

Pengantar Machine Learning

Machine Learning

“Field of study that gives computers the ability to **learn without** being **explicitly programmed**”

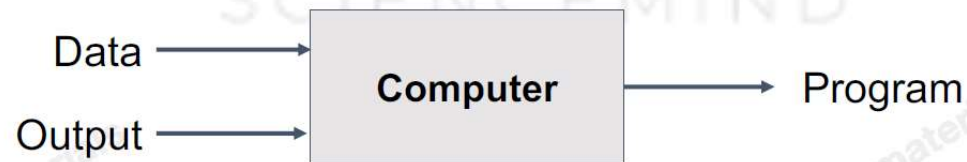
- **Arthur Samuel (1959)**

Traditional Programming vs Machine Learning

Traditional Programming



Machine Learning



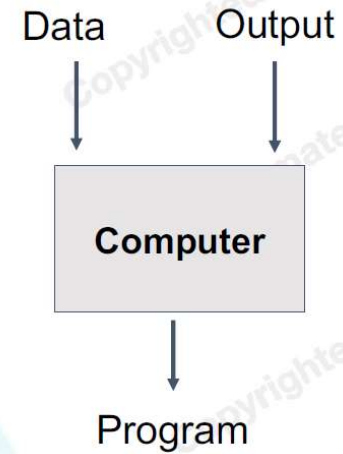
Traditional Programming

Jarak = kecepatan x waktu

```
def hitung_jarak(kecepatan, waktu):  
    return kecepatan*waktu
```

rumus diprogram secara manual oleh manusia

Data - waktu (jam)	Output - jarak (km)
2	20.02
1.8	18.01
3.1	30.98
2.7	27.02
1.5	15.01
2	19.99
3	30.02



```
def hitung_jarak(waktu):  
    return 9.995*waktu+0.01688
```

kedua konstanta
ditemukan oleh komputer

Komponenten



Unstructured data

The university has 5600 students.
John's ID is number 1, he is 18 years old and already holds a B.Sc. degree.
David's ID is number 2, he is 31 years old and holds a Ph.D. degree. Robert's ID is number 3, he is 51 years old and also holds the same degree as David, a Ph.D. degree.

Semi-structured data

```
<University>
  <Student ID="1">
    <Name>John</Name>
    <Age>18</Age>
    <Degree>B.Sc.</Degree>
  </Student>
  <Student ID="2">
    <Name>David</Name>
    <Age>31</Age>
    <Degree>Ph.D. </Degree>
  </Student>
  ....
</University>
```

