SINESTIK II:

Seminar Nasional Tehnik Elektro, Sistem Informasi & Tehnik Informasi

Surabaya, 26 Maret 2022

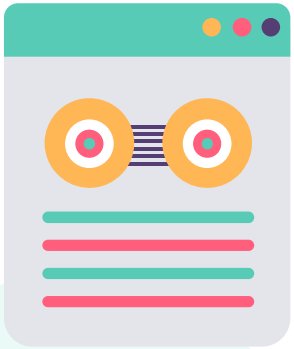
AVAN S.Si, M.Si

Statitician Modeler Bank Mandiri
2017-sekarang

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

Risk Perbankan

BISNIS



GAS

REM

RISK

Scope Project

Aplication 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

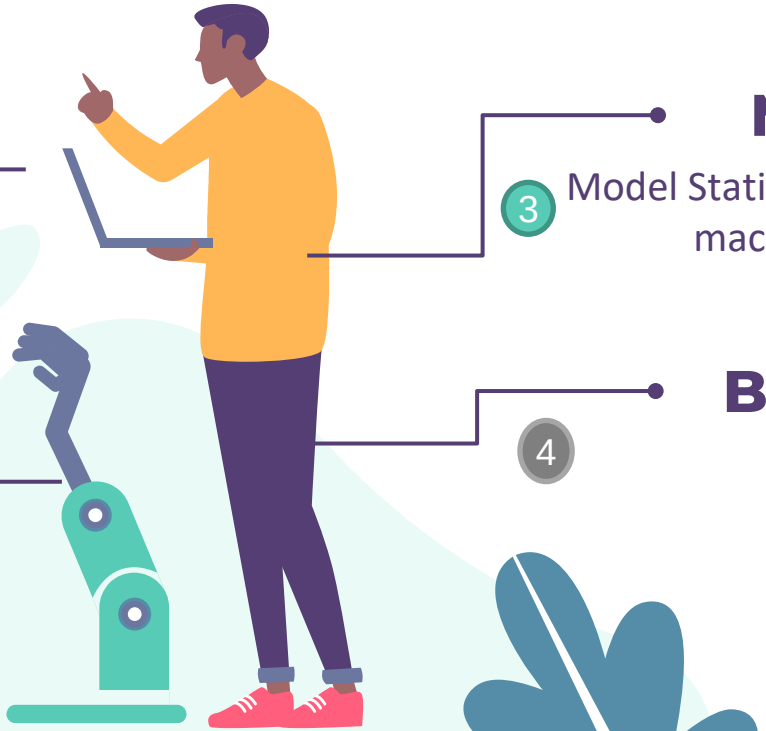
Eksternal Data

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

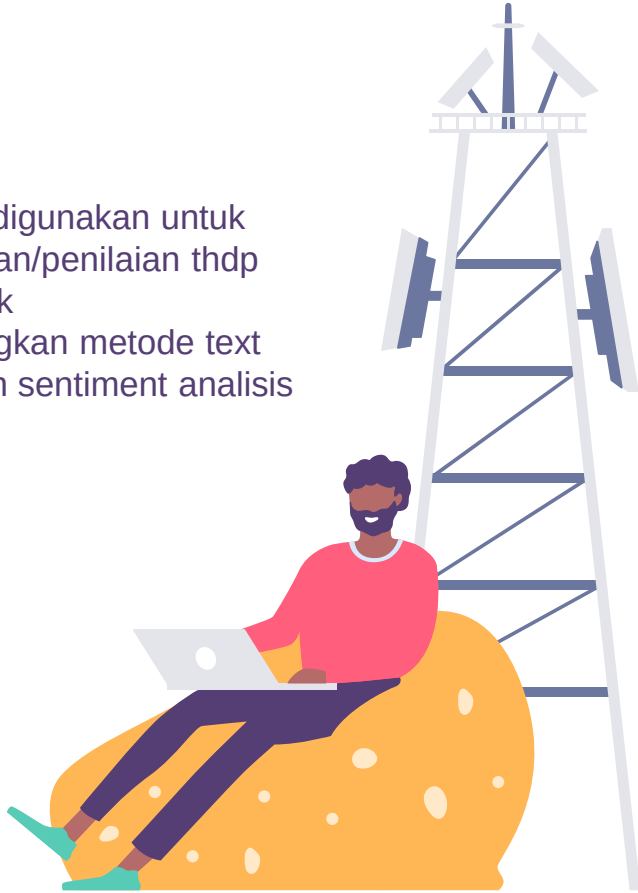
Manfaat

- 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 .

