



Institut Informatika & Bisnis  
**DARMAJAYA**  
Yayasan Alfian Husin



**Kampus  
Merdeka**  
INDONESIA JAYA

**MERDEKA  
BELAJAR**

DATA SCIENCE DARMAJAYA  
“YOUR BEST FUTURE IN DATA”

PERTEMUAN KE: 6

# **SUPERVISED LEARNING KLASIFIKASI & REGRESI**

**KULIAH**

OLEH: NURJOKO



# Learning Objectives

- Learning Objective 1
- Learning Objective 2
- Learning Objective 3
- Learning Objective 4
- Learning Objective 5
- Learning Objective 6
- Learning Objective 7
- Learning Objective 8
- Learning Objective 9
- Learning Objective 10



# Pendahuluan

- Klasifikasi dalam Supervised Learning merujuk pada jenis masalah di mana model machine learning atau algoritma bertugas untuk mengkategorikan atau mengelompokkan data input ke dalam kategori atau label yang telah ditentukan sebelumnya.
- Dalam masalah klasifikasi, setiap sampel data memiliki label atau kategori target yang sesuai, dan tujuannya adalah untuk memprediksi label atau kategori tersebut untuk data yang belum pernah dilihat sebelumnya.

## Contoh-contoh masalah klasifikasi meliputi:

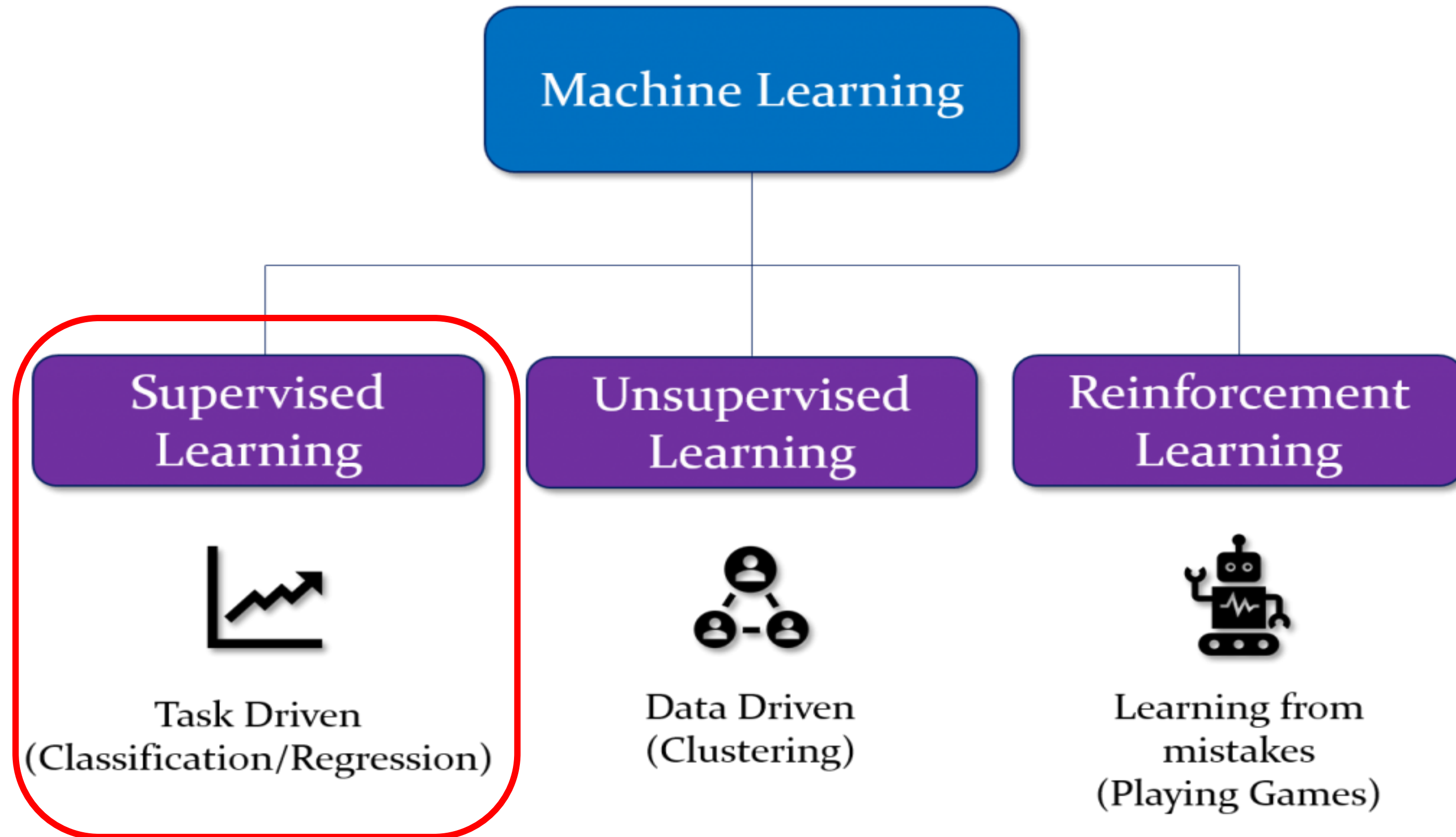
- Klasifikasi email sebagai "spam" atau "bukan spam."
- Identifikasi gambar sebagai "kucing," "anjing," atau "burung."
- Penentuan apakah pasien memiliki penyakit tertentu atau tidak berdasarkan data medis.
- Klasifikasi produk dalam kategori seperti "pakaian," "elektronik," atau "makanan."

