

# **SINESTIK II:**

## **Seminar Nasional Teknik Elektro, Sistem Informasi & Teknik Informasi**

Surabaya, 26 Maret 2022





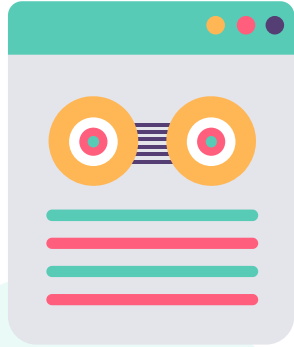
# **AVAN S.Si, M.Si**

**Statistician Modeler Bank Mandiri**  
2017-sekarang

**Peneliti Lembaga Survei Nasional PRC**  
(Politica Research & Consulting)  
2019-sekarang

# Risk Perbankan

**BISNIS**



**GAS**



**REM**



**RISK**



# Scope Project

## Application Scoring (Screening)



## Monitoring

1. Probababilty of Default Debitur
2. Exposure at default Debitur
3. Loss Given Default
4. Makro ekonomi
5. Early Warning Sistem



# Pergeseran Dunia Modeling

## Eksternal Data

1 Makro, vintage dll

## Media Informasi

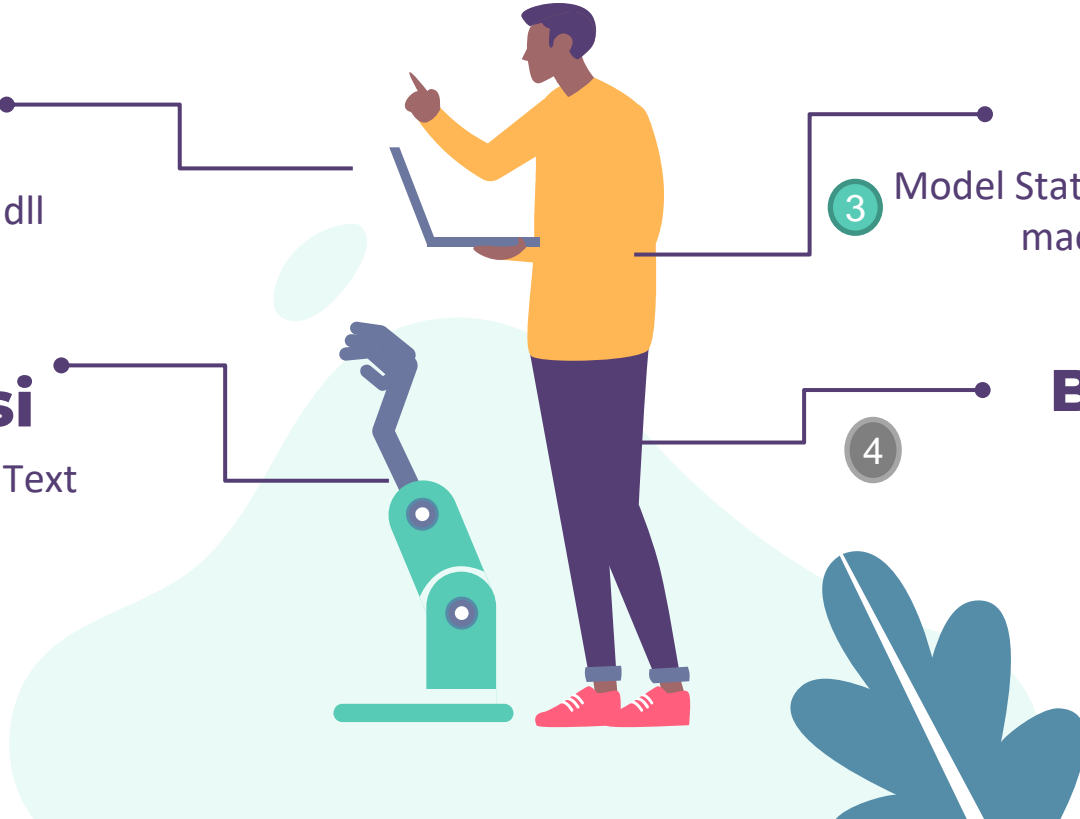
2 Data mining -> Text mining

