatkan dua titik centroid secara sembarangan
2. Untuk setiap data tentukan mana centroid yang terdekat
3. Pindahkan centroid ke tengah kelompoknya
4. Ulangi Langkah 2 dan 3

Tahapan dalam machine learning



Gathering Data (mengumpulkan data)

Data mentah bisa berupa Excel, Ms Access, file teks dan lain-lain. Langkah ini membentuk dasar pembelajaran masa depan. Semakin banyak variasi, kepadatan dan volume data yang relevan, semakin baik prospek pembelajaran untuk mesin.

Situs akses data:

<https://databoks.katadata.co.id/>

<https://www.kaggle.com/datasets>

<https://archive.ics.uci.edu/ml/datasets.php>

<https://www.bps.go.id/>

<https://data.jakarta.go.id/>

Gathering Data

Source : Data Bunga - Kaggle

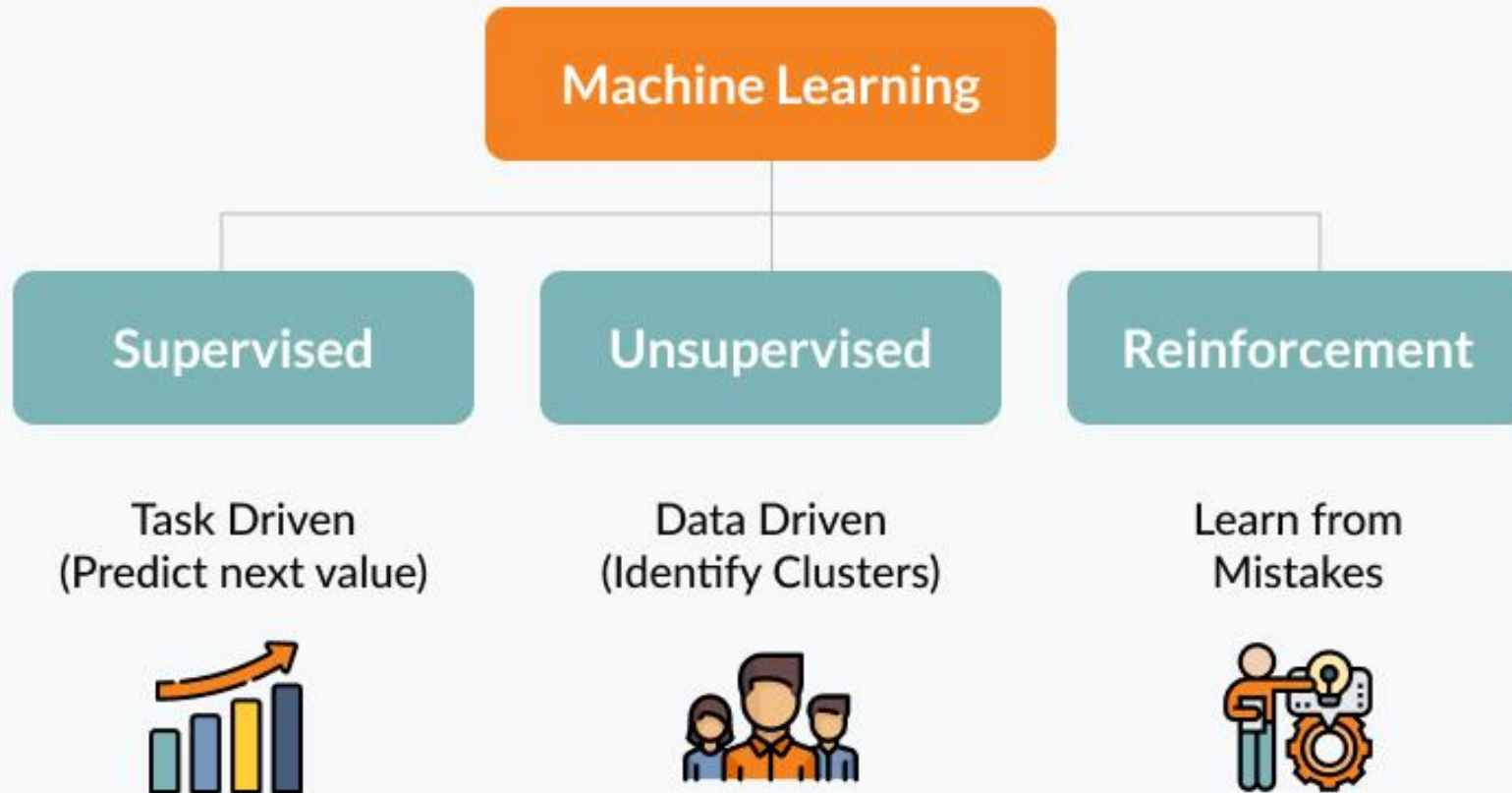
Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
1	5.1	3.5	1.4	0.2	Iris-setosa
2	4.9	3.0	1.4	0.2	Iris-setosa
3	4.7	3.2	1.3	0.2	Iris-setosa
4	4.6	3.1	1.5	0.2	Iris-setosa
5	5.0	3.6	1.4	0.2	Iris-setosa
6	5.4	3.9	1.7	0.4	Iris-setosa
7	4.6	3.4	1.4	0.3	Iris-setosa
8	5.0	3.4	1.5	0.2	Iris-setosa
9	4.4	2.9	1.4	0.2	Iris-setosa
10	4.9	3.1	1.5	0.1	Iris-setosa
11	5.4	3.7	1.5	0.2	Iris-setosa
12	4.8	3.4	1.6	0.2	Iris-setosa