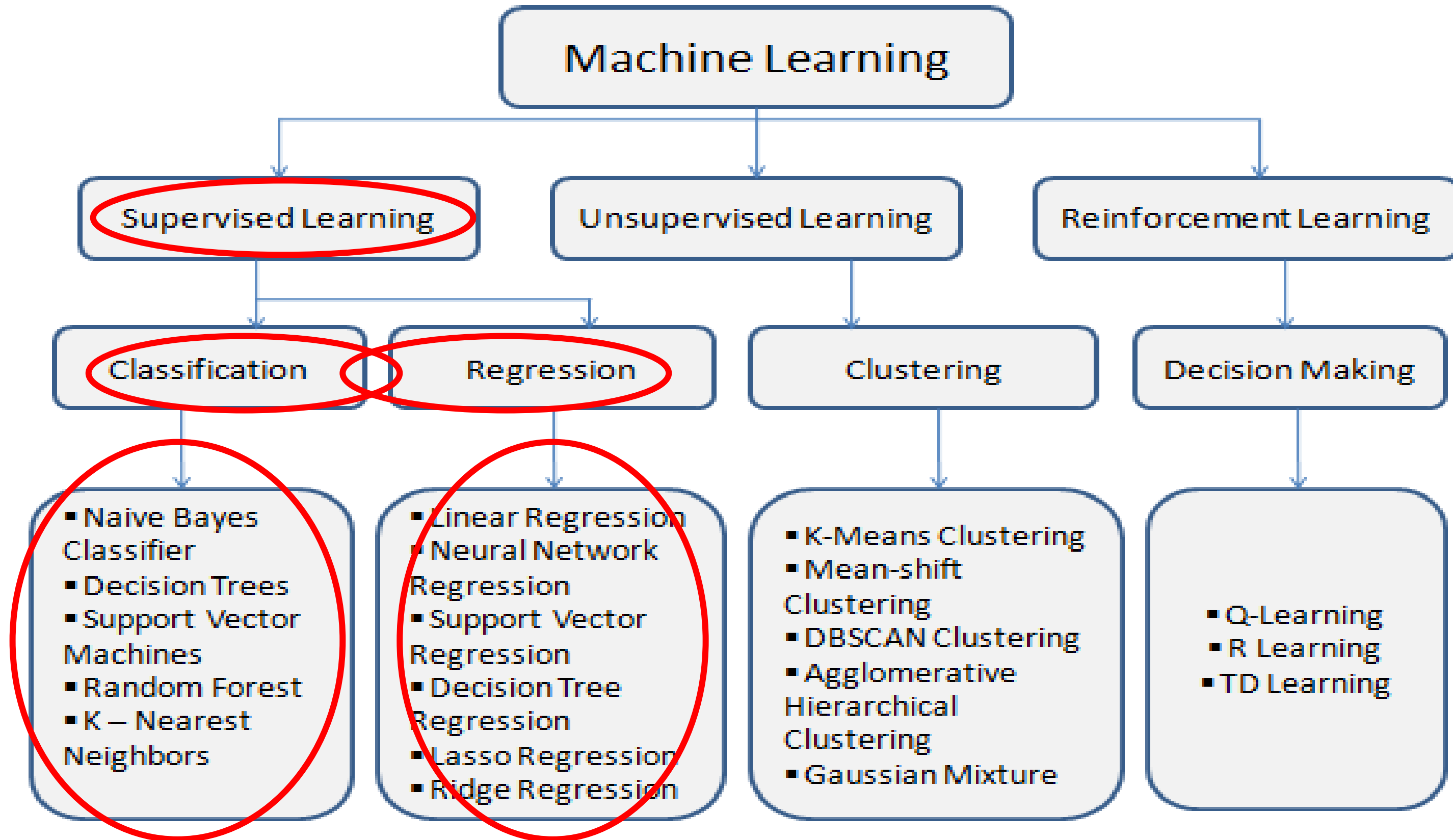
## Peran Algoritma Klasifikasi dalam Analisis Data

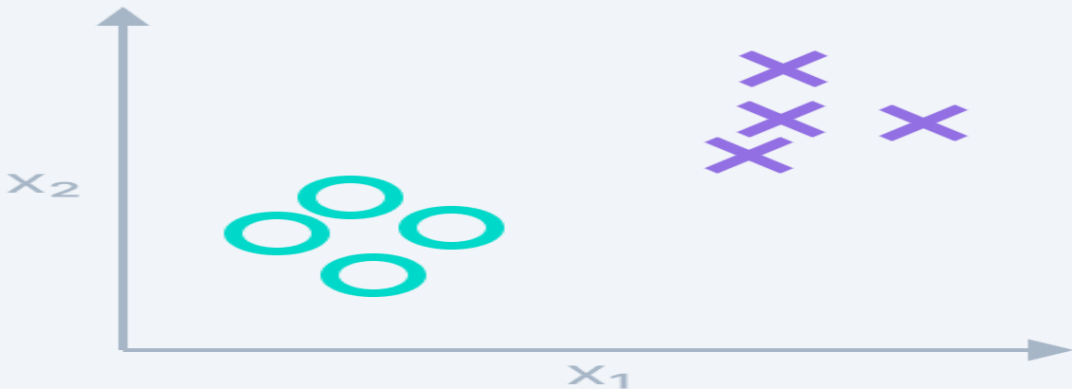
Algoritma klasifikasi dalam analisis data berperan untuk mengkategorikan atau mengelompokkan data ke dalam label atau kategori yang relevan, yang membantu dalam pemahaman, pengambilan keputusan, dan memecahkan berbagai masalah seperti:

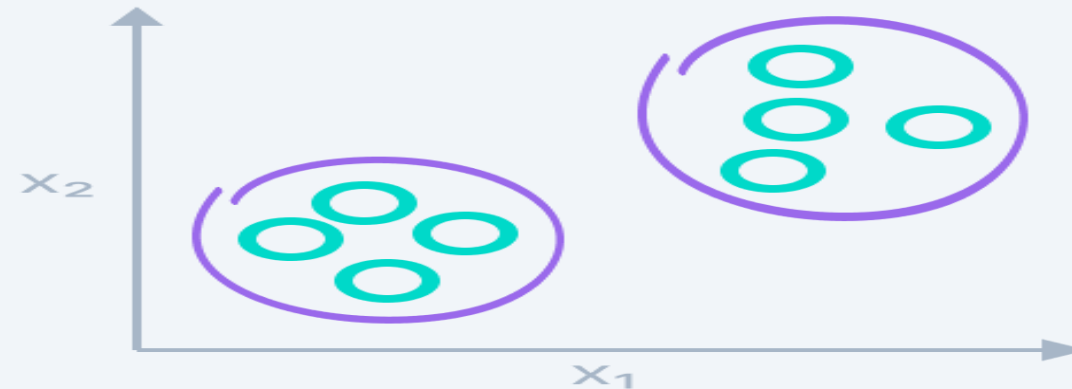
- Pengklasifikasi Data
- Prediksi dan Generalisasi
- Pengambilan Keputusan
- Identifikasi Pola yang Penting
- Pengelompokan Data
- Deteksi Anomali
- Optimasi Proses Bisnis
- Pemahaman Lebih Mendalam