## Metode

3 Model Statistik -> Model machine learning

## Big data

4



# 1 Eksternal Data

## Internal Data :

- Demografi
- Behavior Payment
- Transaksi

## Kerjasama dgn Eksternal

- Vintage
- Ekspor
- Merchant

## Eksternal Data

- Makro Ekonomi
- Kredit perbankan Lain (Slik)
- Media sosial
- Telkom, Listrik, Ekspor dst

## Terintegrasi



2

# Text Mining

## Background

- Meningkatnya penggunaan social media
- Kebiasaan berbagi informasi diruang public
- Berkembangnya opini publik.
- Kemudahan akses informasi

## Manfaat

- Informasi digunakan untuk pengawasan/penilaian thdp suatu objek
- Dikembangkan metode text mining dan sentiment analisis



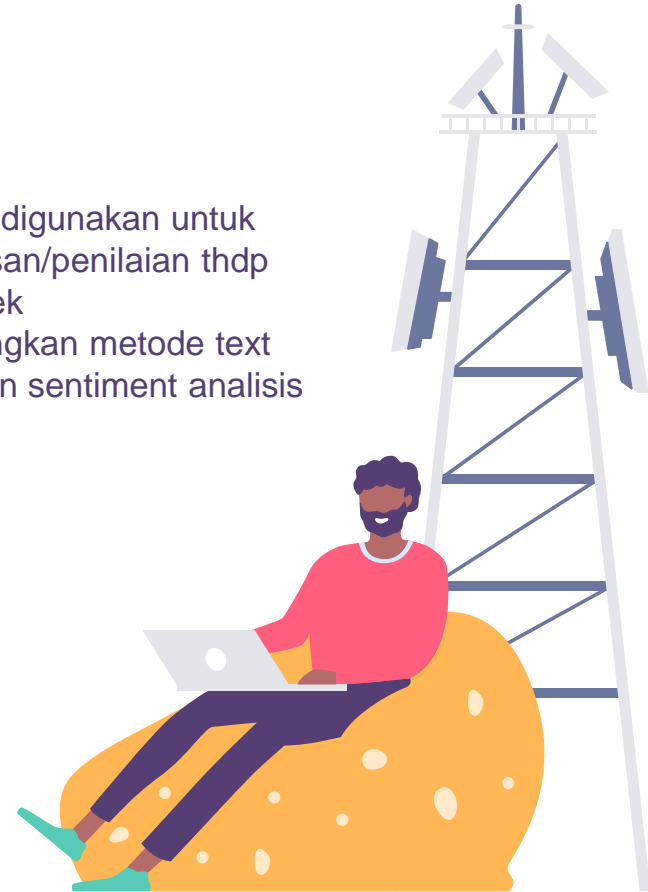
Content di media social hampir seluruhnya text



