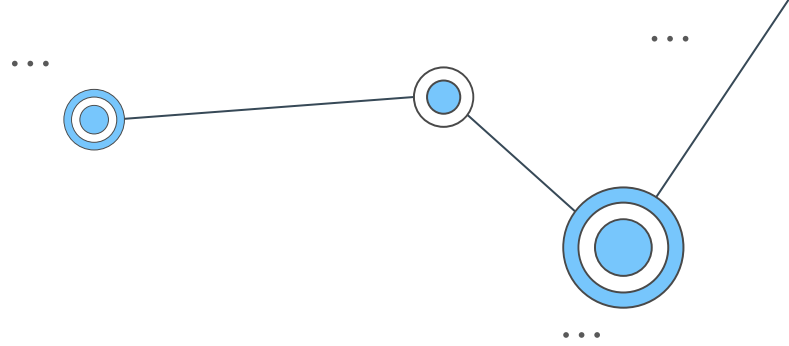




Introduction to Machine Learning with Python

Regita Putri Permata



About Me



- Undergraduate, Statistics ITS (2015-2019)
- Master Degree Program, Statistics, ITS (2019- On-going)
- Data Science Indonesia, East Java Region

**DATA
SCIENCE
INDONESIA**

Table of Contents

01

Introduction to Machine Learning

...

02

Application of Machine Learning in Indonesia and world

...

03

Function and purpose of machine learning

...

04

Steps Machine Learning in Python

...

05

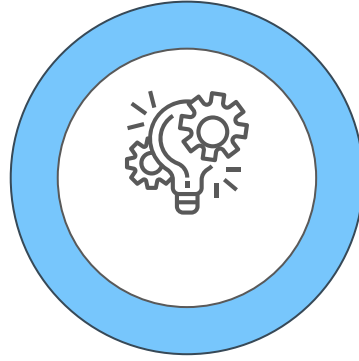
constraints in modeling

...

06

Tips and Tricks to machine learning with Python

...

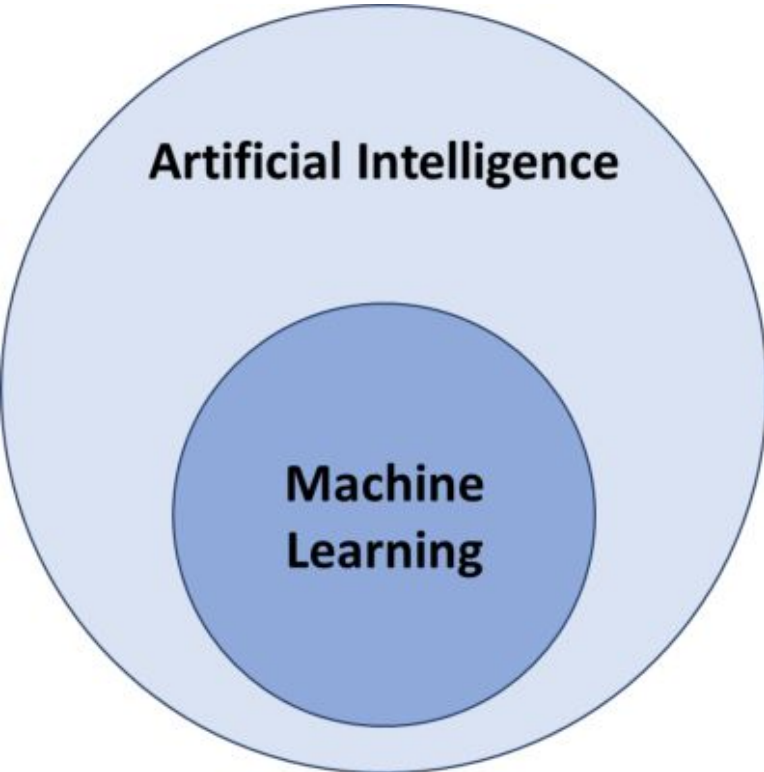
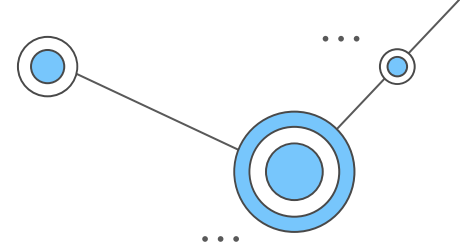


What is Machine Learning ?