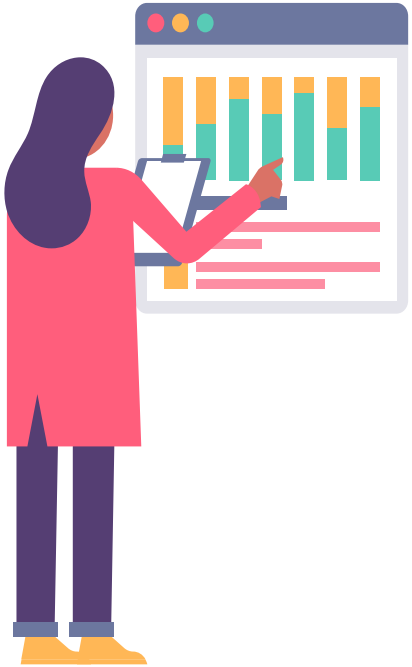
Terstruktur

Tidak terstruktur

- Data Colllection
- Data Cleaning
- Text Mining
- Clustering
- Visualitation

Sentiment Analisis

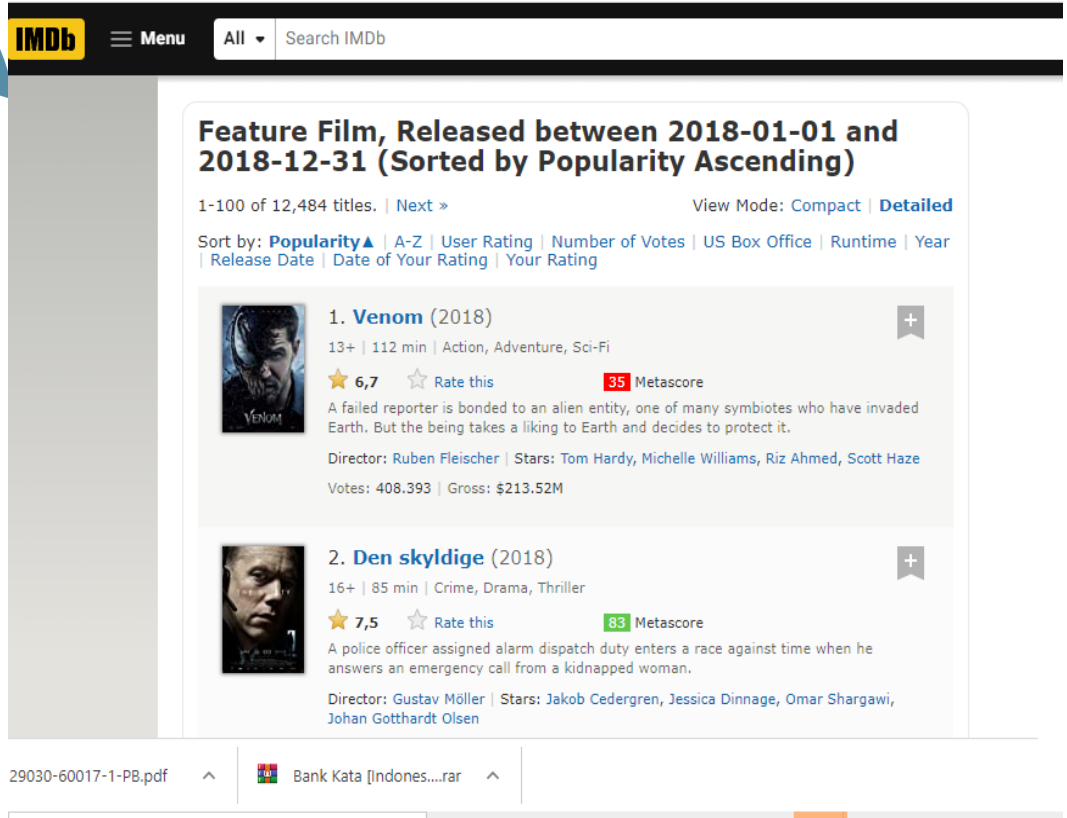
- Sentiment analysist adalah tahapan lebih lanjut dalam tahap text Mining,
- Teknik untuk mengidentifikasi bagaimana sebuah sentimen diekspresikan menggunakan teks dan dikatagorikan dalam sentimen positif maupun sentiment negatif.
- Sebagai adl kamus yang berisi list kata bersentimen positif , negative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