Informasi tersembunyi di extract dlm text

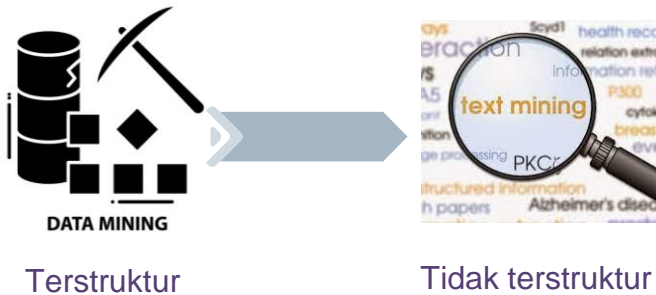


Pengentahuan baru untuk strategi bisnis



# Text Mining

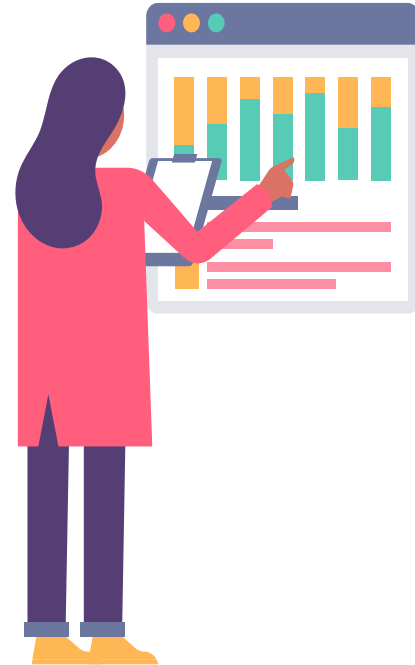
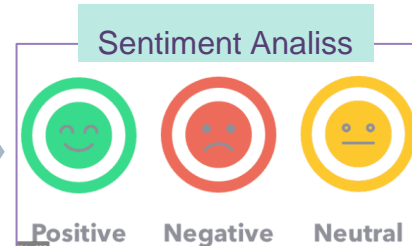
- Text Mining : proses untuk mengambil informasi dari teks yang ada
- mencari pola-pola yang ada di teks-teks dalam Bahasa natural yang **tidak terstruktur** seperti buku, email, artikel, halaman web dll, video, rekaman suara dan gambar,
- banyak digunakan untuk clustering dan Clasification .





# Sentiment Analysis

- Sentiment analysis adalah tahapan lebih lanjut dalam tahap text Mining,
- Teknik untuk mengidentifikasi bagaimana sebuah sentimen diekspresikan menggunakan teks dan dikategorikan dalam sentimen positif maupun sentiment negatif.
- Sebagai adl kamus yang berisi list kata bersentimen positif , negative atau netral



# Kasus Text Mining

IMDb Menu All Search IMDb

## Feature Film, Released between 2018-01-01 and 2018-12-31 (Sorted by Popularity Ascending)

1-100 of 12,484 titles. | Next > View Mode: Compact | Detailed

Sort by: **Popularity** | A-Z | User Rating | Number of Votes | US Box Office | Runtime | Year | Release Date | Date of Your Rating | Your Rating

**1. Venom (2018)** +

13+ | 112 min | Action, Adventure, Sci-Fi

★ 6,7 ☆ Rate this 35 Metascore

A failed reporter is bonded to an alien entity, one of many symbiotes who have invaded Earth. But the being takes a liking to Earth and decides to protect it.

Director: Ruben Fleischer | Stars: Tom Hardy, Michelle Williams, Riz Ahmed, Scott Haze

Votes: 408.393 | Gross: \$213.52M

**2. Den skyldige (2018)** +

16+ | 85 min | Crime, Drama, Thriller

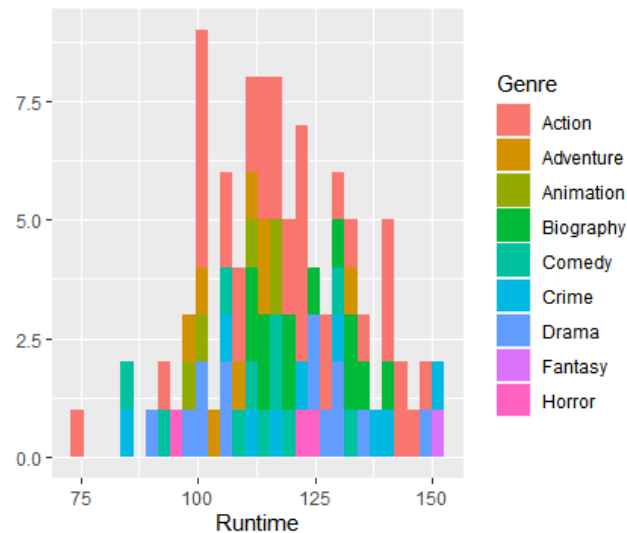
★ 7,5 ☆ Rate this 83 Metascore

A police officer assigned alarm dispatch duty enters a race against time when he answers an emergency call from a kidnapped woman.