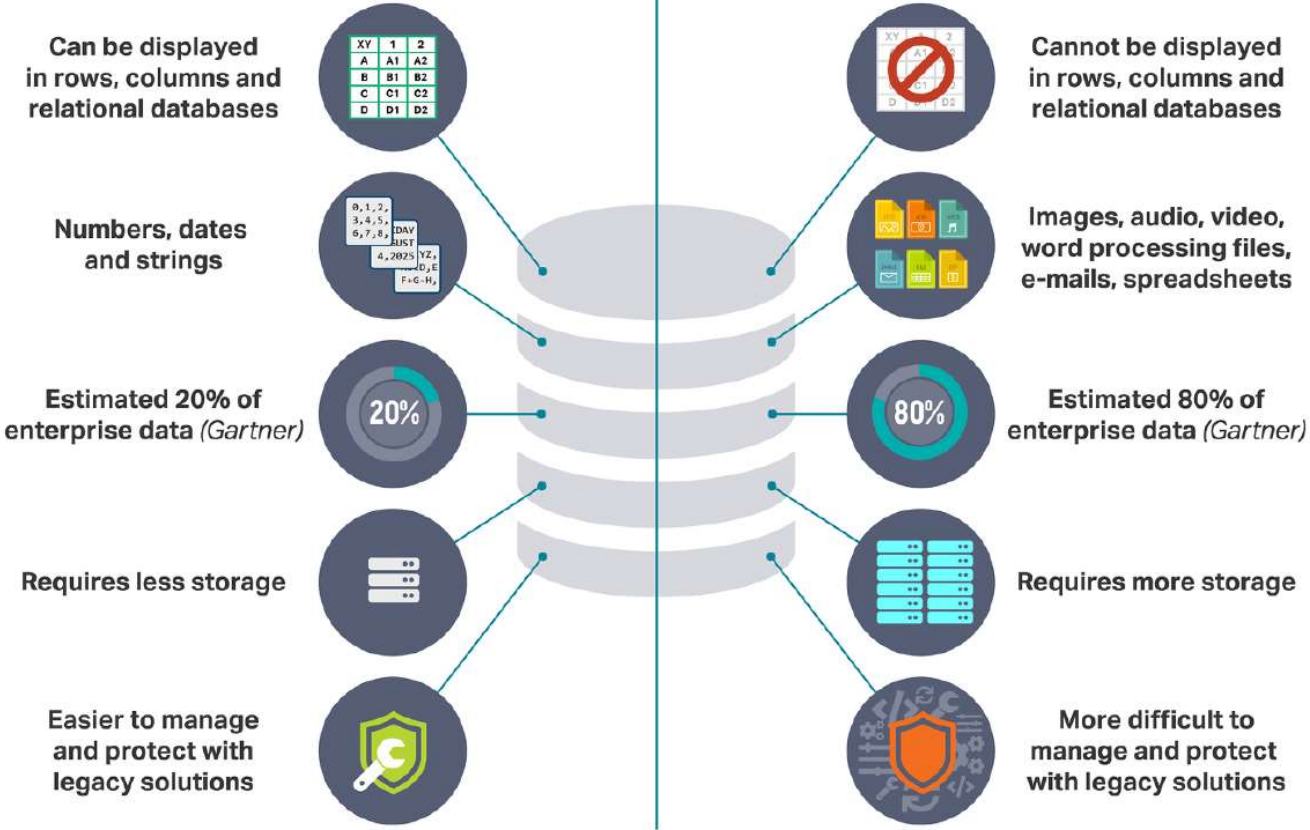
Structured data

ID	Name	Age	Degree
1	John	18	B.Sc.
2	David	31	Ph.D.
3	Robert	51	Ph.D.
4	Rick	26	M.Sc.
5	Michael	19	B.Sc.