# Types of Machine Learning

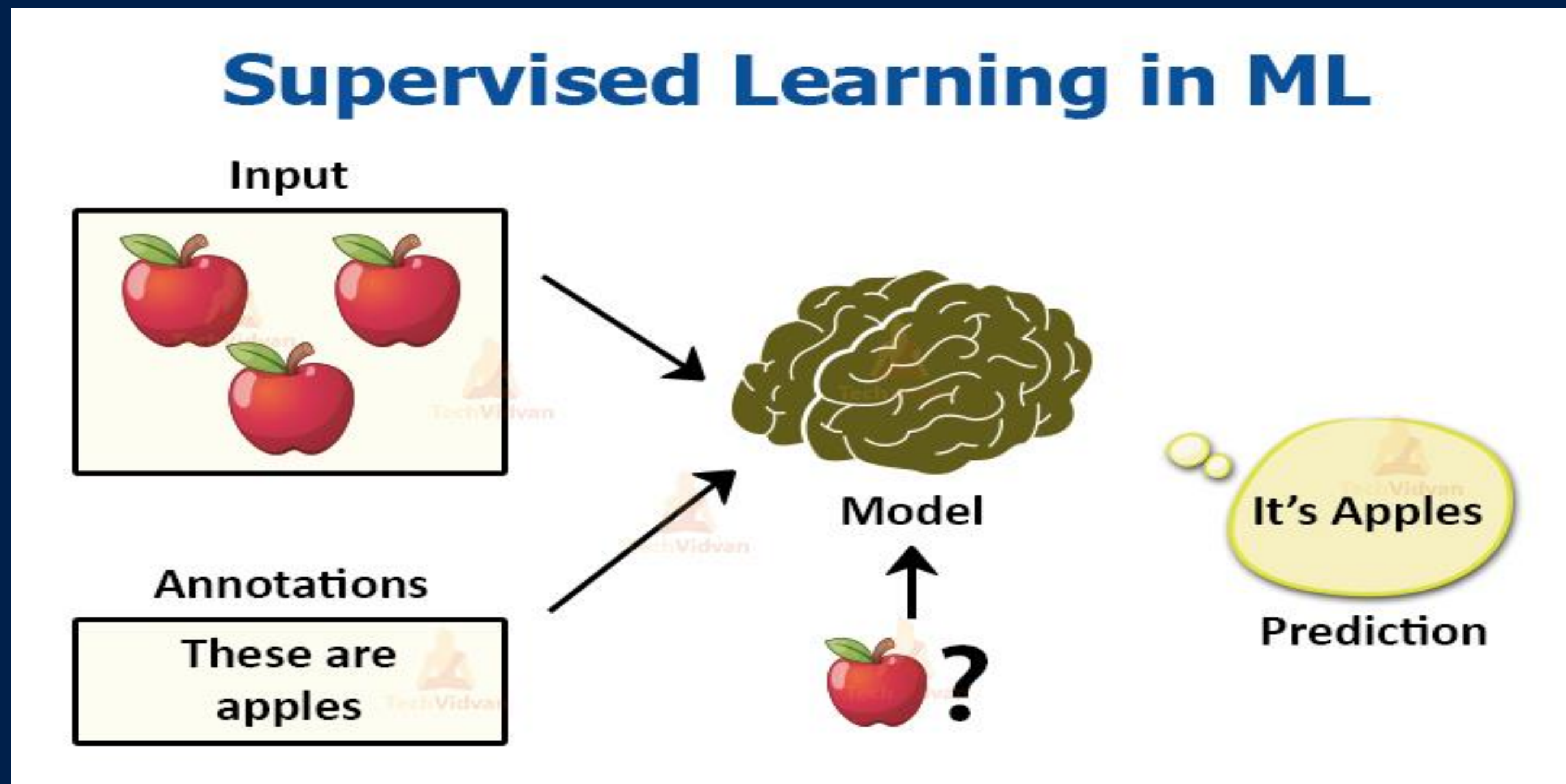


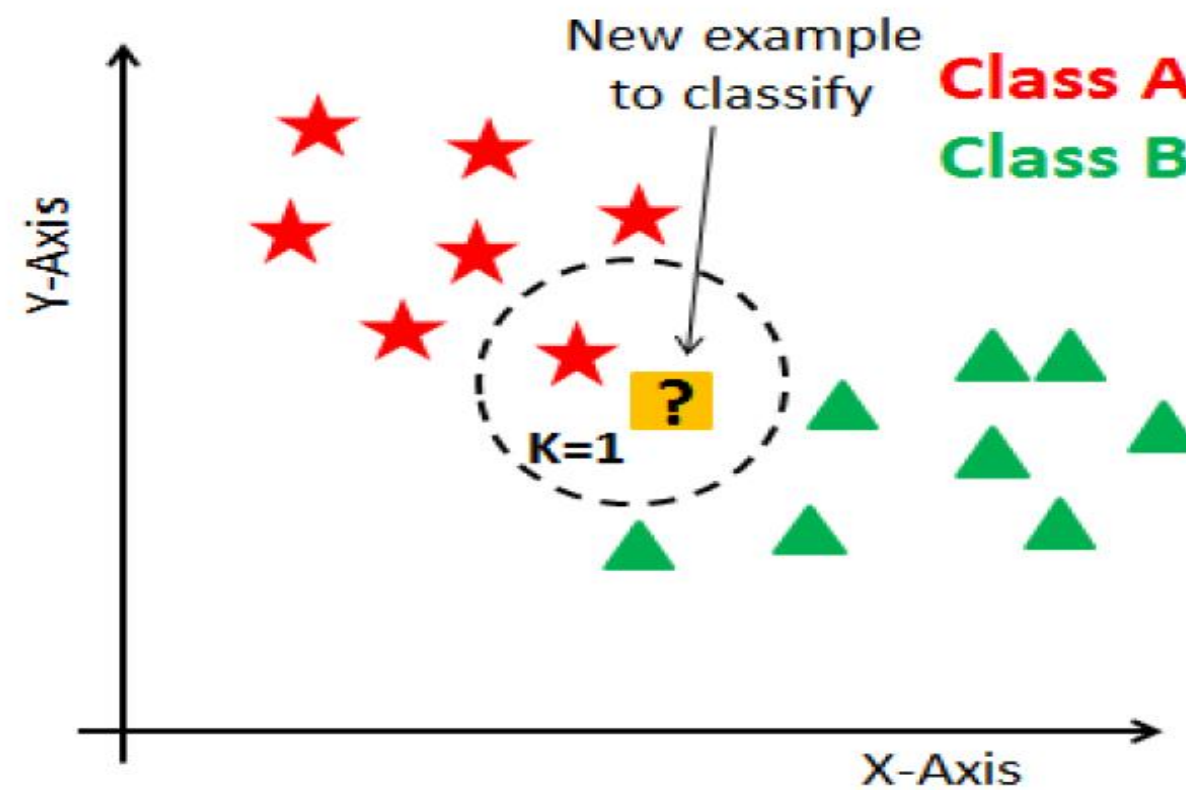


Supervised learning
Input data is labeled
Has a feedback mechanism
Data is classified based on the training dataset
Divided into Regression & Classification
Used for prediction
Algorithms include: decision trees, logistic regressions, support vector machine
A known number of classes