Source: Data Kesehatan balita - Jurnal

Balita ke-	Tinggi Badan (TB)	Berat Badan (BB)
Balita 1	65	5,8
Balita 2	65	7,2
Balita 3	60	5
Balita 4	60	8
Balita 5	52	5,8
Balita 6	51	5
Balita 7	54	3,5
Balita 8	52,5	7,8
Balita 9	70	4,2
Balita 10	71	6,2
Balita 11	72,5	7
Balita 12	71,5	8,5
Balita 13	55	5,5
Balita 14	57	4,8
Balita 15	52	6,5
Balita 16	46,5	5,7
Balita 17	95	12
Balita 18	82	9,7
Balita 19	75	8
Balita 20	99	11
Balita 21	99	7,8
Balita 22	97,5	10
Balita 23	88	9,4

Persiapan Data

Types of Machine Learning

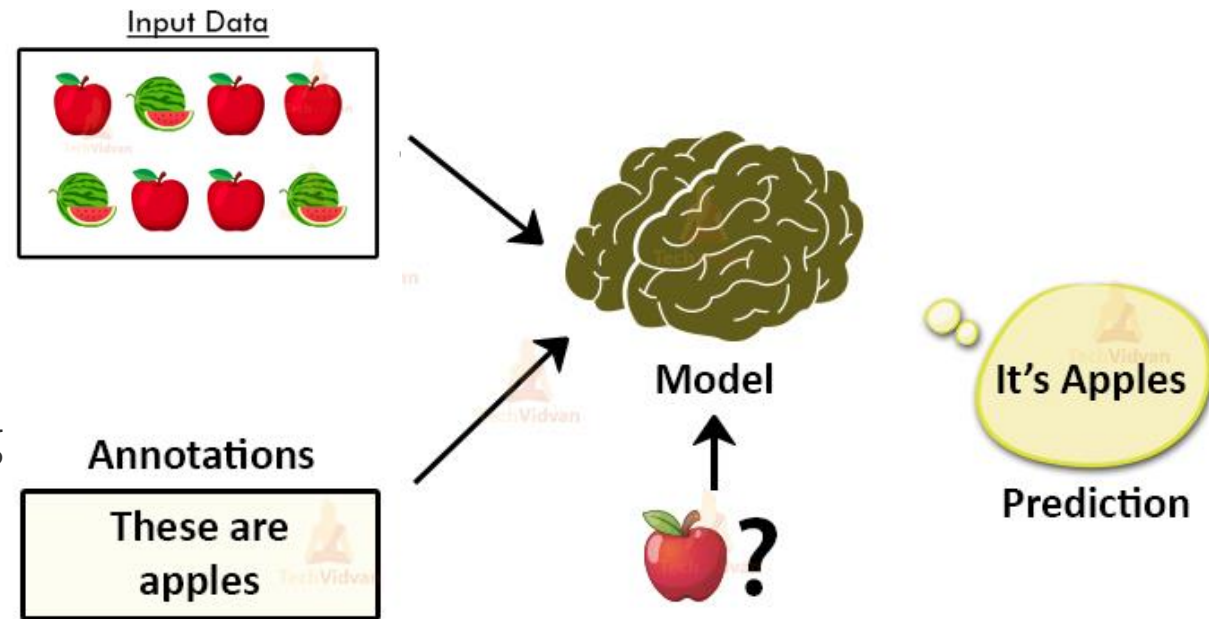


Supervised

Algoritma supervised ditandai dengan:

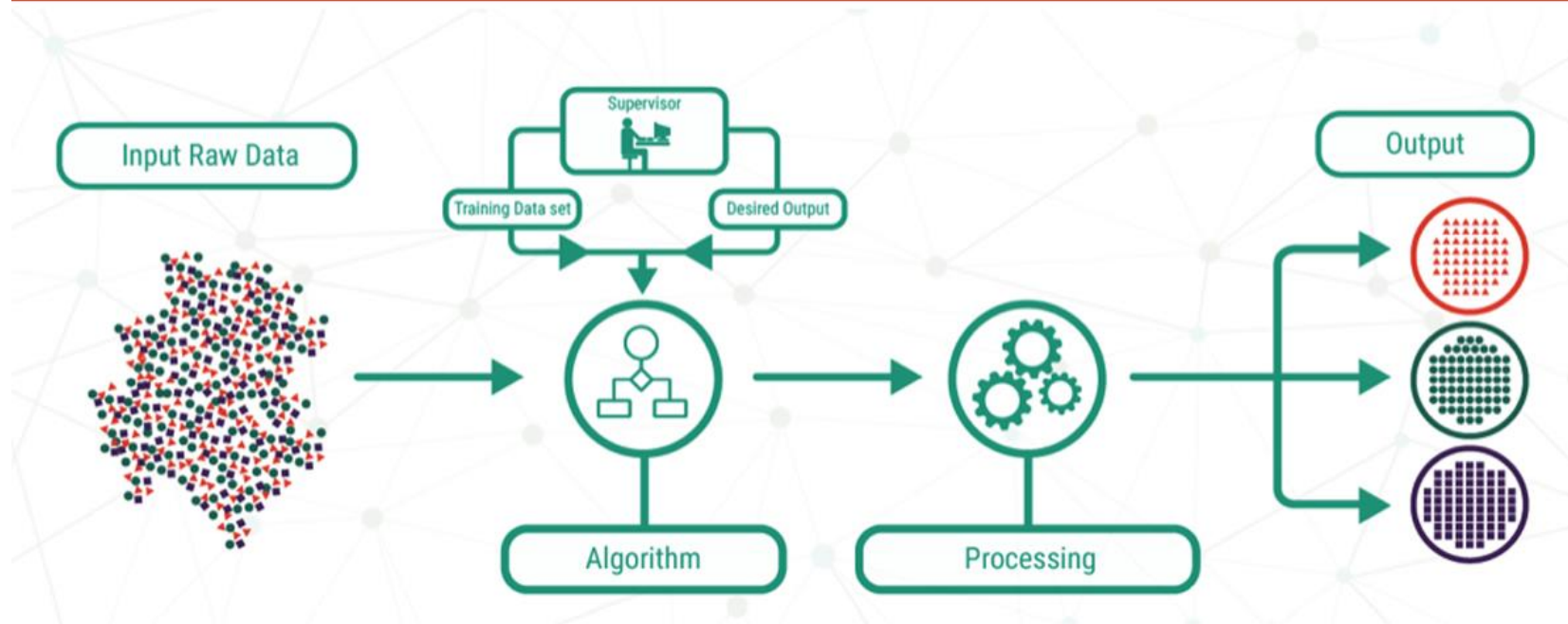
1. Data memiliki label
2. Data yang sudah ditandai sebagai jawaban yang benar. Hal ini yang menyebabkan algoritma ini disebut dengan algoritma yang diawasi.
3. Algoritma supervised learning belajar dari data training berlabel yang dapat dimanfaatkan untuk memprediksi hasil untuk data yang tidak terduga.

Supervised Learning in ML



Supervised Learning

Supervised Learning

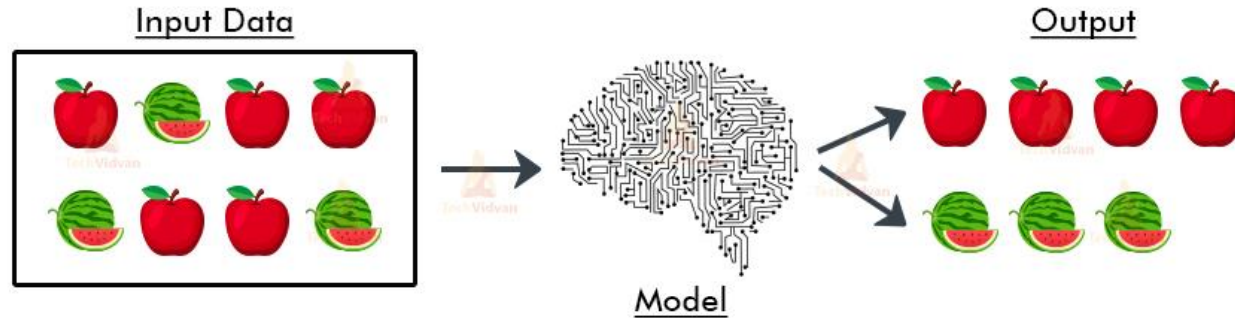


Unsupervised

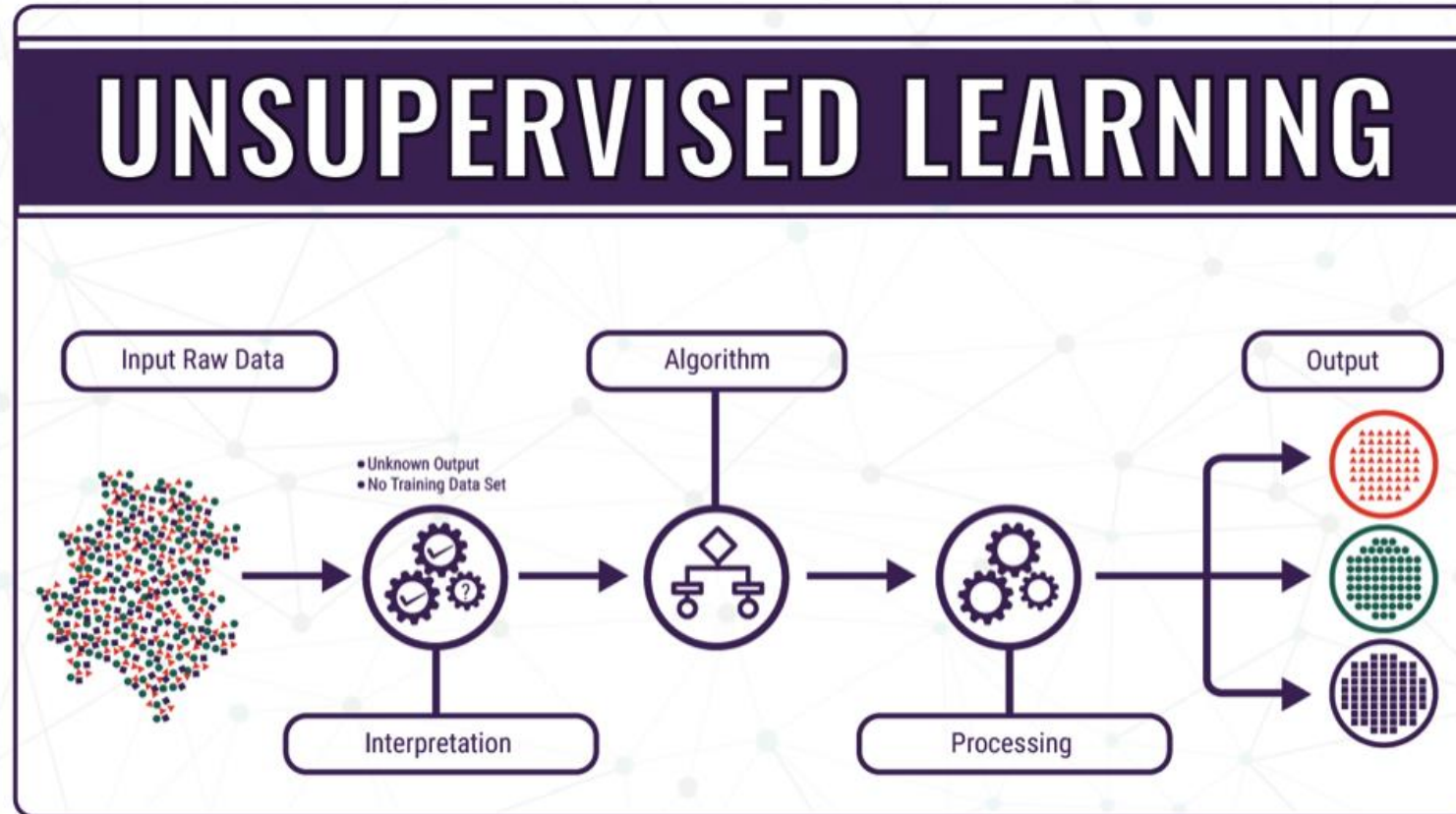
Algoritma supervised ditandai dengan:

1. Data tidak memiliki label
2. Teknik pembelajaran mesin, dimana kita tidak perlu mengawasi modelnya, kita hanya perlu mengizinkan model untuk bekerja sendiri mencari informasi.
3. Algoritma unsupervised learning bisa menemukan semua jenis pola yang tidak diketahui di dalam data.
4. Metode unsupervised learning akan menemukan fitur yang berguna untuk proses kategorisasi. Algoritma unsupervised learning berlangsung secara real time. Artinya, semua data input dianalisis dan kemudian hasil analisisnya baru diberi label.

Unsupervised Learning in ML



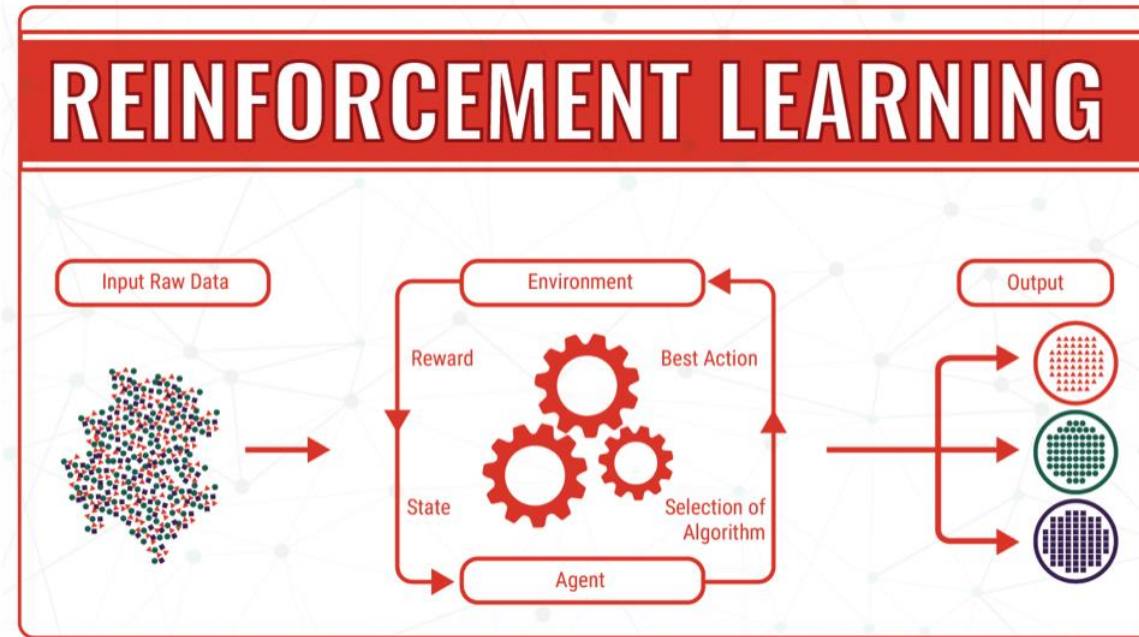
Unsupervised Learning



Reinforcement

Algoritma reinforcement ditandai dengan:

memungkinkan komputer dapat belajar sendiri dari lingkungan. Komputer akan melakukan pencarian sendiri (self-discovery) dengan cara berinteraksi dengan lingkungan. Selama proses training, komputer akan dituntun oleh algoritma untuk melakukan kegiatan trial and error.