Director: Gustav Möller | Stars: Jakob Cedergren, Jessica Dinnage, Omar Shargawi, Johan Gotthardt Olsen

	Runtime	Genre	Rating	Gross_Pendapatan
1	112	Action	6.7	213.52
2	85	Crime	7.5	0.21
3	134	Biography	7.9	216.43
4	106	Crime	6.5	159.34
5	119	Action	7.7	324.59
6	127	Drama	7.3	44.07
7	149	Action	8.4	678.82
8	117	Animation	8.4	190.24
9	90	Drama	7.5	188.02
10	140	Action	7.4	137.69
11	134	Action	7.3	700.06

wing 1 to 12 of 100 entries, 4 total columns

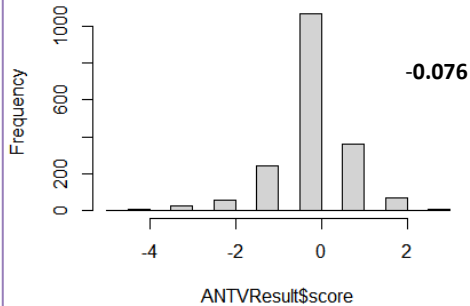


# Kasus Sentimen Analisis

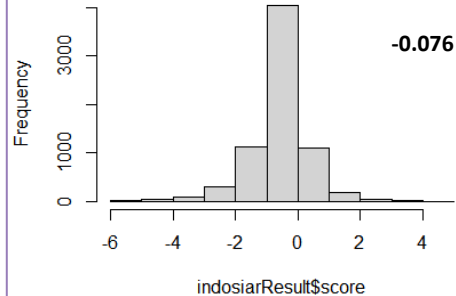
```
###ambil tweet  
some_tweets <- searchTwitter('indosiar|', n=8000)
```



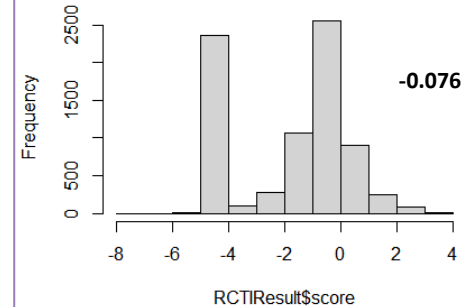
Histogram of ANTVResult\$score



Histogram of indosiarResult\$score



Histogram of RCTIResult\$score



- Melihat opini public terhadap 3 stasiun TV
- Analisa berdasarkan FB, Twitter, IG dan beberapa portal berita

# 3 Model Machine Learning(ML)

**Model Statistika:** Formulasi hubungan antar variabel dalam bentuk persamaan matematis

**Model Mahine Learning** Algoritma yang belajar dari data tanpa bergantung pada persamaan standard

## Kriteria

- Data
- Kompleksitas
- Interpretasi
- Statistika
- Konstrain
  
- Performa
- Orientasi
- Modeler

## Statistika

- ● data kecil perhitungan terbatas
- ● model sederhana
- ● mudah di interpretasikan
- ● Parametrik
- ● Memenuhi asumsi, uji Hipotesis dan selang interval
  
- ● Performa kurang
- ● Lolos uji, dan asumsi baru hasil
- ● Statistician/Matematikawan

## Model Mahine Learning

- Datan berukuran besar
- model kompleks
- sulit di interpretasikan
- Parametrik
- Tanpa asumsi, Uji Hipotesis dan interval
  
- Performa akan lebih baik
- hasil
- semua bisa(Komputasi)





# List Model Machine Learning

## Statistika

Linier  
Regression

Logistic  
Regression

Fractional  
Logistic

Ordinal Logistic

Dll ...

