

Introduction

Machine Learning Fundamentals

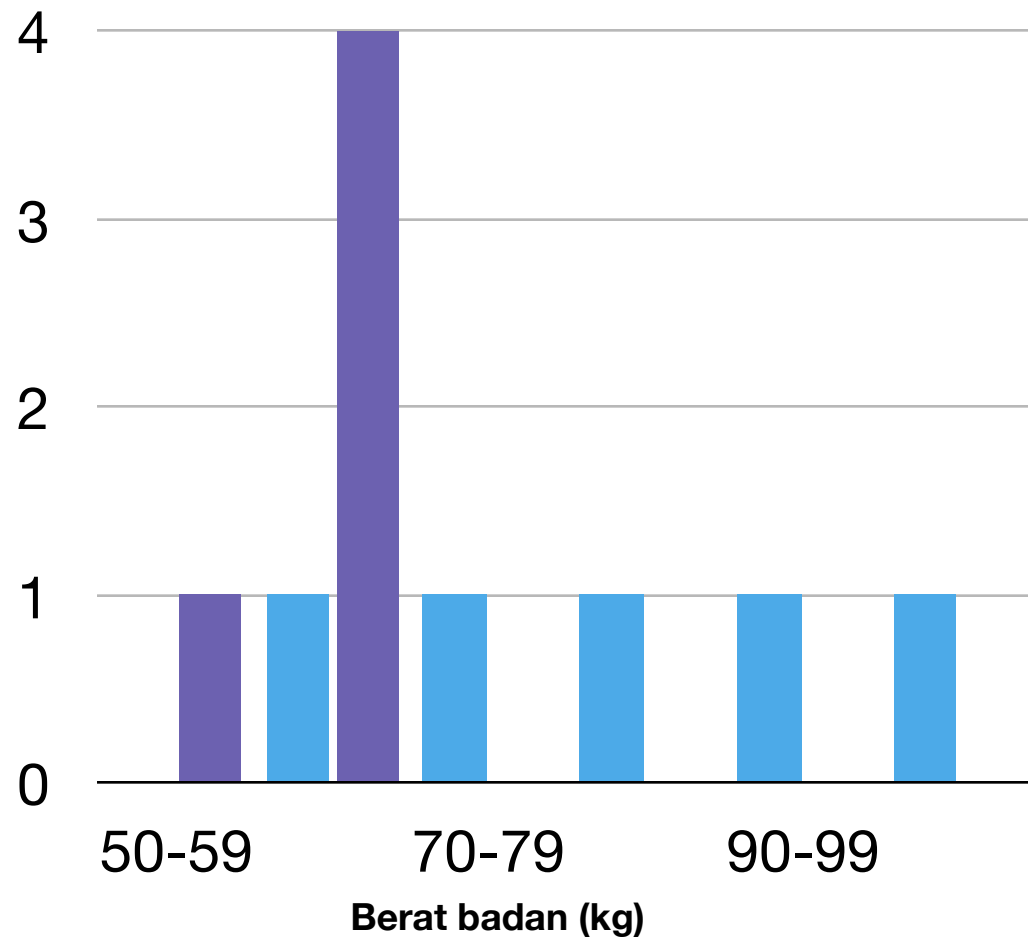
Tujuan

- Mengenalkan konsep Machine Learning
 - Cara memecahkan masalah
 - Memahami masalah yang dapat dipecahkan
- Mengenal data yang akan dianalisa
 - Mem-visualisasikan data supaya dapat dimengerti
 - Membersihkan & mempersiapkan data
- Machine learning menggunakan Python