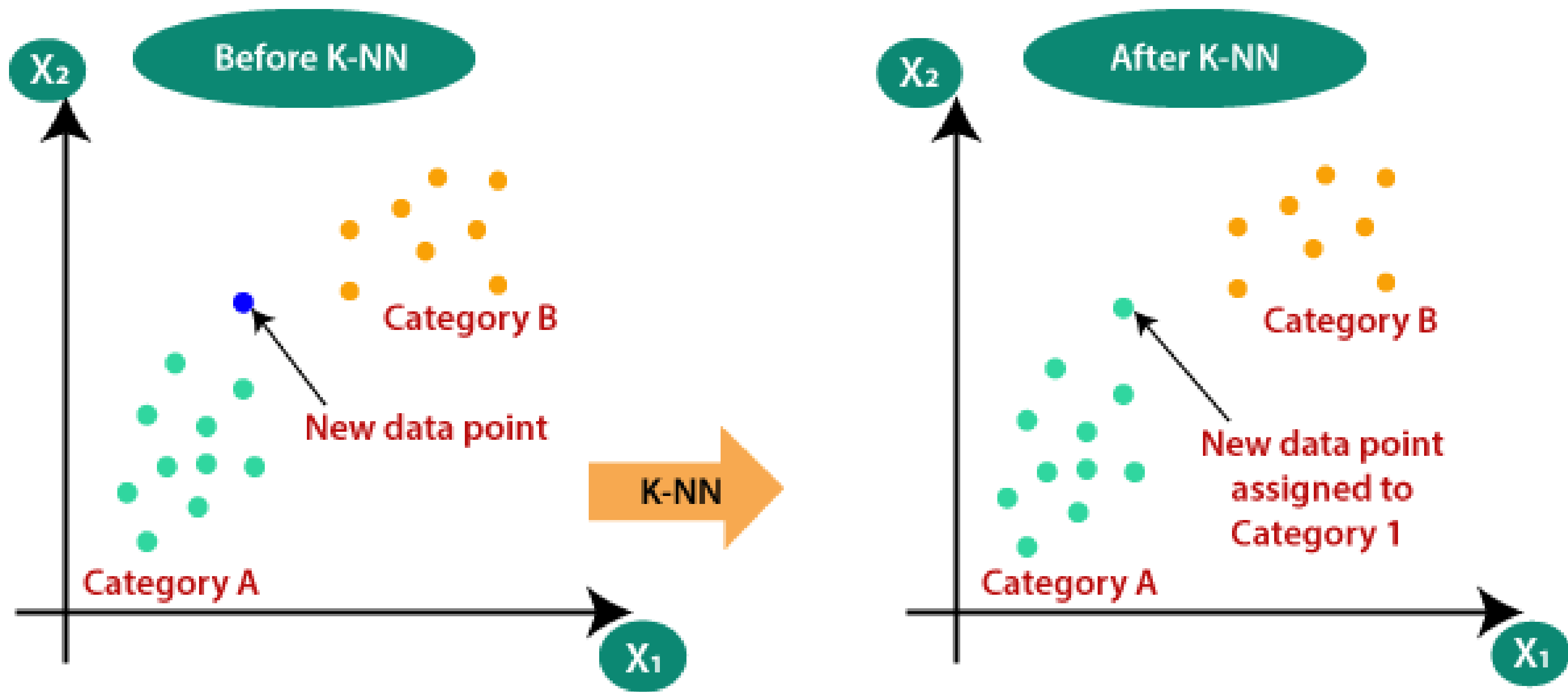
Unsupervised learning
Input data is unlabeled
Has no feedback mechanism
Assigns properties of given data to classify it
Divided into Clustering & Association
Used for analysis
Algorithms include: k-means clustering, hierarchical clustering, apriori algorithm
A unknown number of classes


# Supervised Learning (Klasifikasi)





# K-Nearest Neighbor (KNN)





Institut Informatika & Bisnis  
**DARMAJAYA**  
Yayasan Alfian Husin



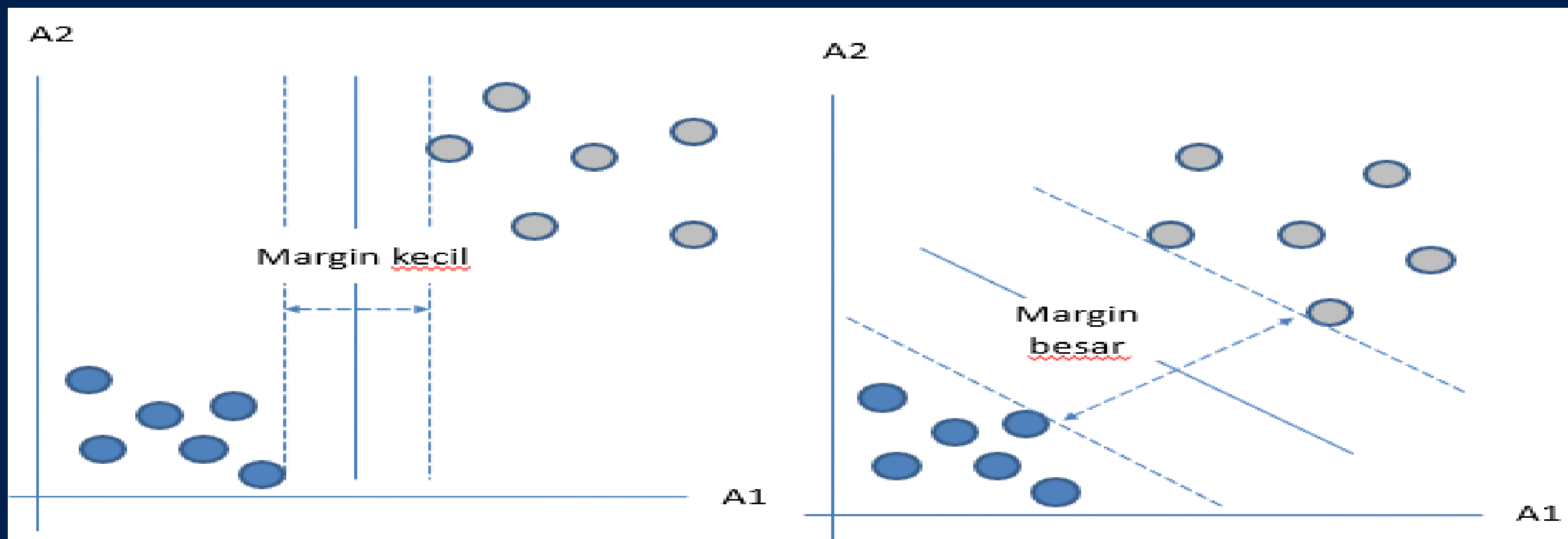
**Kampus  
Merdeka**  
INDONESIA JAYA

**MERDEKA  
BELAJAR**

# 1. Support Vector Machine (SVM)

- SVM merupakan metode klasifikasi jenis terpadu (supervised) karena ketika proses pelatihan diperlukan target pembelajaran tertentu, berbeda dengan FCM yang dalam proses klasifikasi tidak membutuhkan target pelatihan.
- SVM merupakan algoritma yang bekerja menggunakan pemetaan nonlinier untuk mengubah data pelatihan asli ke dimensi yang lebih tinggi.
- Dalam hal ini dimensi baru, akan mencari hyperplane untuk memisahkan secara linier.
- SVM menemukan hyperplane ini menggunakan support vector dan margin.