## Machine Learning

Decision Tree

Random  
Forest

Gradient  
Boosted Trees

Naïve Bayes

K-Means

DBSCAN

K-Nearest  
Neighbors

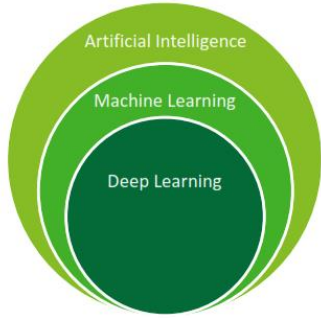
AdaBoost

XGBosst

Bagging Tress

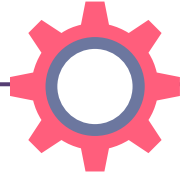
Support Vector  
Machine

Dll ...



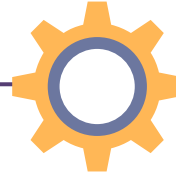
# Machine Learning

**Artificial  
Intellegence**



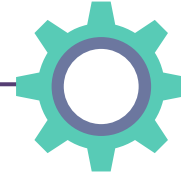
Replikasi kecerdasan  
manusia  
melaksanakan tugas

**Machine  
Learning**



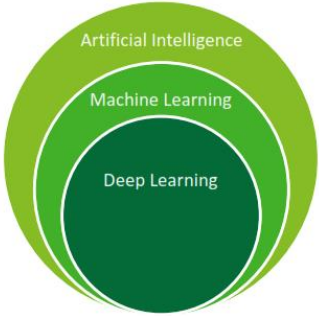
Mengacu AI  
belajar sendiri  
By experience