The study of computer algorithms that improve
automatically through experience.

...

What is Machine Learning ?



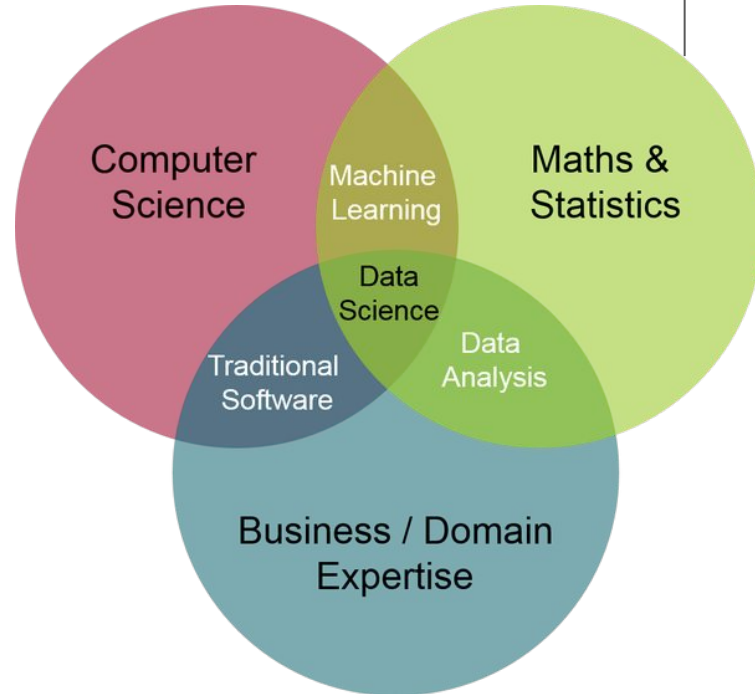
- A huge set of tools for making computers behave intelligently
- Machine learning is the most prevalent **subset of AI**
- Machine learning use **statistical method**

Machine Learning disciplines

Machine Learning is a natural outgrowth of the intersection of **Computer Science** and **Statistics**.

- The question of Computer Science is **“How can we build machines that solve problems ?”**
- The question of Statistics is **“What can be inferred from data plus a set of modeling assumptions ?”**

(Tom M. Mitchell, 2006 The Discipline of Machine Learning)



When do we use machine learning ?

Energy, Feedstock & Utilities

- Power usage analytics
- Seismic data processing
- Smart grid management
- Energy demand & supply optimization

Financial Services

- Risk analytics & regulation
- Customer segmentation
- Your text here
- Credit worthiness evaluation

Travel & Hospitality

- Aircraft scheduling
- Dynamic pricing
- Your text here
- Traffic patterns & congestion management

Manufacturing

- Predictive maintenance or condition monitoring
- Demand forecasting
- Process optimization
- Telematics