Structured Data

VS

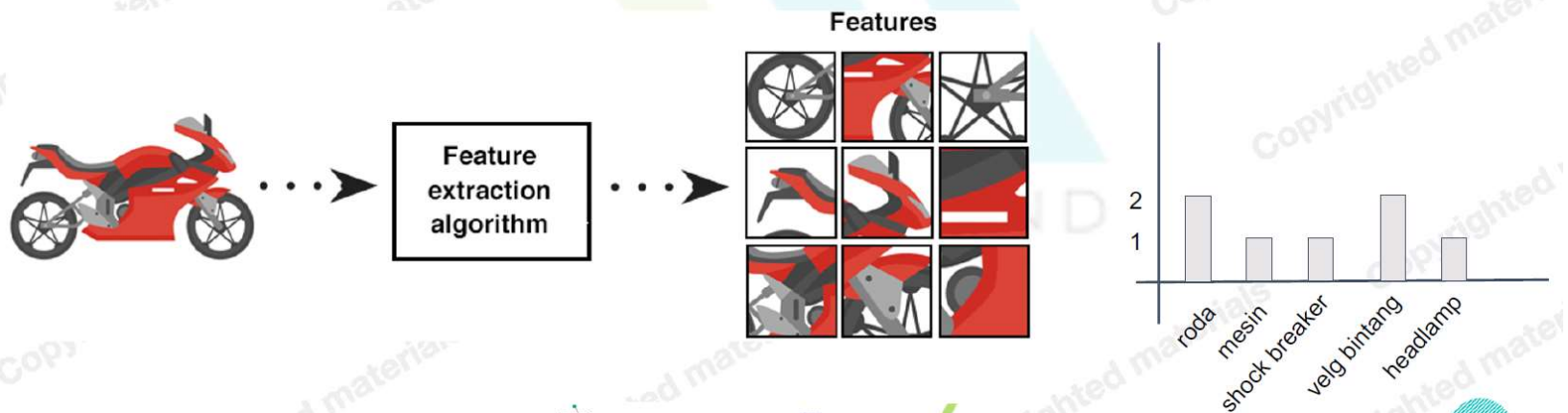
Unstructured Data



Persiapan Data untuk Algoritma ML

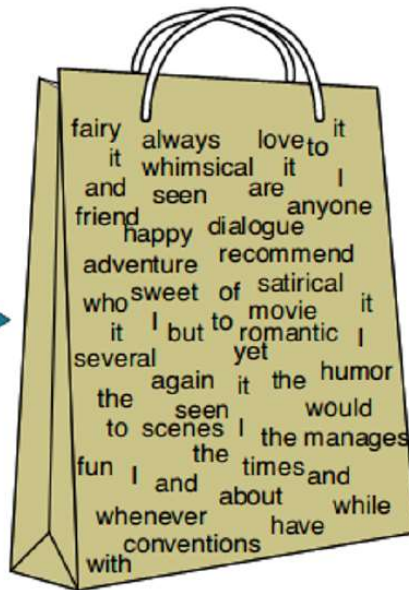
Ekstraksi fitur

- fitur → ciri khusus
- dapat mewakili data asli
- dapat dinyatakan dengan angka

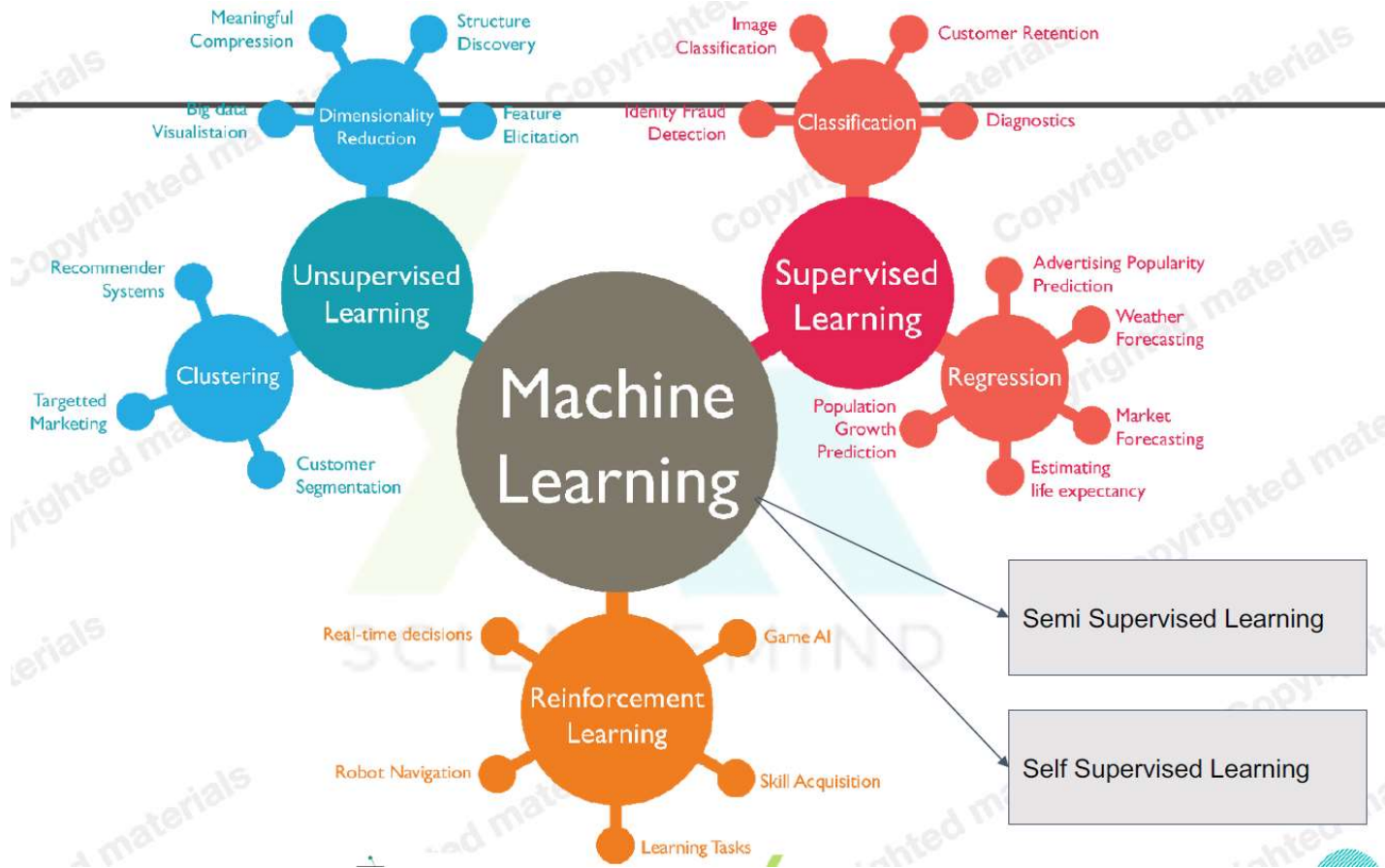


Contoh - Bag of Words

I love this movie! It's sweet, but with satirical humor. The dialogue is great and the adventure scenes are fun... It manages to be whimsical and romantic while laughing at the conventions of the fairy tale genre. I would recommend it to just about anyone. I've seen it several times, and I'm always happy to see it again whenever I have a friend who hasn't seen it yet!



it	6
I	5
the	4
to	3
and	3
seen	2
yet	1
would	1
whimsical	1
times	1