- Waktu pelatihan, SVM kebanyakan lambat tetapi sangat akurat karena kemampuan untuk menangani model-model nonlinier yang kompleks.



Margin Kecil dan Besar

- Secara intuitif, hyperplane dengan margin yang lebih besar lebih akurat dalam mengklasifikasikan data dibanding margin yang lebih kecil.

- Rumus hyperplane :

$$W \cdot X + b = 0 \rightarrow W_0 + W_1X_1 + W_2X_2 = 0$$

Dimana  $W = \{w_1, w_2, \dots, w_n\} \rightarrow$  vector bobot

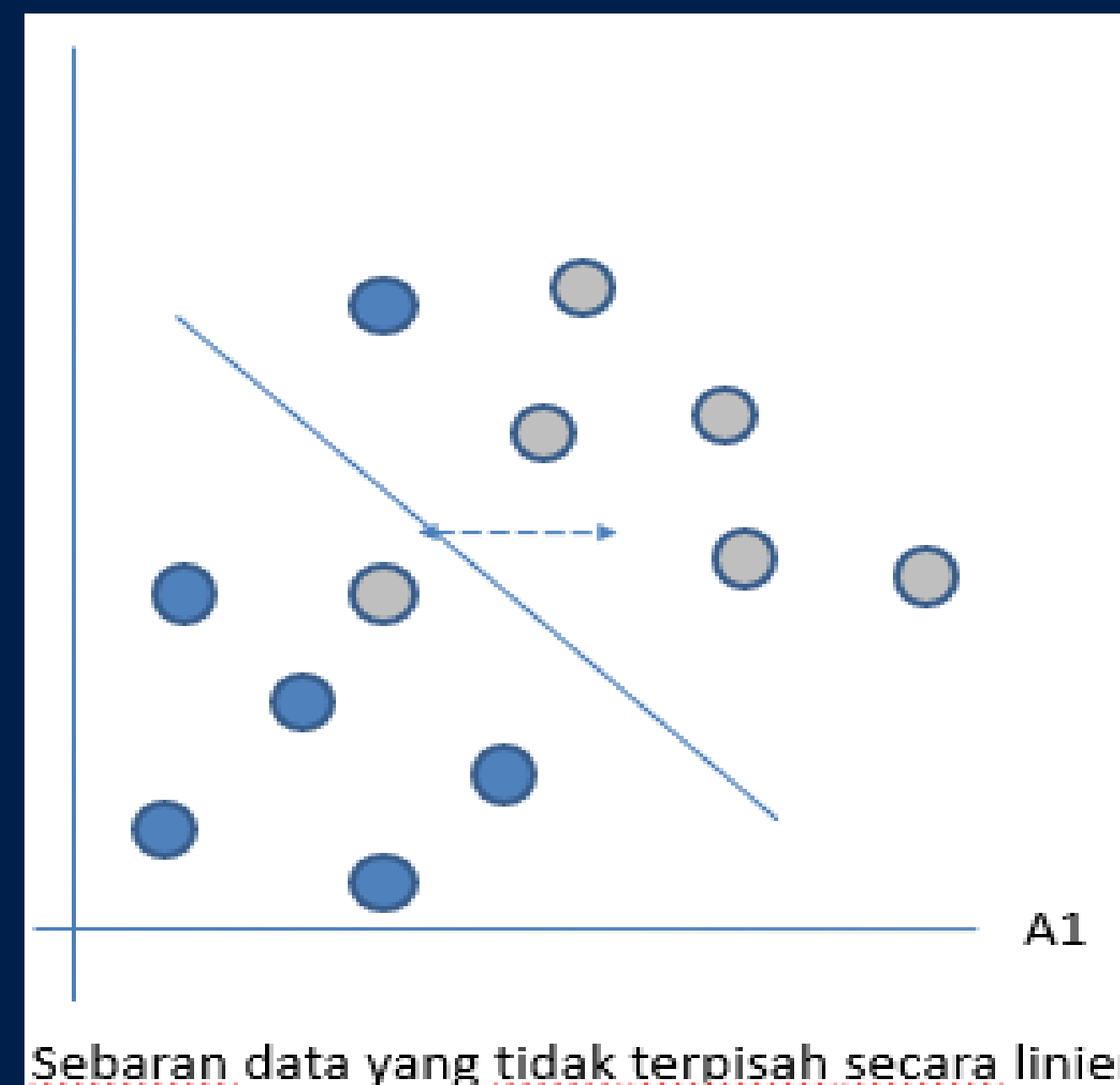
$n$  = jumlah atribut

$b$  = skalar yang digunakan sebagai bias

$X$  = nilai atribut

## Kasus Data yang tidak terpisah secara linier

- Dalam klasifikasi terkadang dijumpai bidang pemisah yang tidak bisa diambil dengan satu garis lurus.
- Oleh karena itu dilakukan pendekatan dengan SVM nonlinier dapat diselesaikan dengan SVM linier.





## Contoh Kasus Kelas Biner

- Terdapat data nilai siswa beserta kelas-kelas yang dimasukinya, yaitu 0=IPA dan 1= IPS.