Retail

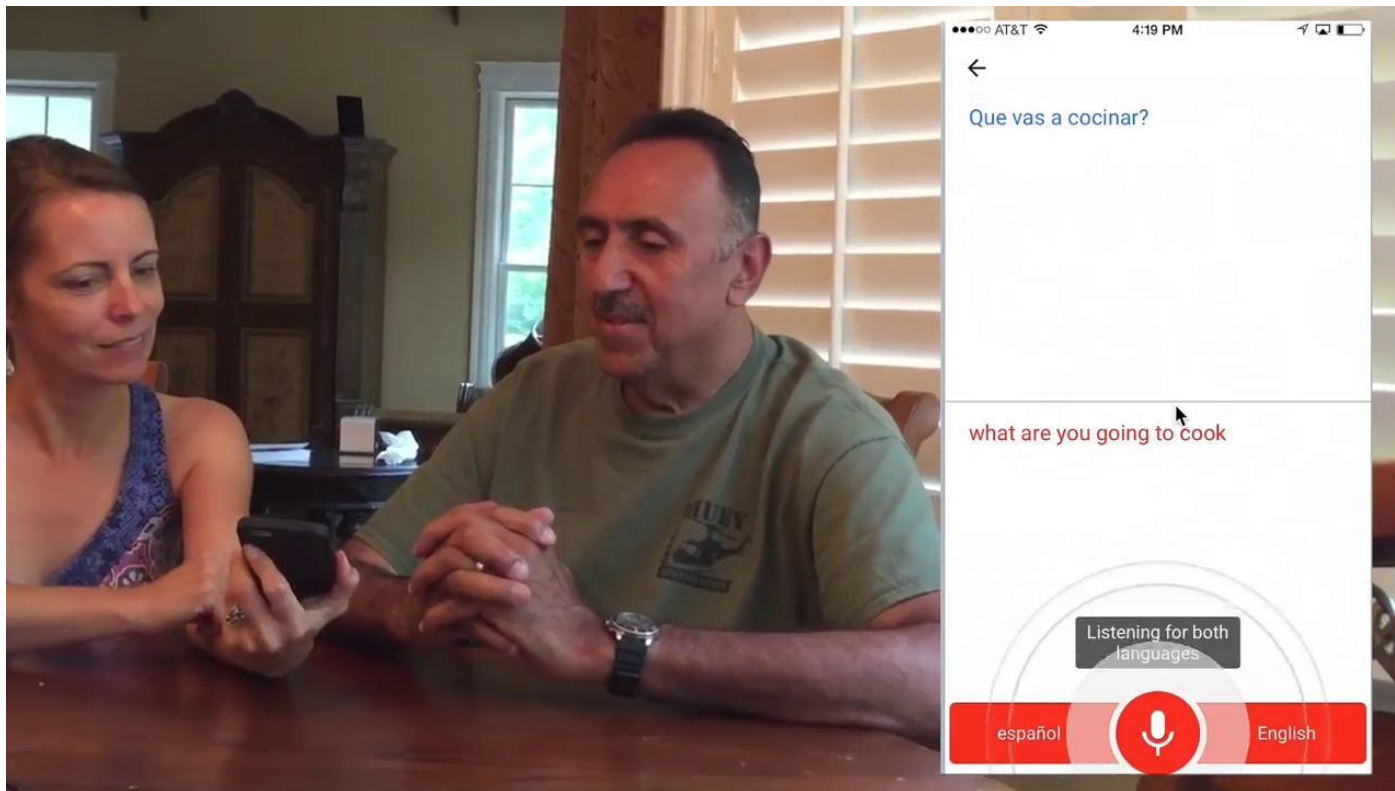
- Predictive inventory planning
- Recommendation engines
- Your text here
- Customer Roil & lifetime value

Healthcare & Life Sciences

- Alerts & diagnostics from real-time patient data
- Your text here
- Proactive health management
- Healthcare provider sentiment analysis



Application Machine Learning



**Voice
recognition**
ex. algorithm
speech to
speech end

Application Machine Learning

The Netflix Recommender System: Algorithms, Business Value, and Innovation

13:3



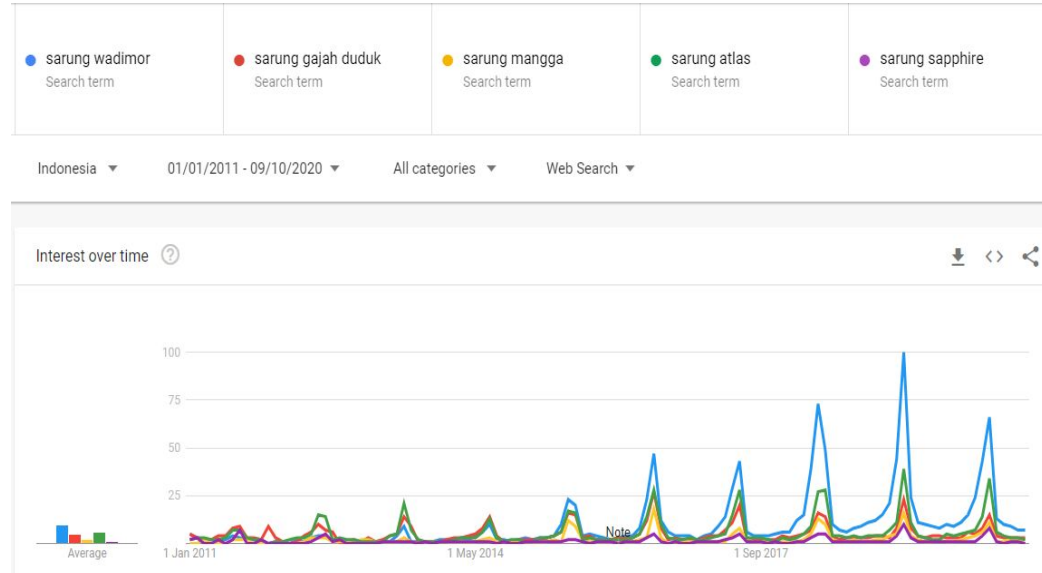
Netflix recommendation system using machine learning for business purpose