sweet	1
satirical	1
adventure	1
genre	1
fairy	1
humor	1
have	1
great	1
...	...



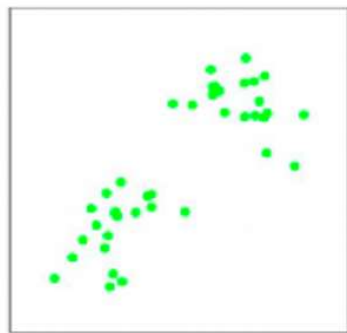
Unsupervised Learning

- Tidak memerlukan label pada proses pelatihan
- Keutamaan:
 - Menemukan pola dari sekumpulan data
 - Membantu menemukan fitur yang berguna
 - Data lebih mudah didapat

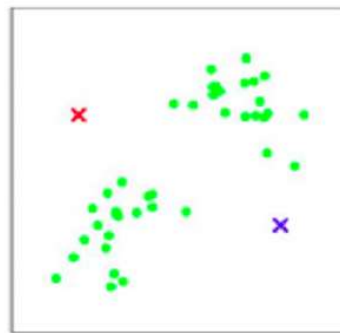
Clustering

- Sampel yang mirip digabungkan ke kelompok yang sama
- Metode yang umum digunakan:
 - K-Means
 - Meanshift
 - DBSCAN
 - Gaussian Mixture Model
 - Agglomerative Hierarchical Clustering

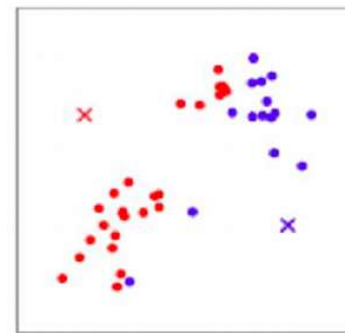
K-means Clustering



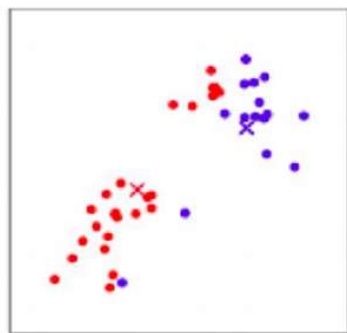
(a)



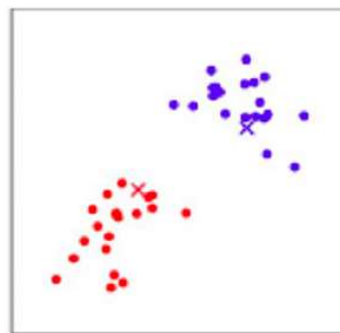
(b)



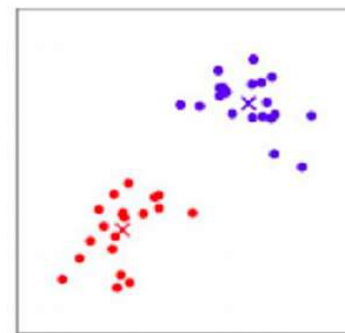
(c)