Kesimpulan

Supervised Learning

X_1	X_2	X_3	X_p	Y

Target

Un-Supervised Learning

X_1	X_2	X_3	X_p	Y

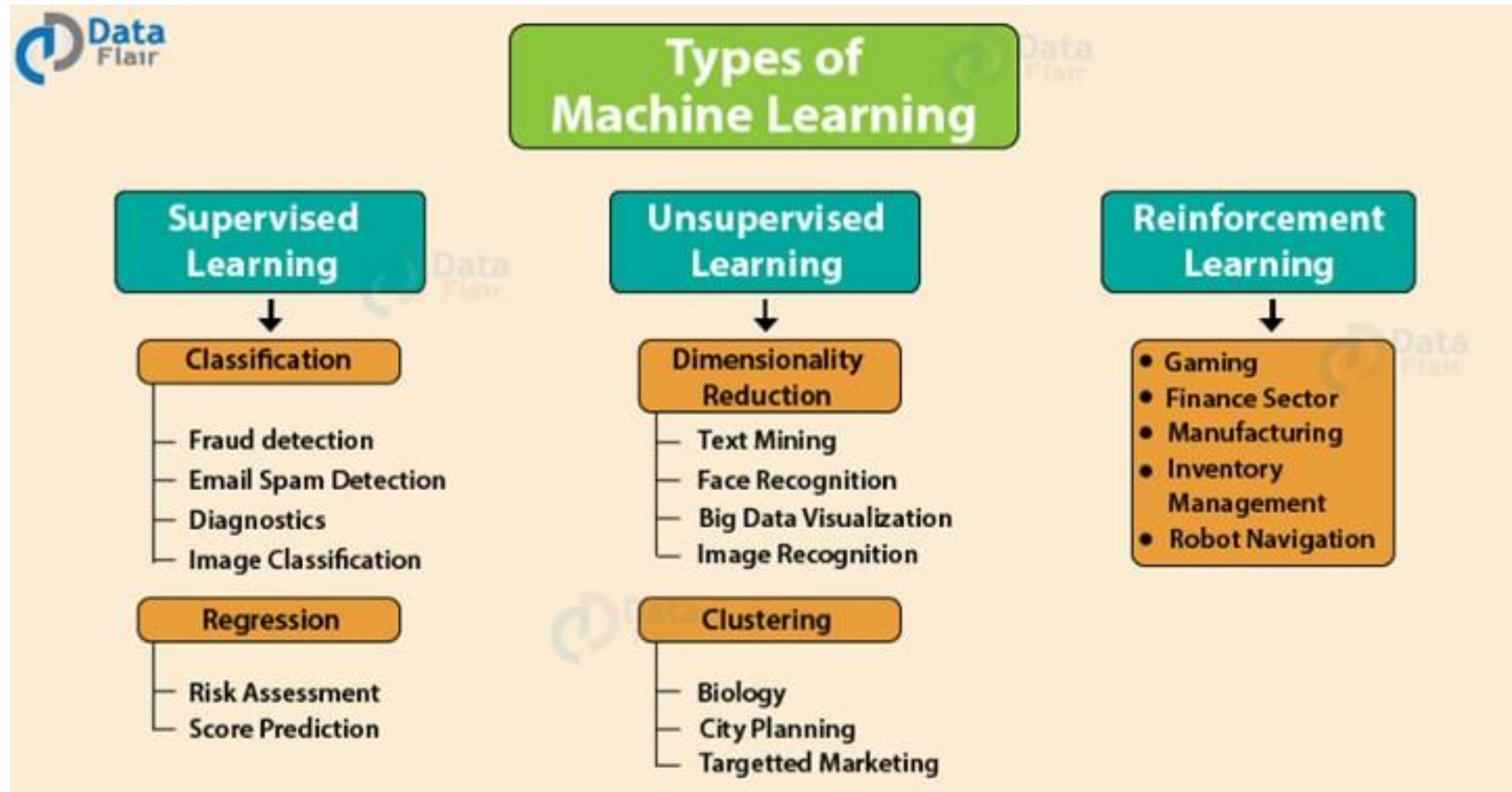
No
Target

Data Selection

Data Preprocessing

Data Transformation

Teknik & Implementasi Machine Learning





Classification

Classification (Klasifikasi) : Klasifikasi adalah sebuah teknik untuk mengklasifikasikan atau mengkategorikan beberapa *item* yang belum berlabel ke dalam sebuah set kelas diskrit.

Berkaitan dengan output data kategorik seperti apakah berwarna merah atau tidak, apakah cuaca cerah atau mendung, sehat atau sakit

Classification

Klasifikasi mencoba mempelajari hubungan antara kumpulan variabel fitur dan variabel target. Dalam klasifikasi, variabel targetnya bertipe kategori.

Feature Variables 				Target Variable 
sepal-length	sepal-width	petal-length	petal-width	class
5.1	3.5	1.4	0.2	Iris-setosa
5.4	3.9	1.7	0.4	Iris-setosa
5.9	3.2	4.8	1.8	Iris-versicolor
6.8	2.8	4.8	1.4	Iris-versicolor
6.9	3.2	5.7	2.3	Iris-virginica
7.4	2.8	6.1	1.9	Iris-virginica
6.2	2.8	4.8	1.8	?