Application Machine learning for prediction

Forecasting the Search Trend of Muslim Clothing in **Indonesia** on **Google Trends** Data Using ARIMAX and **Neural Network**

<https://www.springerprofessional.de/en/forecasting-the-search-trend-of-muslim-clothing-in-indonesia-on-/17200116>

- **The results show that FFNN yields the most accurate forecast.** The forecast results show **that the search trend for Atlas brand is predicted to be the highest of all sarong brands.**
- This prediction can be used as a production plan for the next year for the company concerned



... Application machine learning for Classification ...

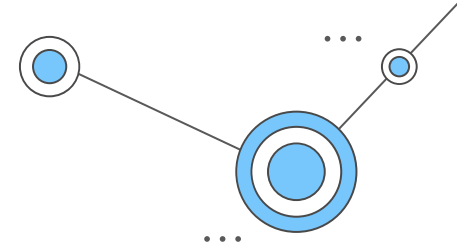
Klasifikasi Sentimen Wisatawan Candi Borobudur pada Situs TripAdvisor Menggunakan Support Vector Machine dan K-Nearest Neighbor

Review of visitor

Amazing place, sunrise was great, but it is worth to spend time to see the temple proper-ly with a guide to understand the complexity of the place and the work done



Concept Machine Learning in Python



Types of Machine Learning

- **Supervised learning**

-Given: training data + desired outputs (labels)

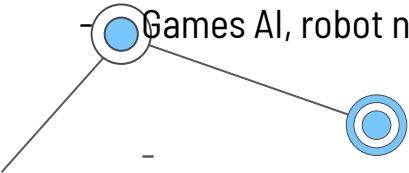
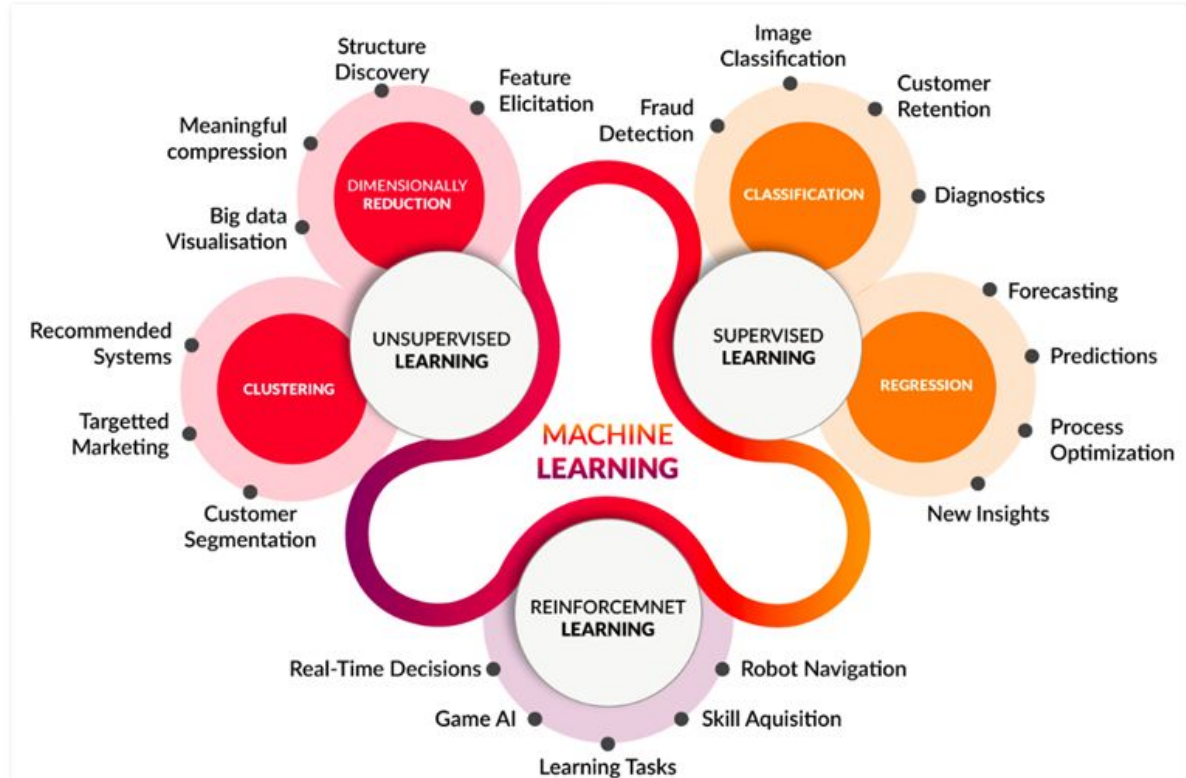
- **Unsupervised learning**