(d)



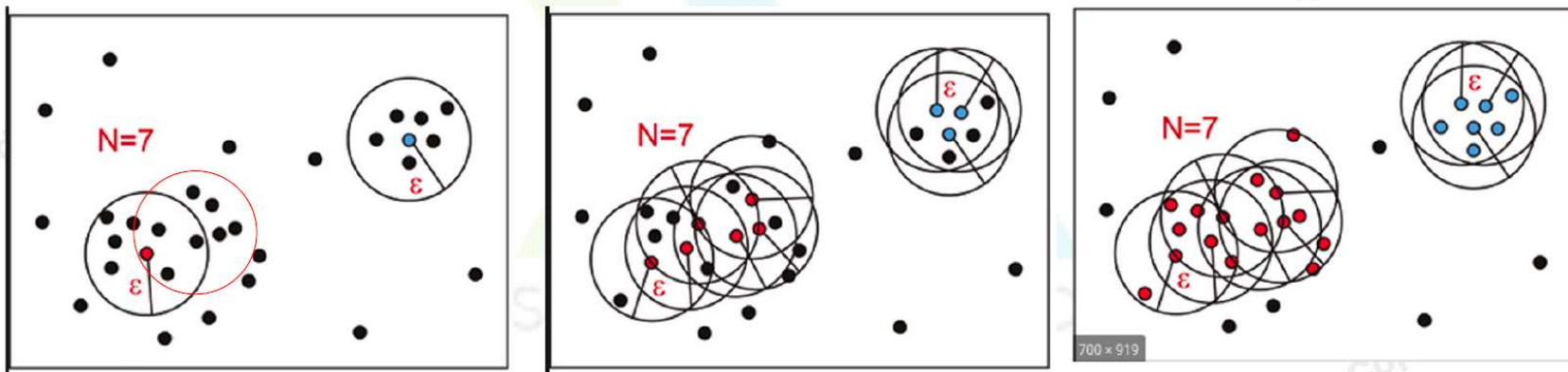
(e)



(f)

DBSCAN

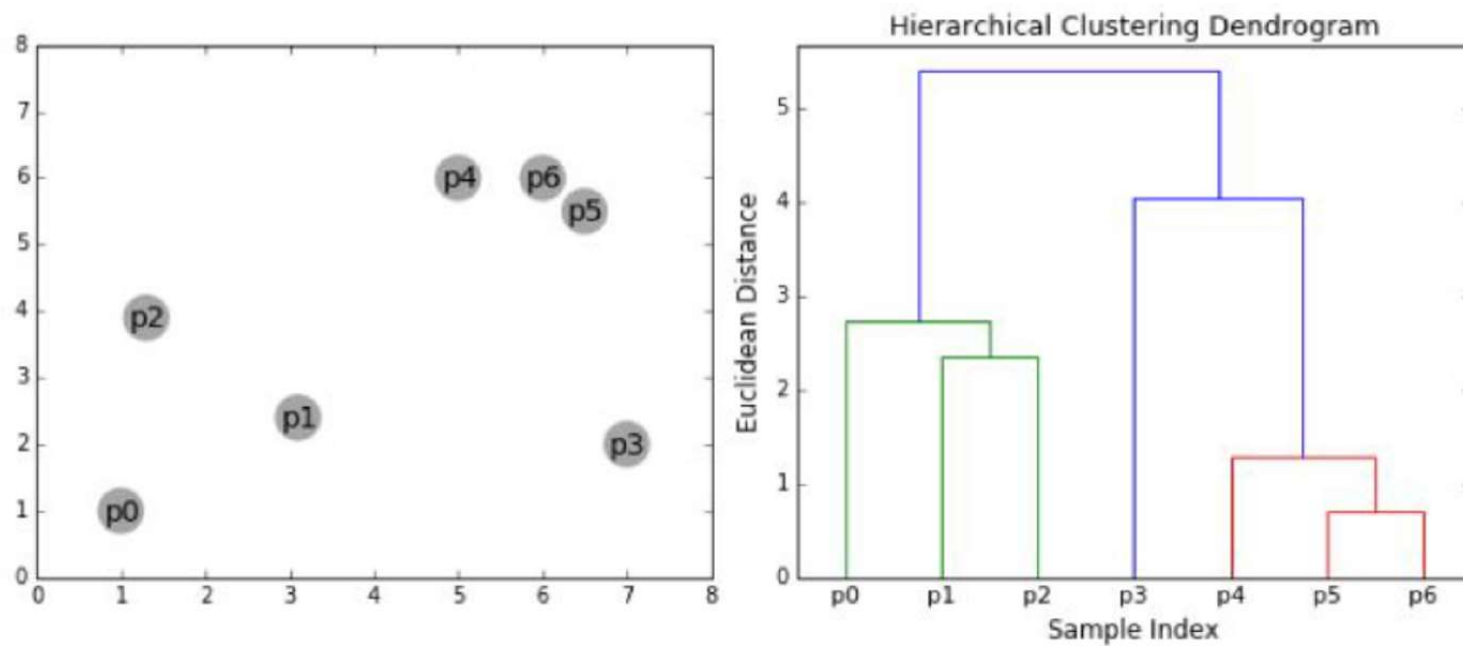
Density-Based Spatial Clustering of Application with Noise