No	<u>Nilai IPA</u>	<u>Nilai IPS</u>	<u>Jurusan</u>
1.	8	8	0
2.	7	8	1
3.	7	7	0
4.	8	9	1
5.	6	7	1
6.	7	9	1
7.	7.5	8	1
8.	7	7.5	1

- Buat dua file berekstensi dat yaitu file data\_svm dan group\_svm. Sbb:

```
8      8
7      8
7      7
8      9
6      7
7      9
7.5    8
7      7.5
```

```
0
1
0
1
1
1
1
1
1
1
1
```

- Pada command window matlab ketik :  
load data\_svm.dat  
load group\_svm.dat, sbb:

```
>> load data_svm.dat
>> load group_svm.dat
>> data_svm

data_svm =

    8.0000    8.0000
    7.0000    8.0000
    7.0000    7.0000
    8.0000    9.0000
    6.0000    7.0000
    7.0000    9.0000
    7.5000    8.0000
    7.0000    7.5000

>> group_svm

group_svm =

    0
    1
    0
    1
    1
    1
    1
    1
    1
```

- Jika `data_svm` dan `group_svm` telah berhasil dibuat aka langkah berikutnya membuat persamaan pemisah dengan ketik :

**`svmStruct=`**

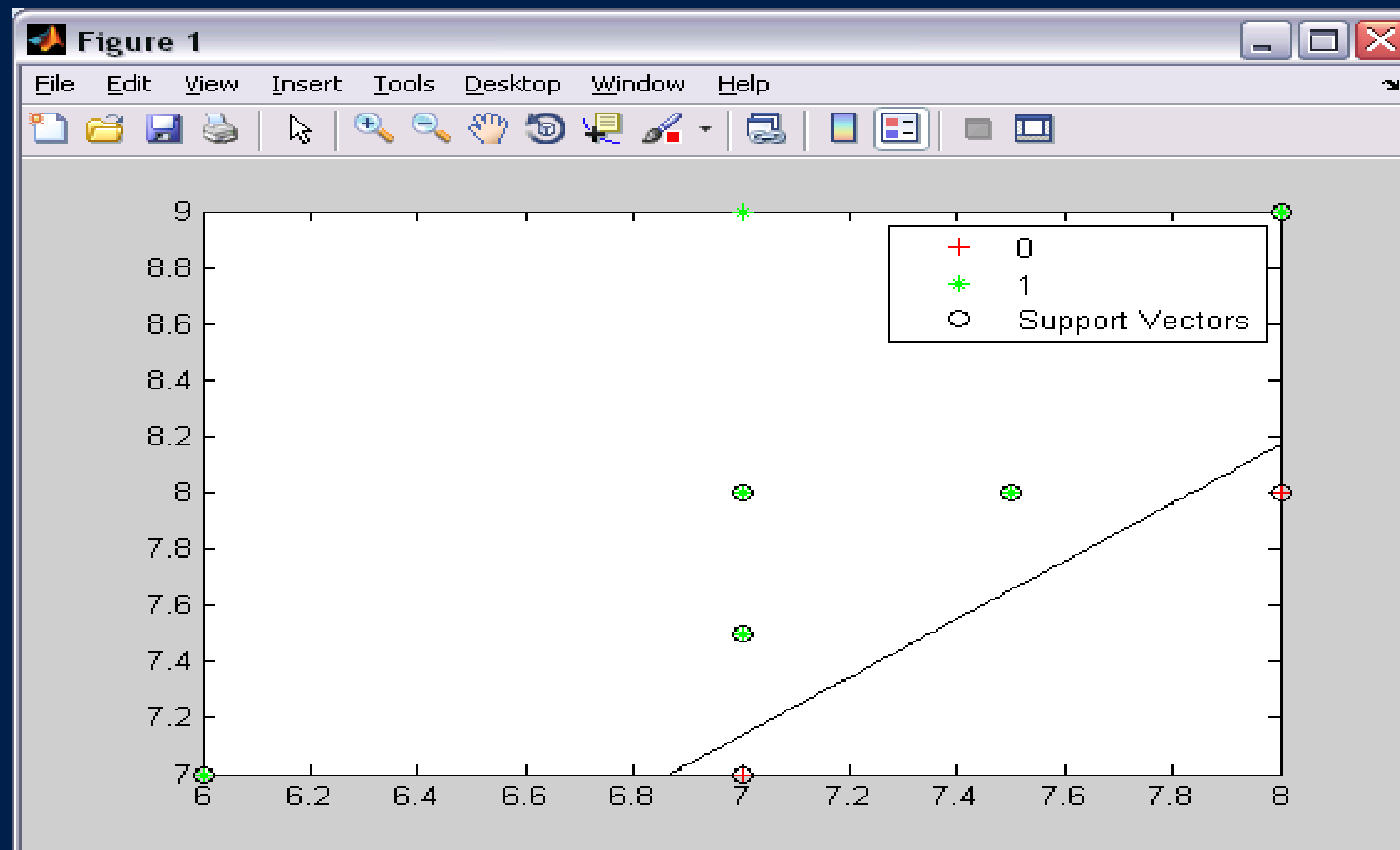
**`svmtrain(data_svm,group_svm,'showplot',true);`**

kemudian simpan , dengan ketik:

save kelassvm.mat, seperti sbb:

```
>> svmStruct = svmtrain(data_svm,group_svm,'showplot',true);  
  
>> save kelassvm.mat
```

- Hasil klasifikasi, sbb :