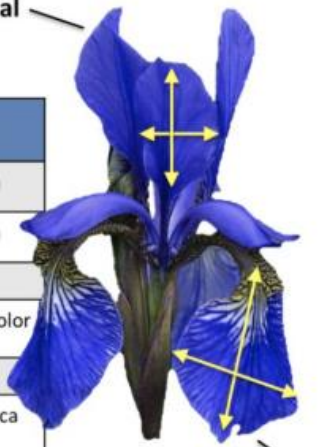
Classification

Samples
(instances, observations)

	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
...					
50	6.4	3.5	4.5	1.2	Versicolor
...					
150	5.9	3.0	5.0	1.8	Virginica

Features
(attributes, measurements, dimensions)

Class labels
(targets)



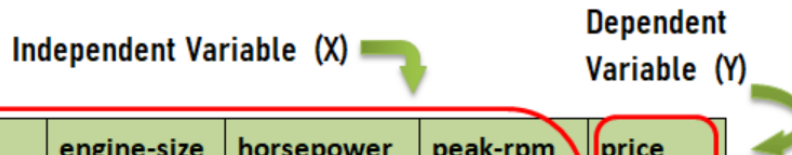
Model classification **bertujuan** untuk menentukan kelas berdasarkan atribut tertentu.

Pada kasus klasifikasi Iris, sebuah model bertugas untuk memprediksi spesies dari sebuah bunga iris berdasarkan atributnya yaitu panjang sepal, lebar sepal, panjangpetal, dan lebar petalnya.

Contohnya panjang petal dari Iris Setosa lebih pendek dari spesies versicolor dan virginica. Maka jika panjang petal pendek maka kemungkinan spesies Iris tersebut adalah Setosa.

Regression

- **Regression** (Regresi) : Suatu teknik analisis untuk mengidentifikasi relasi atau hubungan diantara dua variabel atau lebih.
- Model regresi memprediksi bilangan kontinu.



Independent Variable (X)				Dependent Variable (Y)
body-style	engine-size	horsepower	peak-rpm	price
convertible	130	111	5000	13495
convertible	130	111	5000	16500
hatchback	152	154	5000	16500
sedan	109	102	5500	13950
sedan	136	115	5500	17450
wagon	136	110	5500	18920
hatchback	131	160	5500	?

Dimensionality Reduction

Dimensionality Reduction : adalah proses pengurangan dimensi dari data yang berdimensi besar menjadi data yang berdimensi kecil.

Ada dua teknik dalam dimensionality reduction ini, yaitu :

1. Feature selection, memilih feature yang berpengaruh dari sekumpulan data asli.
2. Feature extraction, membentuk feature baru berdasarkan feature yang lama dengan dimensi yang lebih sedikit dibandingkan dengan sebelumnya.

Clustering

Clustering : Sebuah metode machine learning unsupervised untuk mengelompokkan objek-objek yang memiliki kemiripan, ke dalam sebuah klaster.