Data Karyawan Perusahaan

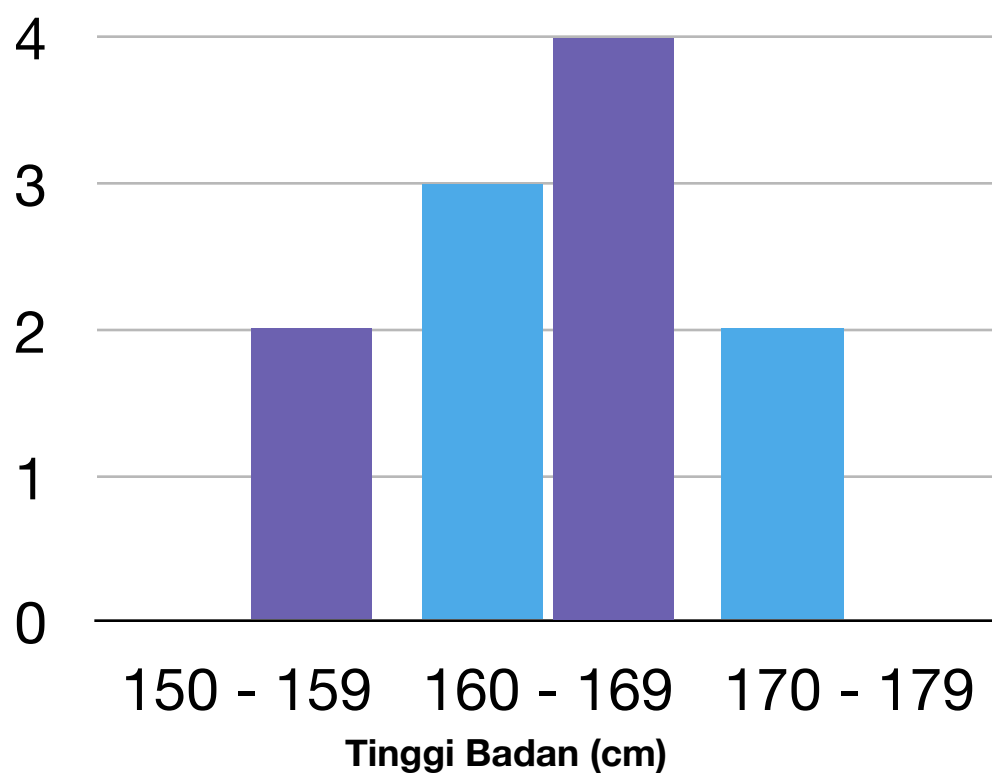
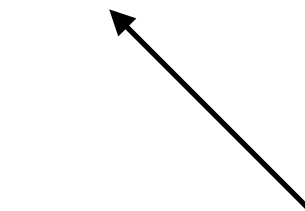
No	Jenis Kelamin	Berat Badan (kg)	Tinggi
1	Pria	80	170
2	Pria	78	168
3	Wanita	60	157
4	Pria	68	165
5	Wanita	58	160
6	Wanita	62	178
7	Pria	90	164
8	Pria	108	172
9	Wanita	64	150
10	Wanita	65	160

- Analisa *insight* yang bisa didapat
- Berat badan pria lebih dari perempuan
- Tinggi badan perempuan lebih dari pria

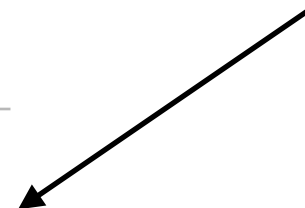


- Analisa *insight* yang bisa didapat

- Berat badan pria lebih dari wanita



- Tinggi badan wanita lebih dari pria



Analisa Data

- **Pria**

- *Tinggi badan:*

- Min:
- Max:
- Mean:
- Std. Dev:

- *Berat badan:*

- Min
- Max:
- Mean:
- Std. Dev

- **Wanita**

- *Tinggi badan:*

- Min:
- Max:
- Mean:
- Std. Dev:

- *Berat badan:*

- Min
- Max:
- Mean:
- Std. Dev

No	Jenis Kelamin	Berat Badan (kg)	Tinggi
1	Pria	80	170
2	Pria	78	168
3	Wanita	60	157
4	Pria	68	165
5	Wanita	58	160
6	Wanita	62	178
7	Pria	90	164
8	Pria	108	172
9	Wanita	64	150
10	Wanita	65	160

Prediksi Jenis Kelamin

- Apabila terdapat orang ke-11 dengan berat badan 60 kg dan tinggi badan 162 cm?
 - Pria atau wanita?
 - Seberapa yakin?
 - Dasarnya apa?
 - Bisa dipakai untuk perusahaan lain?

(Lebih Banyak) Pertanyaan Lagi

- Apa yang akan terjadi apabila data sampel karyawan diperbanyak? (Anggap 1000 data)
- Apa yang akan terjadi apabila ada kesalahan dalam pengambilan data?