- Kemudian diuji data training (data\_svm.dat), dengan ketik: `svmclassify(svmStruct,data)`, sehingga hasilnya akan sama dengan yang ada di `group_svm.dat`, untuk mengetahui hal tersebut silakan dengan data lain yang tidak ada pada `data_svm.dat`, misalkan:

```
>> svmclassify(svmStruct,data_svm)

ans =

     0
     1
     0
     1
     1
     1
     1
     1
```

```
>> siswa=[8 7.5]

siswa =

     8.0000     7.5000

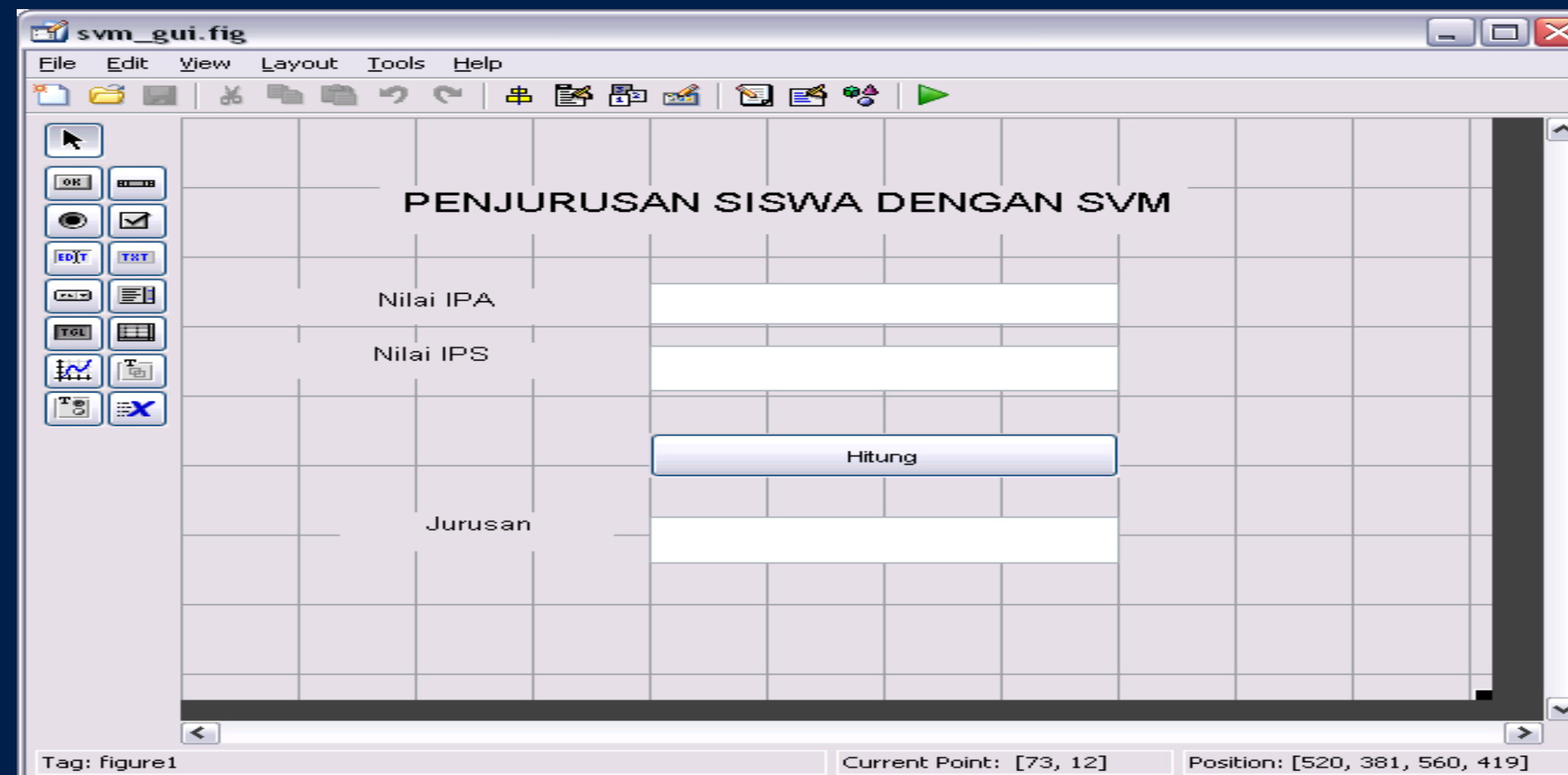
>> svmclassify(svmStruct,siswa)

ans =

     0
```

# Aplikasi Dengan GUI

- Buka matlab, klik FILE-NEW-GUI, kemudian desain sbb:



```
1 function ipa_Callback(hObject, eventdata, handles)
2     ipa=str2double(get(hObject,'String'))
3     handles.ipa=ipa;
4     guidata(hObject,handles)
5
6 function ips_Callback(hObject, eventdata, handles)
7     ips=str2double(get(hObject,'String'))
8     handles.ips=ips;
9     guidata(hObject,handles)
10
11 function hitung_Callback(hObject, eventdata, handles)
12     a=handles.ipa;
13     b=handles.ips;
14     load kelassvm.mat
15     siswa2=[a b]
16     tahap1=svmclassify(svmStruct,siswa2)
17     if tahap1==0
18         set(handles.jurusan,'String','IPA');
19     else
20         set(handles.jurusan,'String','IPS');
21     end
```

svm\_gui

### PENJURUSAN SISWA DENGAN SVM

Nilai IPA:

Nilai IPS:

Jurusan:



# Regresi

**Regresi** adalah teknik statistik yang digunakan untuk memahami dan memodelkan hubungan antara satu atau lebih variabel independen (juga dikenal sebagai variabel prediktor atau fitur) dan variabel dependen (juga dikenal sebagai variabel respons atau target).

Tujuan utama regresi adalah memprediksi atau menjelaskan variabel dependen berdasarkan variabel independen.

## **1. Variabel Independen (Fitur):**

1. Variabel independen adalah variabel atau atribut yang digunakan sebagai dasar untuk melakukan prediksi. Dalam regresi, kita dapat memiliki satu atau lebih variabel independen.

## **2. Variabel Dependen (Target):**

1. Variabel dependen adalah variabel yang ingin diprediksi atau dijelaskan. Itu adalah hasil yang ingin kita pahami berdasarkan variabel independen.

## **3. Hubungan Linier:**

1. Dalam regresi linear, asumsi dasar adalah adanya hubungan linier antara variabel independen dan variabel dependen. Ini berarti perubahan dalam variabel independen secara proporsional mempengaruhi variabel dependen.

## 1. Model Regresi:

1. Model regresi adalah ekspresi matematis yang digunakan untuk menggambarkan hubungan antara variabel independen dan variabel dependen. Model ini biasanya mengambil bentuk persamaan linear, seperti regresi linear sederhana ( $y = mx + b$ ) atau persamaan regresi berganda yang lebih kompleks.

## 2. Koefisien Regresi:

1. Koefisien regresi adalah angka-angka yang digunakan dalam model regresi untuk menggambarkan sejauh mana variabel independen mempengaruhi variabel dependen. Dalam regresi linear sederhana, koefisien  $m$  adalah kemiringan garis regresi, dan  $b$  adalah perpotongan dengan sumbu  $y$ .

## 3. MSE (Mean Squared Error):

1. MSE adalah metrik evaluasi yang umum digunakan dalam regresi untuk mengukur seberapa baik model memfitting data. Ini mengukur rata-rata kuadrat perbedaan antara prediksi model dan nilai sebenarnya dari variabel dependen.

## **1. Overfitting dan Underfitting:**

1. Overfitting terjadi ketika model regresi terlalu kompleks dan menyesuaikan dengan noise dalam data pelatihan, yang dapat menghasilkan kinerja yang buruk pada data pengujian. Underfitting terjadi ketika model terlalu sederhana dan tidak dapat menangkap hubungan yang sebenarnya dalam data.

## **2. Regresi Linear Sederhana:**

1. Ini adalah bentuk dasar dari regresi, di mana ada satu variabel independen dan satu variabel dependen. Regresi linear sederhana menghasilkan garis regresi yang mencoba menggambarkan hubungan linear antara kedua variabel.

## **3. Regresi Linear Berganda:**

1. Dalam regresi linear berganda, ada lebih dari satu variabel independen yang digunakan untuk memprediksi variabel dependen. Ini memungkinkan model untuk menangkap lebih banyak kompleksitas dalam data.



## **1. Regresi Non-Linear:**

1. Regresi non-linear digunakan ketika hubungan antara variabel independen dan variabel dependen tidak bersifat linear. Model regresi non-linear dapat mengambil bentuk fungsi non-linear seperti kuadratik, eksponensial, logaritmik, dan sebagainya.