**Deep  
Learning**



Bagian ML  
Data Besar  
model Kompleks

# Machine Learning



**Model machine Learning** digunakan untuk mendeteksi observasi atas kejadian tertentu berdasarkan historical data



**Finance**

Fraud Detection,  
Early Warning Sistem  
Aplication Scoring



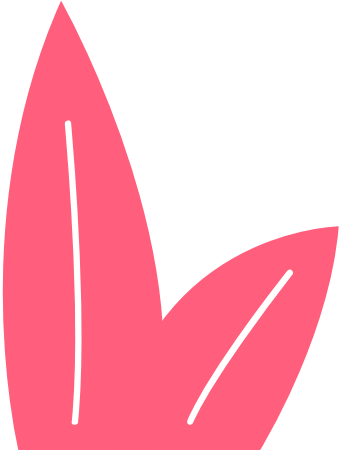
**Healthcare**

Tumor Detection  
Drug Discovery

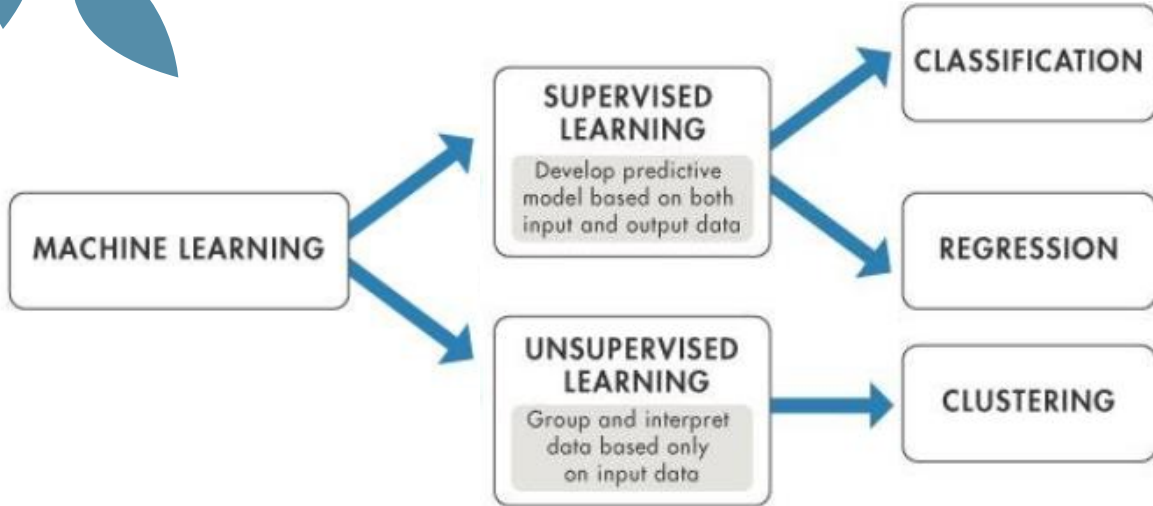


**E-Commerce**

Personalized Item  
Recommendations

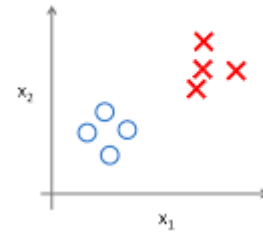


# Klasifikasi ML

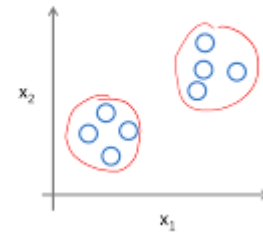


**Supervised Learning** : Target variabel diketahui  
**Unsupervised Learning** : Target variabel tidak diketahui

Supervised Learning

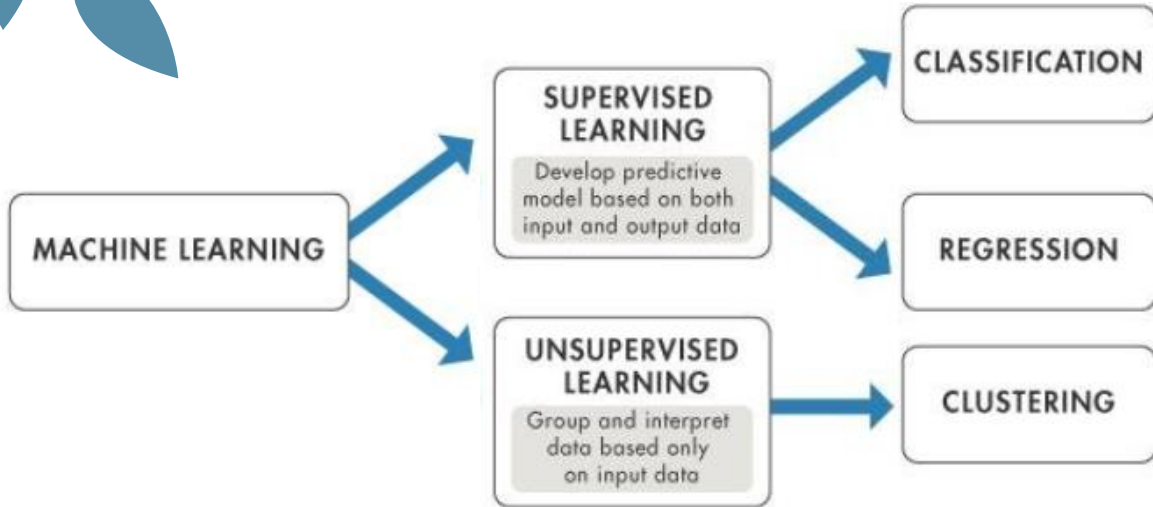


Unsupervised Learning



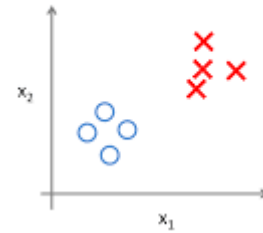


# Klasifikasi ML

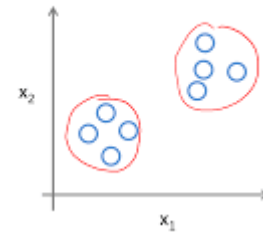


**Supervised Learning** : Target variabel diketahui  
**Unsupervised Learning** : Target variabel tidak diketahui

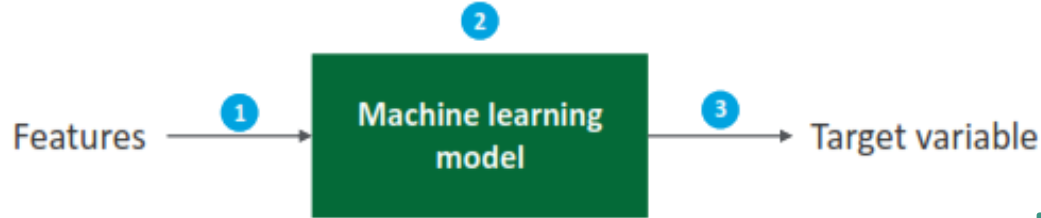
Supervised Learning



Unsupervised Learning



# Arsitektur ML



## Komponen

### Features

- Variabel bebas yang digunakan untuk memprediksi