Definisi

- Data di tabel dikenal sebagai ***training set***
- Data yang belum pernah dilihat sebelumnya dikenal sebagai ***validation set*** ataupun ***testing set***
- Langkah utama dimana kita menganalisa data dikenal sebagai ***Exploratory Data Analysis (EDA)***
- Berat badan serta tinggi badan merupakan ***features***
 - *Feature* merupakan atribut/karakteristik pada data yang dianalisa
 - *Feature* dianalisa untuk mengeluarkan hasil output yang diinginkan

Definisi

- Berat badan serta tinggi badan merupakan **features**
- *Feature* merupakan atribut/karakteristik pada data yang dianalisa
- *Feature* dianalisa untuk mengeluarkan hasil output yang diinginkan

No	Jenis Kelamin	Berat Badan (kg)	Tinggi
1	Pria	80	170
2	Pria	78	168
3	Wanita	60	157
4	Pria	68	165
5	Wanita	58	160
6	Wanita	62	178
7	Pria	90	164
8	Pria	108	172
9	Wanita	64	150
10	Wanita	65	160

Machine Learning

- Memiliki dasar dari ilmu statistik
- Membuat suatu **prediksi** berdasarkan dari suatu set data (**dataset**)
- **Features** digunakan menjadi dasar daripada prediksi yang dilakukan
- **Features** diolah menggunakan **algoritma** yang diinginkan untuk bisa menemukan **pattern/pola** dalam data sehingga menghasilkan suatu prediksi

Machine Learning

- **Contoh algoritma:** k-Nearest Neighbor, SVM, Naive Bayes, Logistic Regression
- Hasil pembelajaran oleh algoritma disimpan menjadi suatu file (**model**) lalu digunakan untuk melakukan prediksi terhadap data yang belum pernah dilihat sebelumnya
- Kemampuan dalam membuat prediksi **dievaluasi** menggunakan *validation set* ataupun *testing set*

Merumuskan masalah



Mengumpulkan data



Menganalisa data
(EDA)



Training



Evaluation

Machine Learning Tasks

- **Classification**
 - Diketahui tinggi dan berat seseorang, apa jenis kelamin orang tersebut?
 - Features: Tinggi & berat
 - Output: kategori

No	Jenis Kelamin	Berat Badan (kg)	Tinggi
1	Pria	80	170
2	Pria	78	168
3	Wanita	60	157
4	Pria	68	165
5	Wanita	58	160
6	Wanita	62	178
7	Pria	90	164
8	Pria	108	172
9	Wanita	64	150
10	Wanita	65	160

Machine Learning Tasks

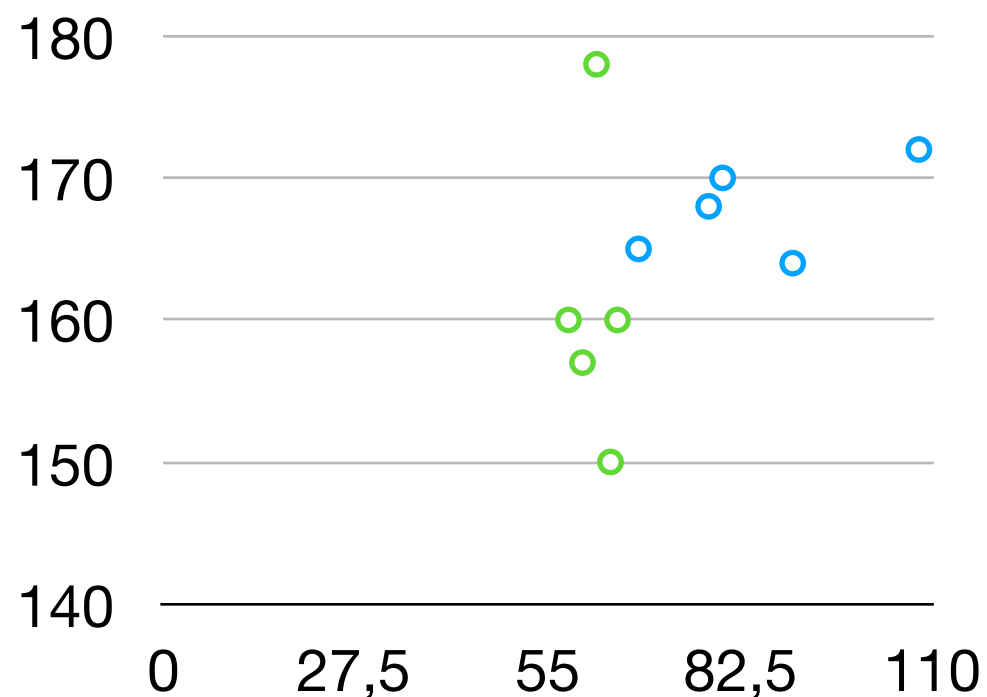
- **Regression**
 - Diketahui tinggi dan berat seseorang, berapakah kapasitas paru-parunya?
 - Features: Tinggi & berat
 - Output: data numerik