## **2. Validasi Model:**

1. Validasi model adalah langkah penting dalam regresi. Ini melibatkan pengujian seberapa baik model berkinerja pada data yang belum terlihat, seperti data pengujian, untuk memastikan bahwa model tidak hanya memfitting data pelatihan.

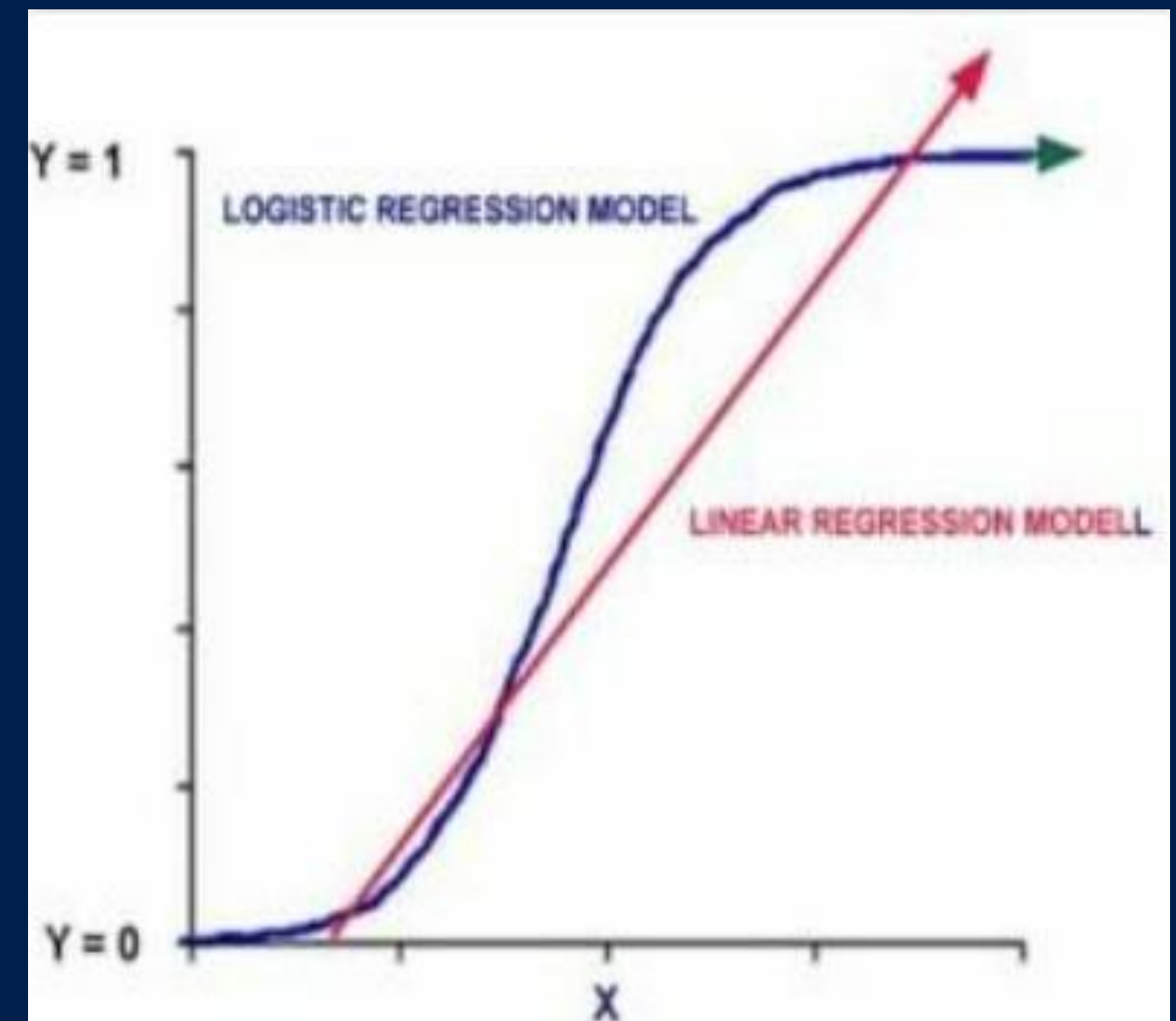
## 2. Regresi Logistik (Logistic Regression)

### Definisi Regresi Logistik

- Regresi logistik (logistic regression) sebenarnya sama dengan analisis regresi berganda, hanya variabel terikatnya (dependent variable) merupakan dummy variabel (0 dan 1).
- Regresi logistik tidak memerlukan asumsi normalitas dan asumsi klasik yang lain, meskipun screening data outliers tetap dapat dilakukan.
- Contoh kasus Regresi Logistik: Ingin diketahui apakah konsumen akan membeli makanan di rumah makan berdasarkan penilaian konsumen terhadap lokasi, pelayanan, pendapatan, kebersihan, selera dan harga. Dalam kasus ini hanya ada 2 kemungkinan respon konsumen, yaitu konsumen **membeli** dan **tidak membeli**.

## Regresi Logistik

- Tidak seperti regresi linier biasa, regresi logistik tidak mengasumsikan hubungan antara variabel independen dan dependen secara linier.
- Regresi logistik merupakan regresi non linier dimana model yang ditentukan akan mengikuti pola kurva linier seperti gambar di samping.



# Model Regresi Logistik

Model yang digunakan pada regresi logistic adalah:

$$\text{Log} \left( \frac{P}{1 - p} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

- Dimana  $p$  adalah kemungkinan bahwa  $Y = 1$ , dan  $X_1, X_2, X_3$  adalah variabel independen, dan  $b$  adalah koefisien regresi.
- Regresi logistik akan membentuk variabel prediktor/respon ( $\log(p/(1-p))$ ) yang merupakan kombinasi linier dari variabel independen.
- Nilai variabel prediktor ini kemudian ditransformasikan menjadi probabilitas dengan fungsi logit



# Jenis Regresi Logistik

Regresi Logistik terdiri dari 2 jenis yaitu :

1. Binary logistic regression Yaitu regresi logistik dimana variabel terikat memiliki 2 kategori
2. Multinomial Logistic Regression Yaitu regresi logistik dimana variabel terikat memiliki lebih dari 2 kategori



# CONCLUSION

Fill in .....



# REFERENCES

Fill in IEEE Style



Institut Informatika & Bisnis  
**DARMAJAYA**  
Yayasan Alfian Husin



**Kampus  
Merdeka**  
INDONESIA JAYA

**MERDEKA  
BELAJAR**

# THANK YOU!!

DATA SCIENCE DARMAJAYA "YOUR BEST FUTURE IN DATA"