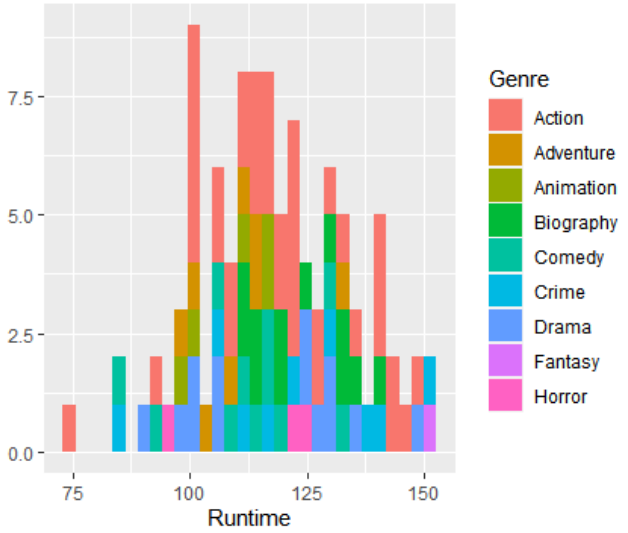


Kasus Text Mining

	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

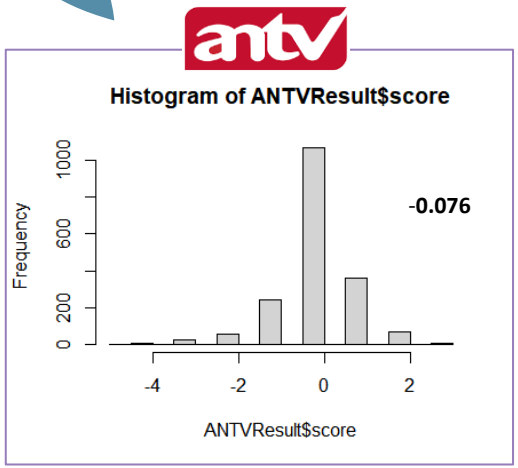
owing 1 to 12 of 100 entries, 4 total columns