Gaussian Mixture Model (GMM)



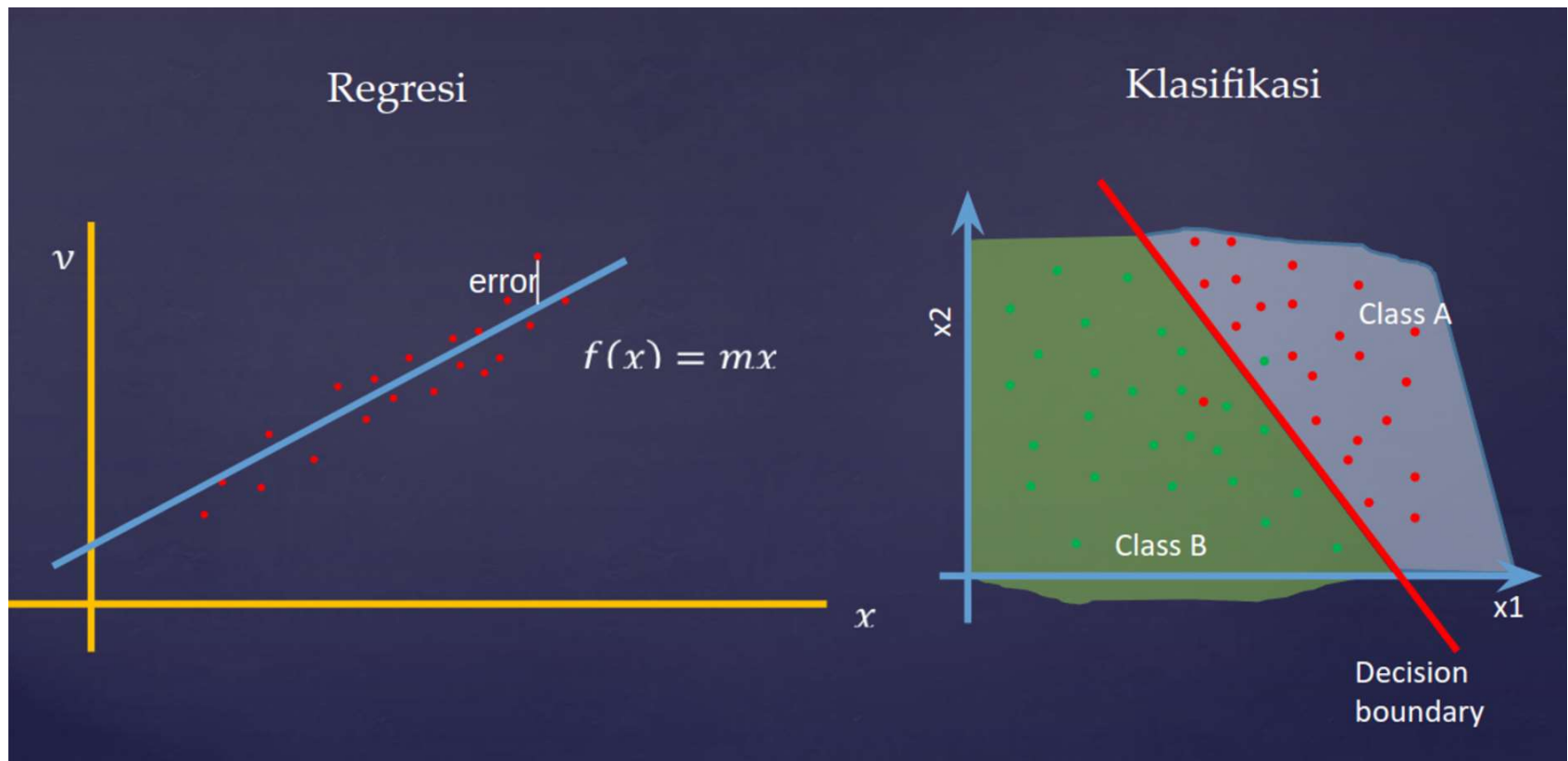
Agglomerative Hierarchical Clustering



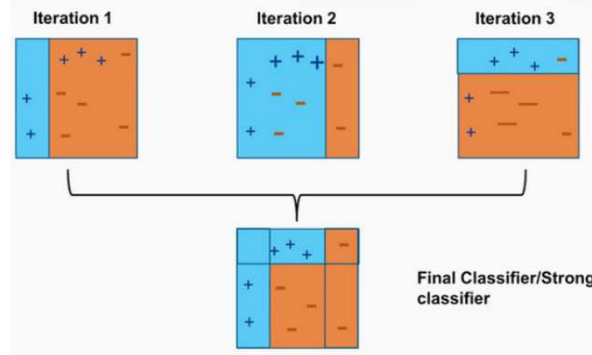
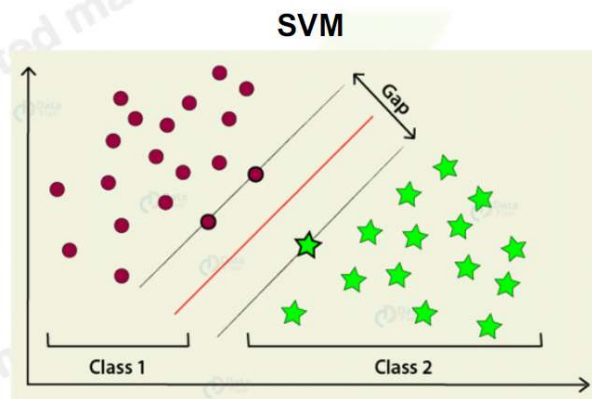