No	Berat Badan (kg)	Tinggi	Kapasitas Paru-paru (l)
1	80	170	4.6
2	78	168	4.8
3	60	157	3.2
4	68	165	3.1
5	58	160	3.0
6	62	178	3.1
7	90	164	4.5
8	108	172	4.0
9	64	150	3.8
10	65	160	4.5

Machine Learning Tasks

- **Clustering**

- Diketahui tinggi dan berat seseorang, apa jenis kelamin orang tersebut?
- Features: Tinggi & berat
- Output: kategori kelompok



No	Jenis Kelamin	Berat Badan (kg)	Tinggi
1	Pria	80	170
2	Pria	78	168
3	Wanita	60	157
4	Pria	68	165
5	Wanita	58	160
6	Wanita	62	178
7	Pria	90	164
8	Pria	108	172
9	Wanita	64	150
10	Wanita	65	160

Machine Learning Tasks

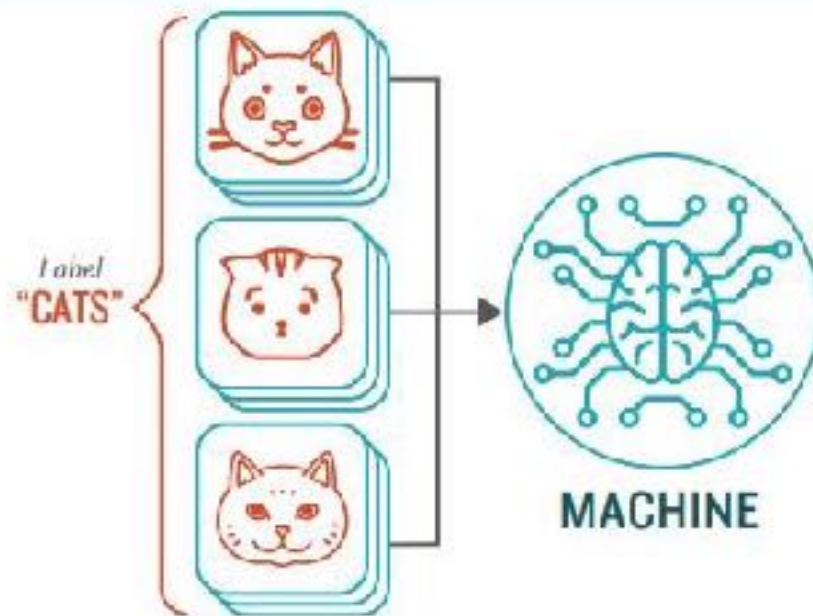
- **Time-Series Prediction**
 - Diketahui perkembangan tinggi badan seseorang, berapakah indeks massa tubuh (perbandingan antara berat dan tinggi badan) 1 tahun kedepan?
 - Input: data historis
 - Output: data numerik

No	Tanggal	Berat Badan (kg)	Tinggi	Body Mass Index
1	2019-01	80	170	22.3
2	2019-02	78	168	21.2
3	2019-03	60	157	18.4
4	2019-04	68	165	19.0
5	2019-05	58	160	18.4
6	2019-06	62	178	19.0
7	2019-07	90	164	23.0
8	2019-08	108	172	38.2
9	2019-09	64	150	18.3
10	2019-10	65	160	18.5