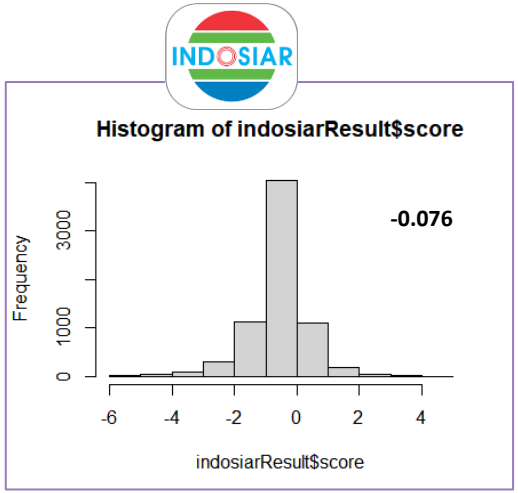


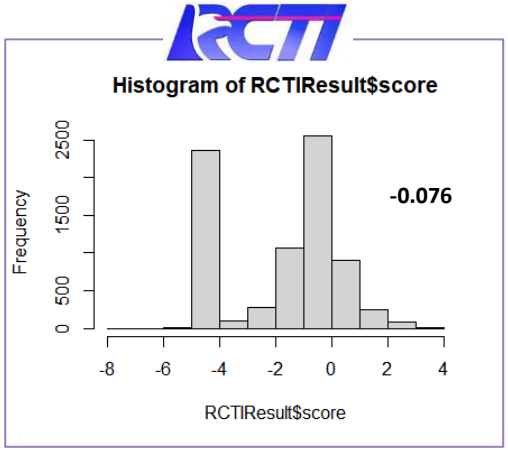


Kasus Sentimen Analisis

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







- 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 persamaaan matematis

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

Kriteria	Statistika	Model Mahine Learning
Data	data kecil perhitungan terbatas	Datan berukuran besar
Kompleksitas	model sederhana	model kompleks
Interpretasi	mudah di interpretasikan	sulit di interpretasikan
Statistika	Parametrik	Parametrik
Konstrain	Memenuhi asumsi, uji Hipotesis dan selang interval	Tanpa asumsi, Uji Hipotesis dan interval
Performa	Performa kurang	Performa akan lebih baik
Orientasi	Lolos uji, dan asumsi baru hasil	hasil
Modeler	Statistician/Matematikawan	semua bisa(Komputasi)