Supervised Learning



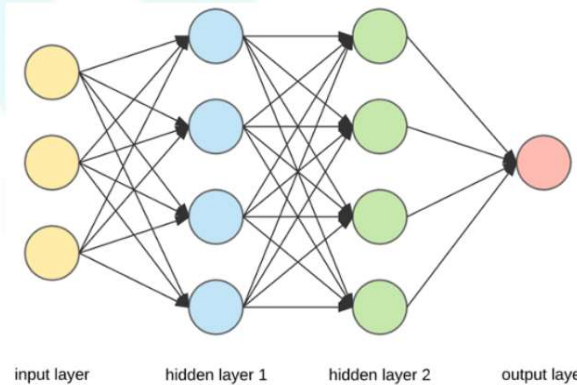
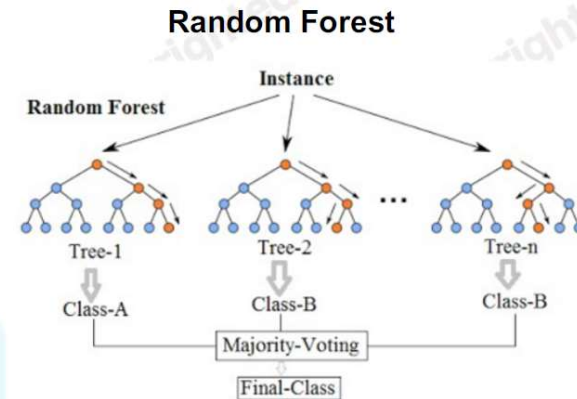
Supervised Learning



Beberapa Metode Populer



AdaBoost



Neural Network