How **Supervised** Machine Learning Works

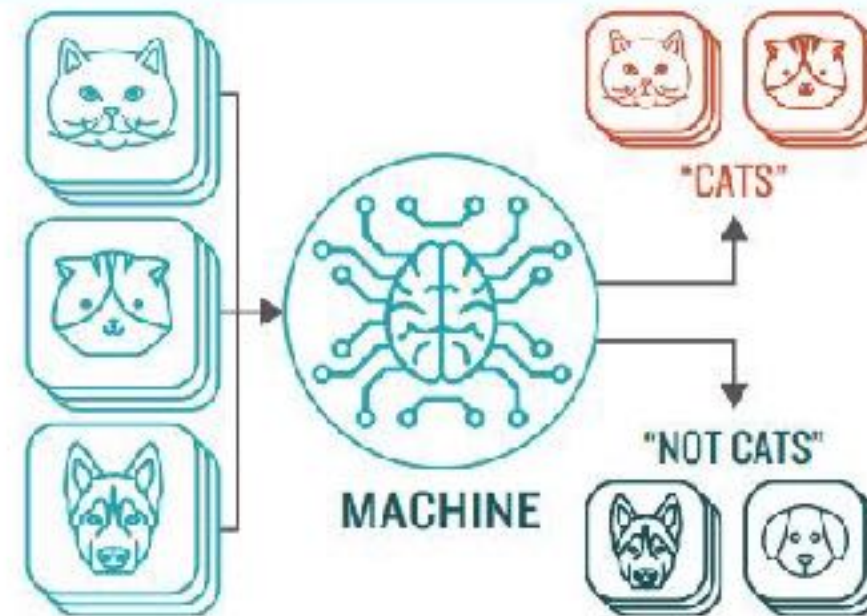
STEP 1

Provide the machine learning algorithm categorized or "labeled" input and output data from to learn

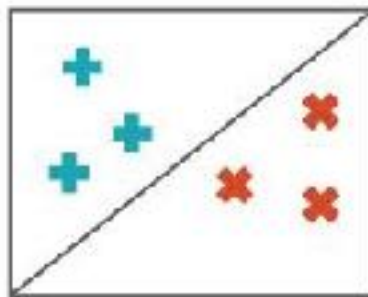


STEP 2

Feed the machine new, unlabeled information to see if it tags new data appropriately. If not, continue refining the algorithm

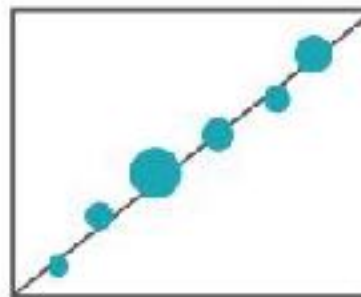


TYPES OF PROBLEMS TO WHICH IT'S SUITED



CLASSIFICATION

Sorting items into categories



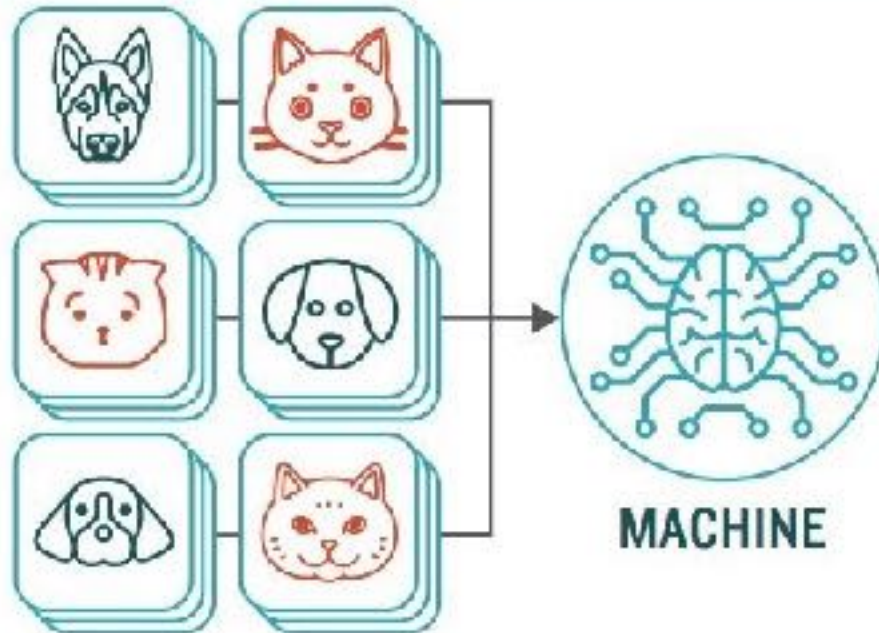
REGRESSION

Identifying real values (dollars, weight, etc.)