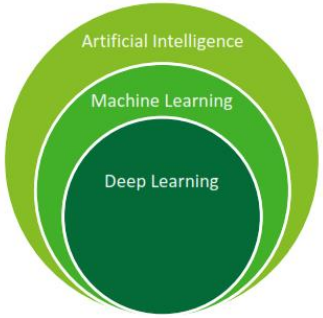
List Model Machine Learning

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

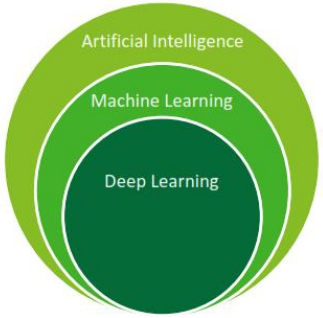
Machine Learning

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

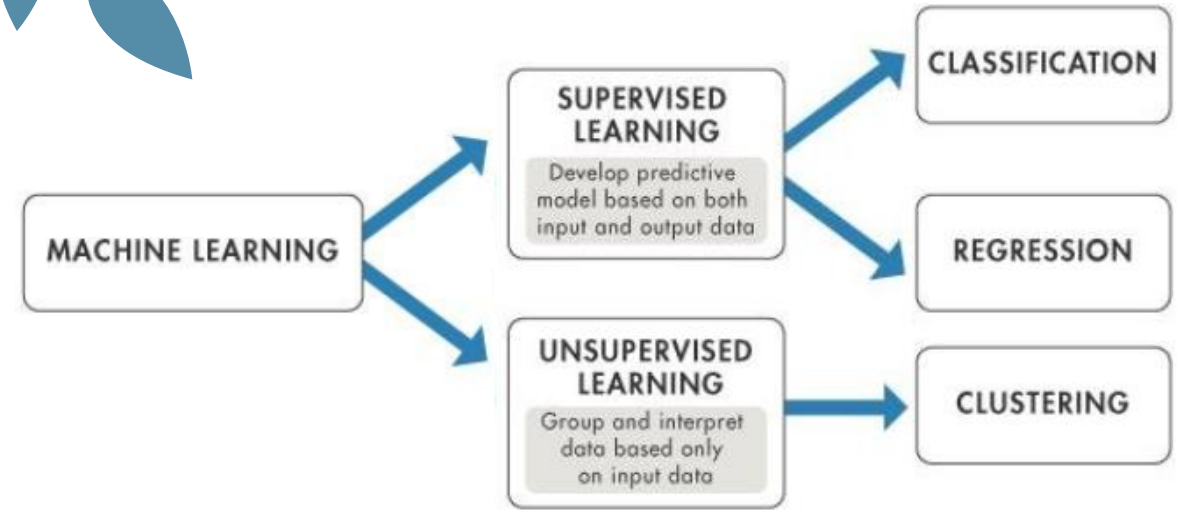
Healthcare

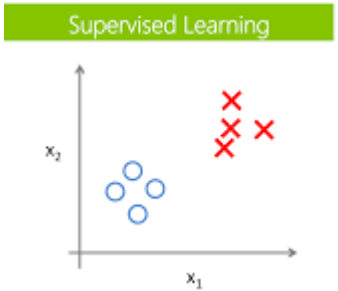
Tumor Detection
Drug Discovery

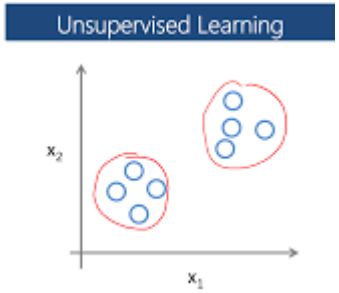
E-Commerce

Personalized Item
Recommendations

Klasifikasi ML

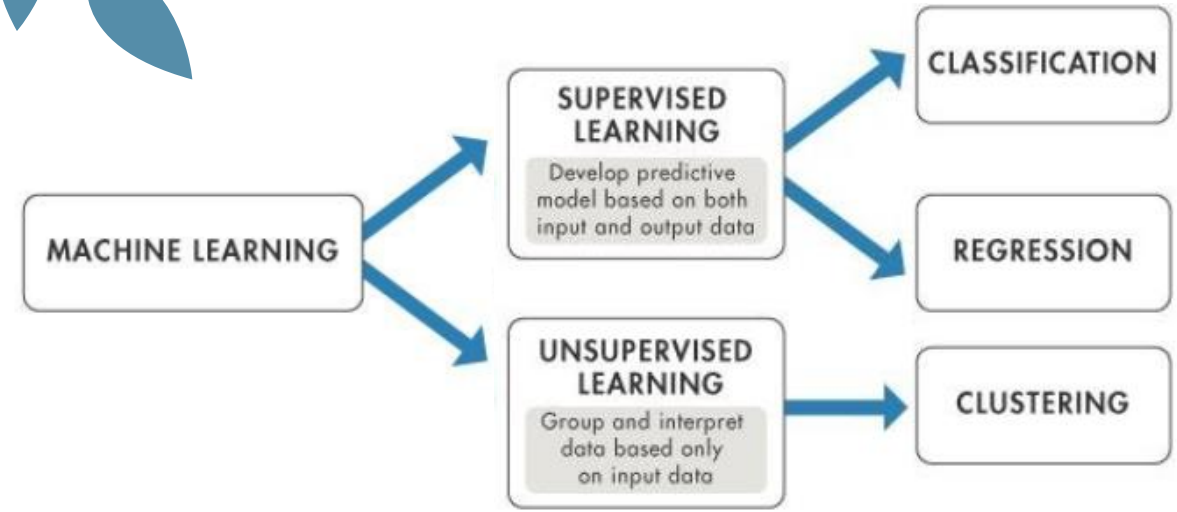


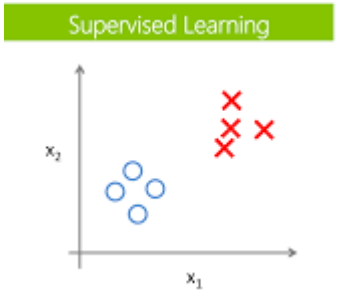


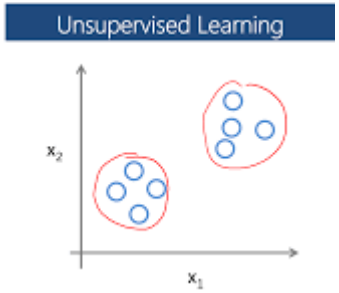


Supervised Learning : Target variabel diketahui
Unsupervised Learning : Target variabel tidak diketahui

Klasifikasi ML

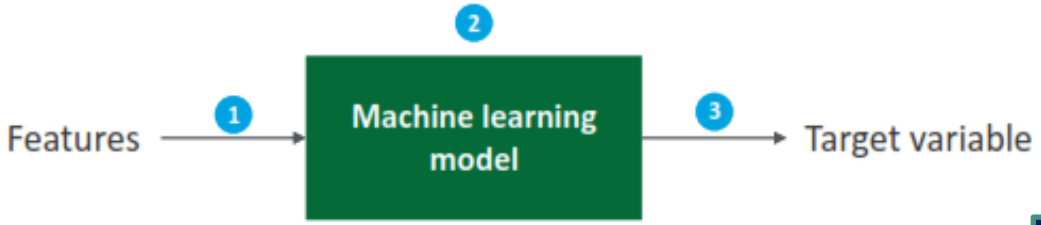






Supervised Learning : Target variabel diketahui
Unsupervised Learning : Target variabel tidak diketahui