- faktor yang diyakini memiliki pengaruh /hubungan

### Machine Learning Model

- Arsitektur model machine learning,
- model ML di bangun berdasarkan experience
- learning dari histiocal data (data training),

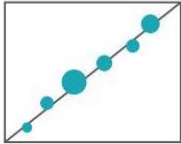
### Target Variabel

Variabel yang akan kita prediksi

## PROSES

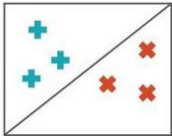
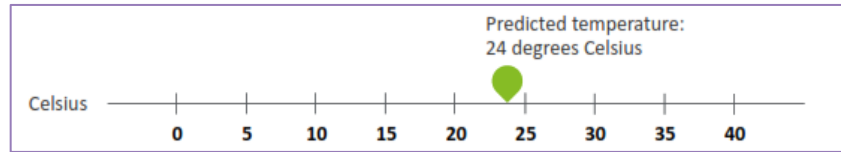
1. Model ML menerima features sbg input
2. Model ML melakukan kalkulasi menggunakan basis data input/learns by historical data
3. Model Mmenghasilkan output (target variabel)

# Supervised Learning



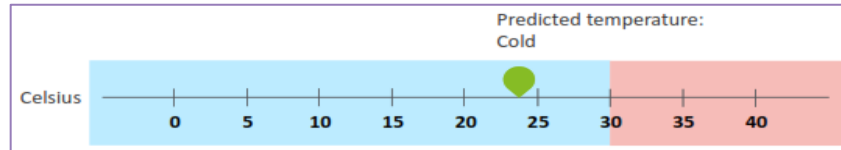
## REGRESSION

Identifying real values  
(dollars, weight, etc.)



## CLASSIFICATION

Sorting items  
into categories

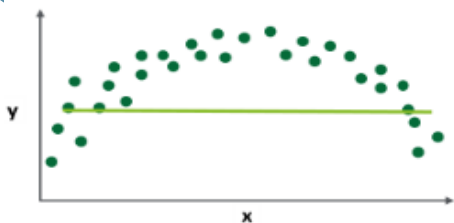


Target Variabel menentukan model ML yang akan digunakan :

- Regression : memprediksi nilai tertentu
- Classification : memprediksi Class

# Issue ML

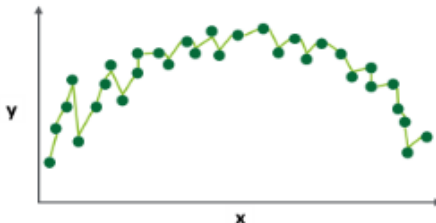
a



High bias  
Underfitting

- Model terlalu simple
- model tidak bisa menangkap pola dari dataset
- model high bias dan high varians,