How **Unsupervised** Machine Learning Works

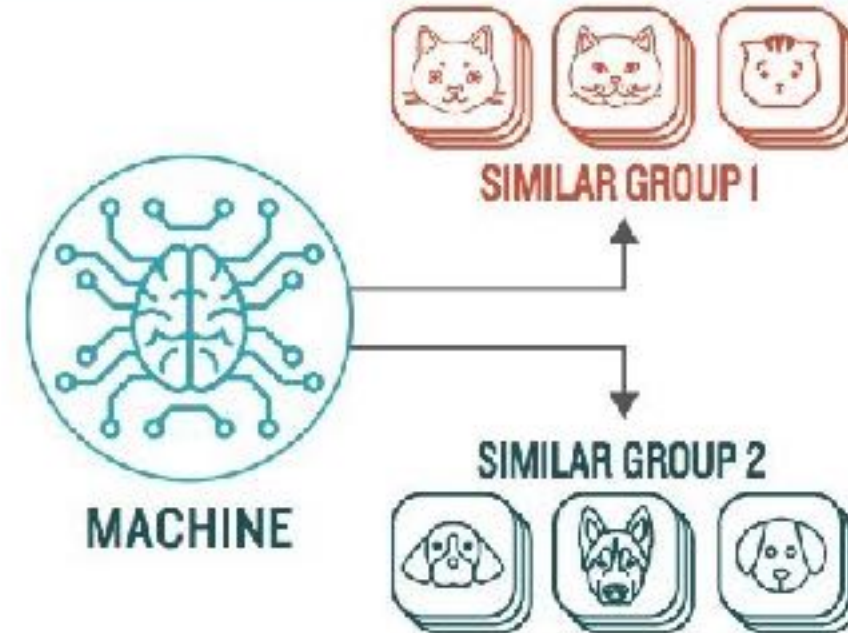
STEP 1

Provide the machine learning algorithm uncategorized, unlabeled input data to see what patterns it finds

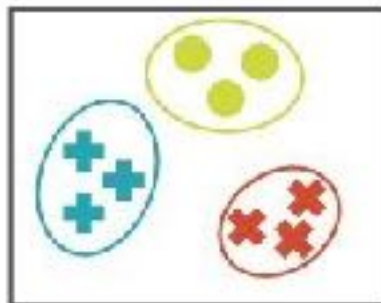


STEP 2

Observe and learn from the patterns the machine identifies



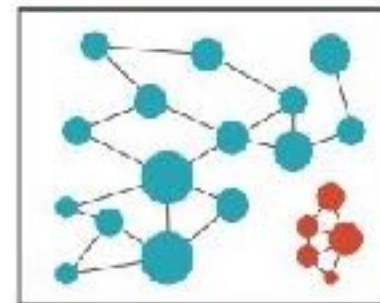
TYPES OF PROBLEMS TO WHICH IT'S SUITED



CLUSTERING

Identifying similarities in groups

For Example: Are there patterns in the data to indicate certain patients will respond better to this treatment than others?



ANOMALY DETECTION

Identifying abnormalities in data

For Example: Is a hacker intruding in our network?

Machine Learning Data

- **Numeric Tabular**
 - Diketahui tinggi dan berat badan seseorang, apakah kita bisa mengetahui jenis kelaminnya?
- **Images (Computer Vision)**
 - Diketahui gambar seseorang, apakah kita bisa mengetahui jenis kelaminnya?

No	Jenis Kelamin	Berat Badan (kg)	Tinggi
1	Pria	80	170
2	Pria	78	168
3	Wanita	60	157
4	Pria	68	165
5	Wanita	58	160
6	Wanita	62	178
7	Pria	90	164
8	Pria	108	172
9	Wanita	64	150
10	Wanita	65	160



Machine Learning Data

- **Text (Natural Language Processing)**
 - Diketahui suatu topik yang seseorang katakan, apakah kita bisa mengetahui jenis kelamin orang tersebut?
 - Contoh:
 - *“Saya olah raga mengangkat beban setiap hari”*
 - *“Cobalah produk kecantikan untuk menghaluskan kulit anda”*

Kenapa Python?

- Banyak library siap pakai yang tersedia di Python
- Di-support oleh komunitas dengan baik
- Performa bahasa yang baik, serta penulisan kode relatif mudah dimengerti

Kenapa Python?

- Python hanyalah **alat** untuk melakukan machine learning.
- Bisa Python tidak bisa machine learning? **Mungkin**
- Tidak bisa Python tetapi bisa machine learning? **Mungkin**
- Tidak bisa keduanya? **Mungkin**
- Tujuan utama: **memahami** (konsep) & **mempraktekan** (Python)

Kesimpulan

- Keyword penting:
 - Dataset
 - Training set
 - Testing set & Validation set
 - Exploratory Data Analysis (EDA)
 - Features
 - Classification, regression, clustering, time-series prediction
 - Supervised & unsupervised learning