Arsitektur ML



Komponen

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

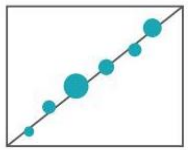
Machine Learning Model
- Arsitektur model machine learning,
- model ML di bangun berdasarkan experience
- learning dari histirocal data (data training),

Target Variabel
Variabel yang akan kita prediksi

PROSES

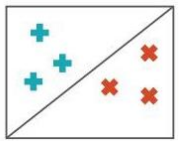
1. Model ML menerima features sbg input
2. Model ML melakukan kalkulasi menggunakan basis data input\learns by historical data
3. Model Mlmenghasilkan output (target variabel)

Supervised Learning



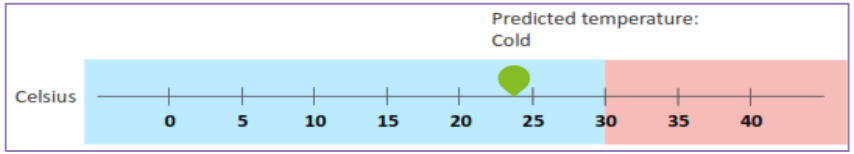
REGRESSION

Identifying real values
(dollars, weight, etc.)



CLASSIFICATION

Sorting items
into categories

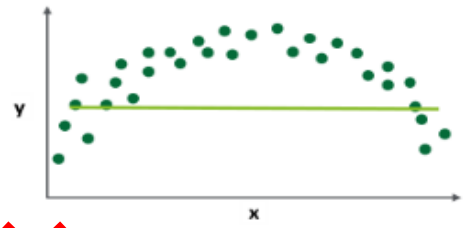


Target Variabel menentukan model ML yang akan digunakan :