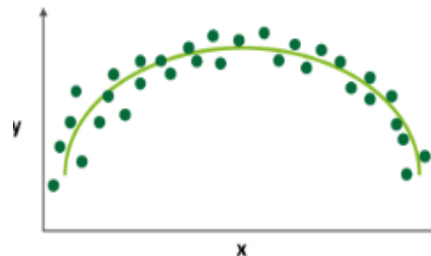
b



High variance  
Overfitting

- Model terlalu kompleks
- Menangkap noises data training,
- Model hanya merefleksikan data training bukan populasi actual
- ini membuat varians dari model besar/over fitting.
- Akurasi tidak stabil ; bagus di training & turun drastis di testing

c

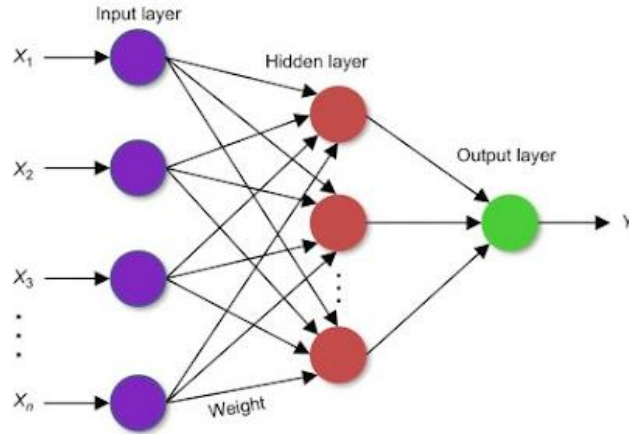


Low bias, low variance  
Balanced

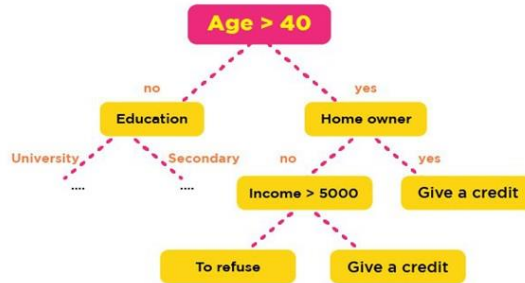
- model ideal, bias dan variance kecil
- Model stabil training dan testing

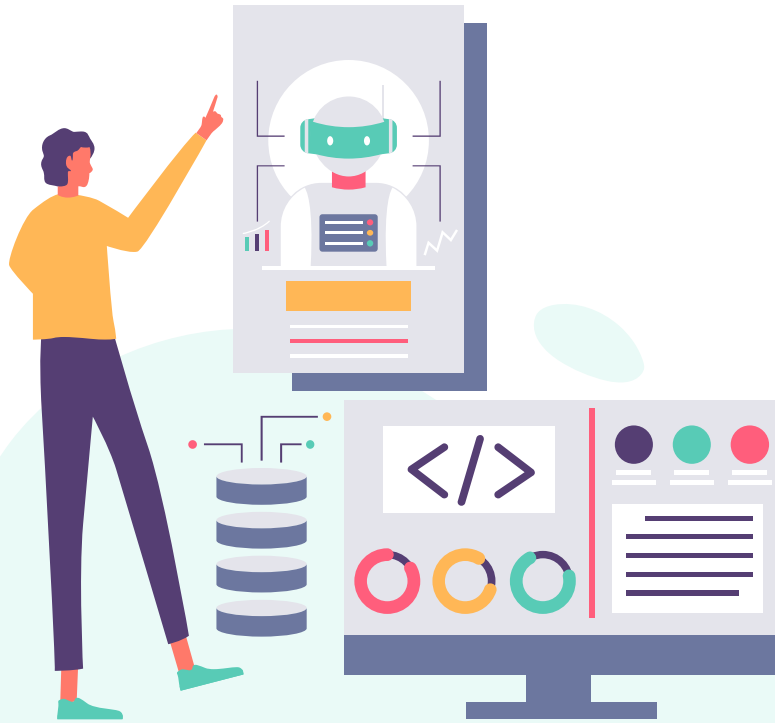
# Model ML

## Neural Networks



## Decision Tree





# THANK YOU

“Perkembangan teknologi tidak bisa dihindari, pilihan kita hanya 2 : tumbuh dan berkembang bersama atau tertinggal, tergilas dan mati”