fig 1: before applying
k-means clustering

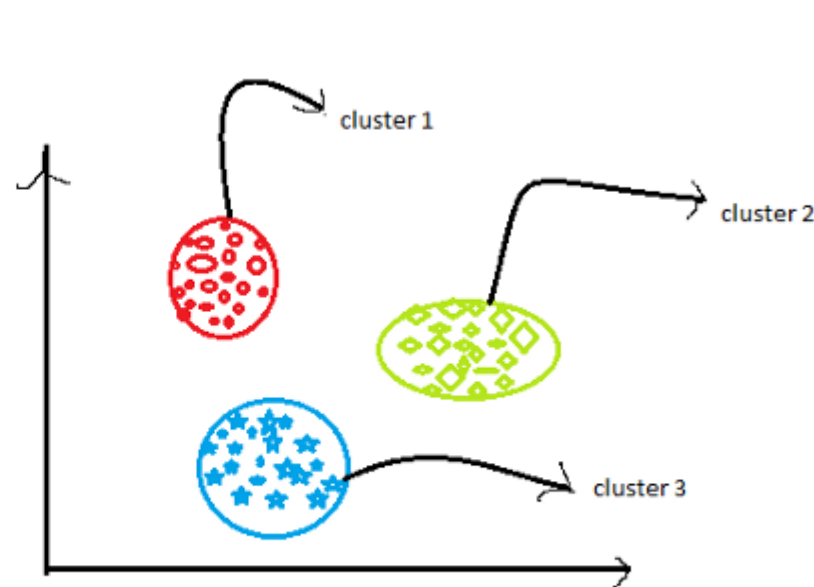


fig 2: After applying K-
means clustering

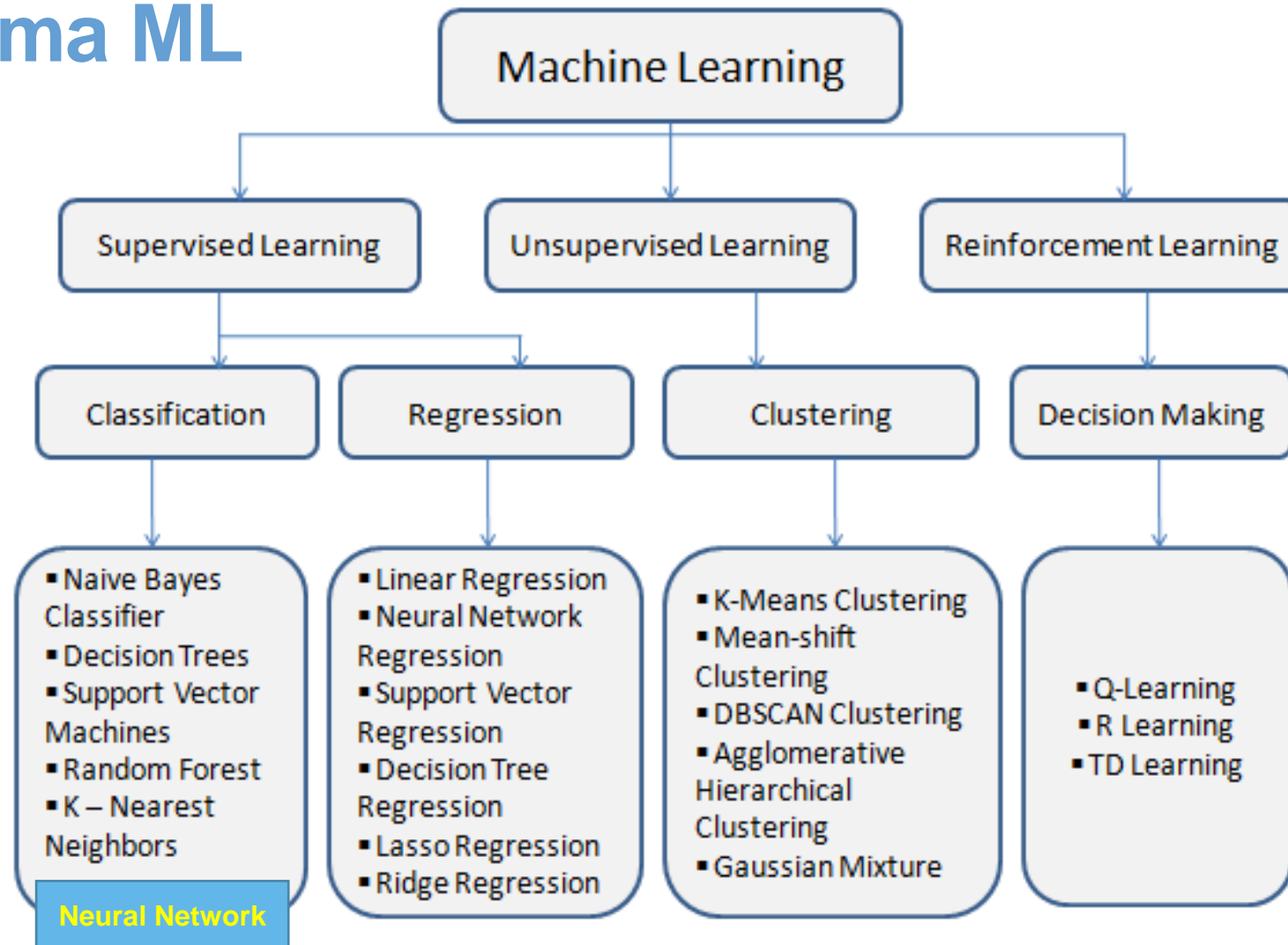
Association rule

Association rule untuk menemukan atribut yang muncul dalam satu waktu. Association rule mining adalah suatu prosedur untuk mencari hubungan antar item dalam suatu data set yang ditentukan.

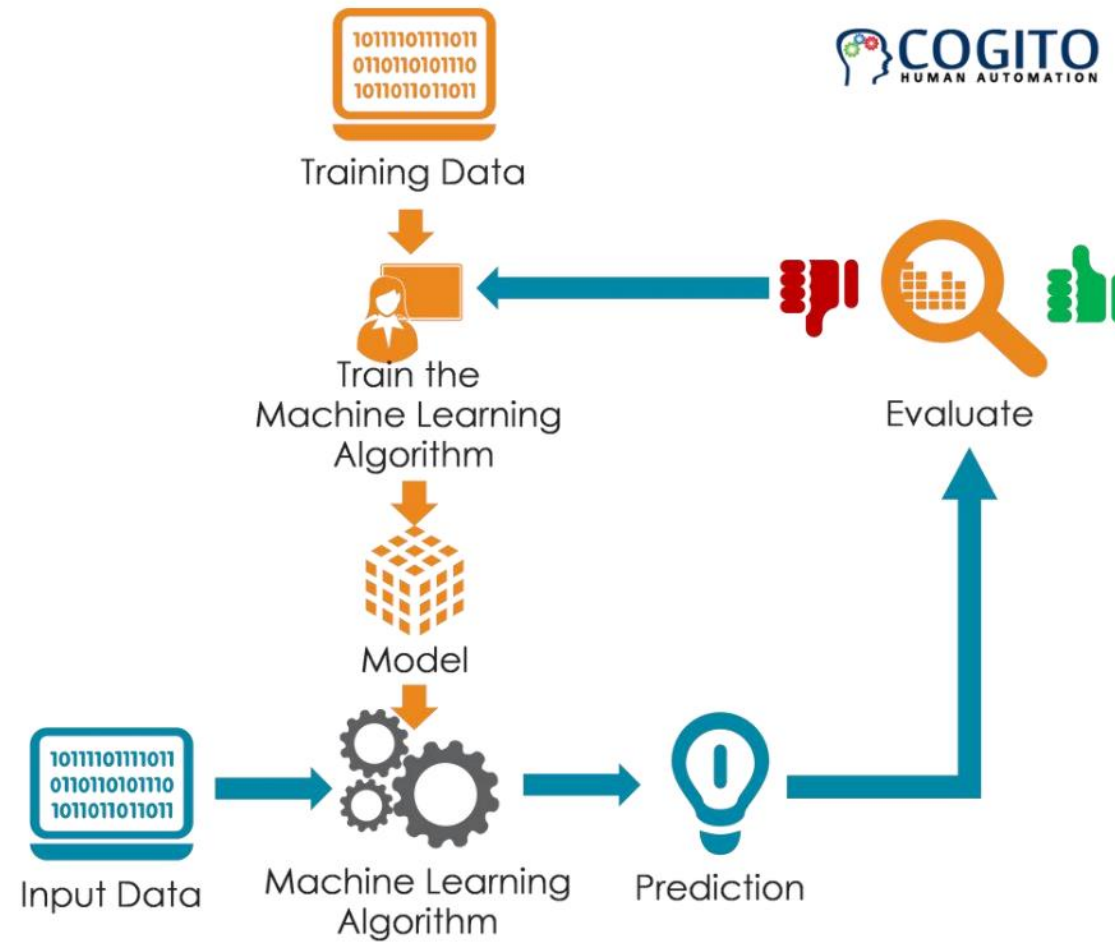
Association rule meliputi dua tahap:

1. Mencari kombinasi yang paling sering terjadi dari suatu item set.
2. Mendefinisikan condition dan result (conditional association rule)

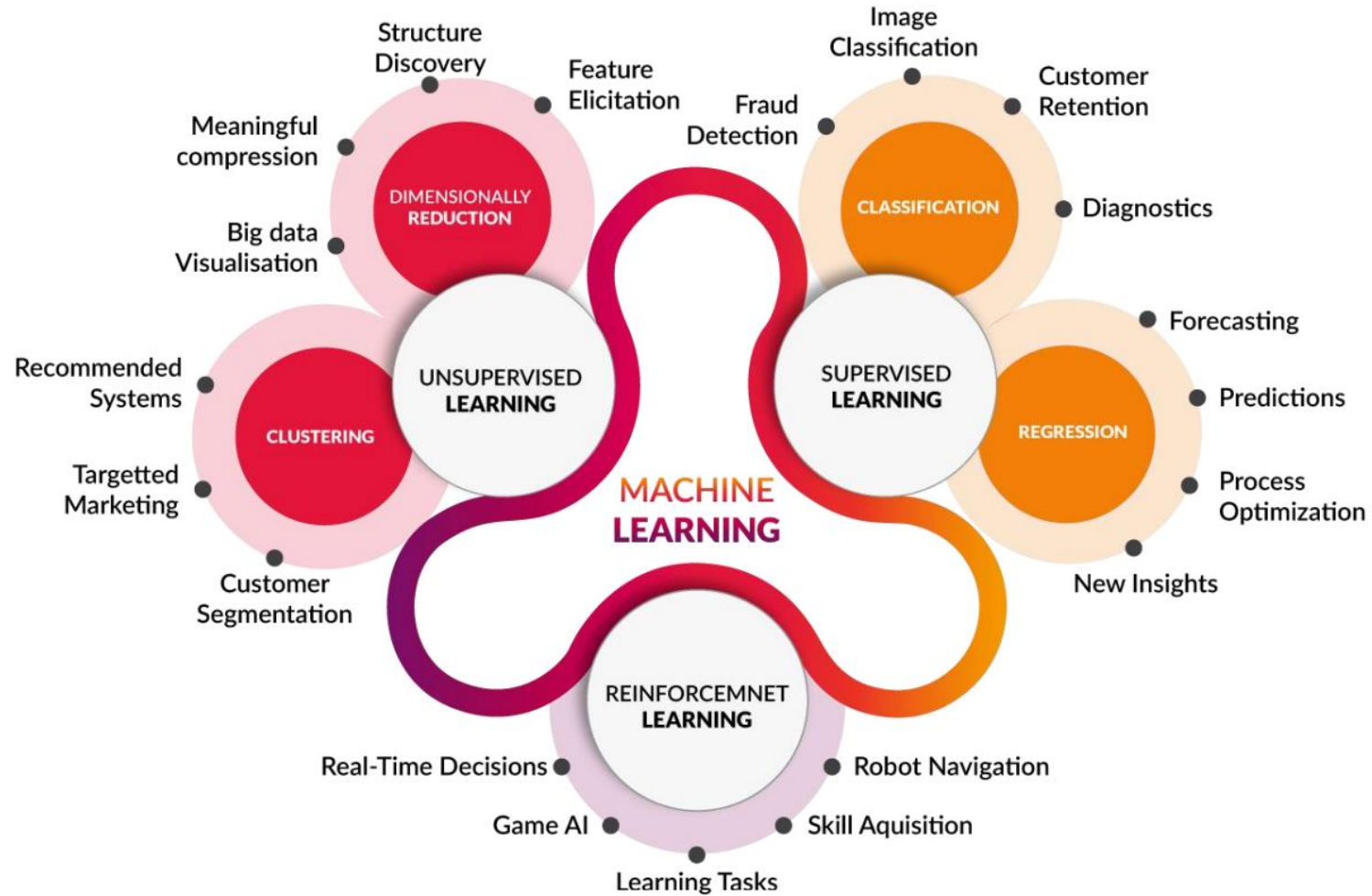
Algoritma ML



Tahapan ML



Penerapan ML



Neural Network (Jaringan Syaraf Tiruan)