- Regression : memprediski nilai tertentu
- Classification : memprediski Class

Issue ML

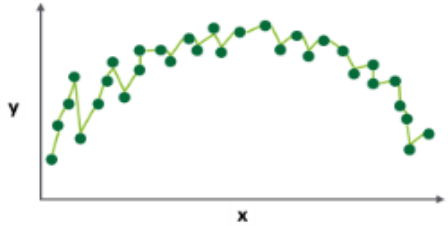
a



High bias
Underfitting

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

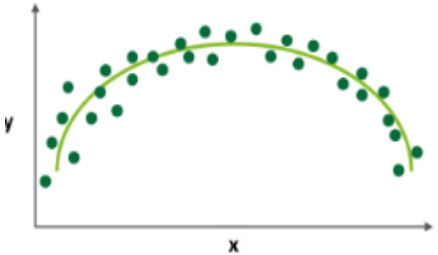
b



High variance
Overfitting

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

c

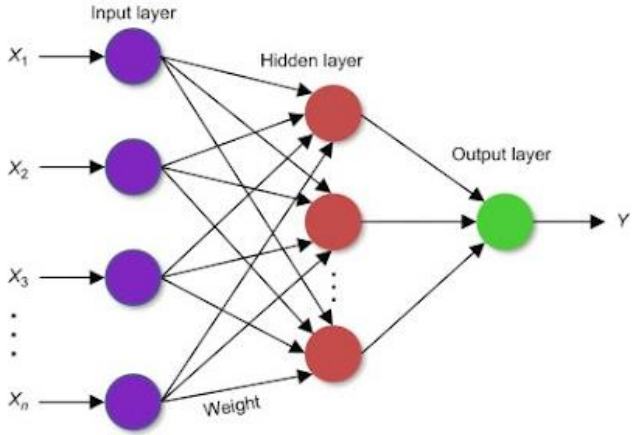


Low bias, low variance
Balanced

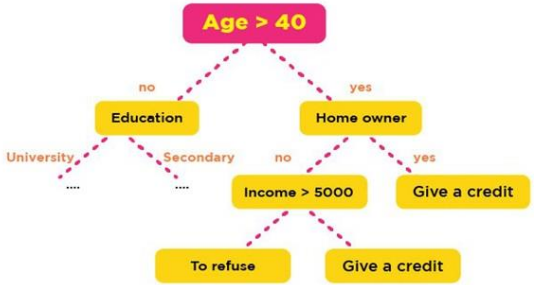
- model ideal, bias dan variance kecil
- Model stabil training dan testing

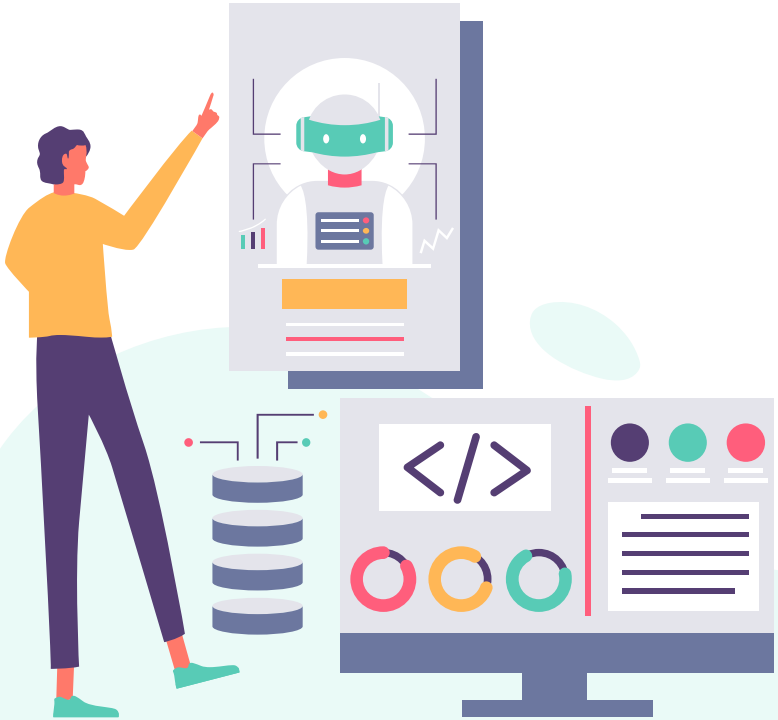
Model ML

Neural Networks



Decision Tree





THANK YOU

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