-Given: training data (without desired outputs)

- **Reinforcement learning**

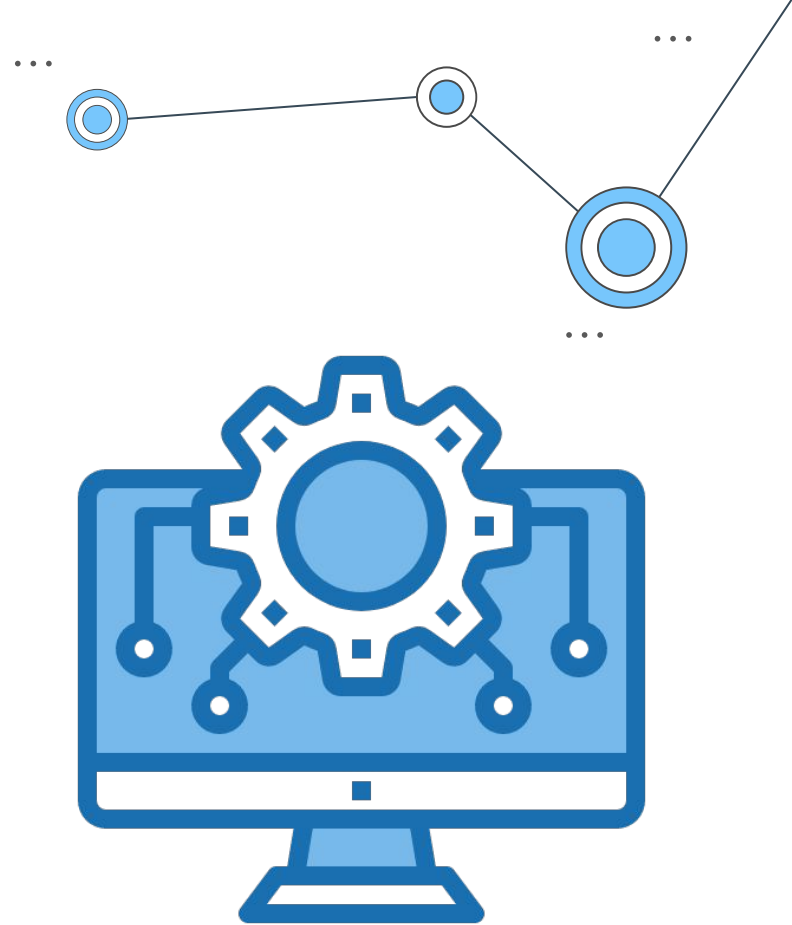
-Rewards from sequence of actions

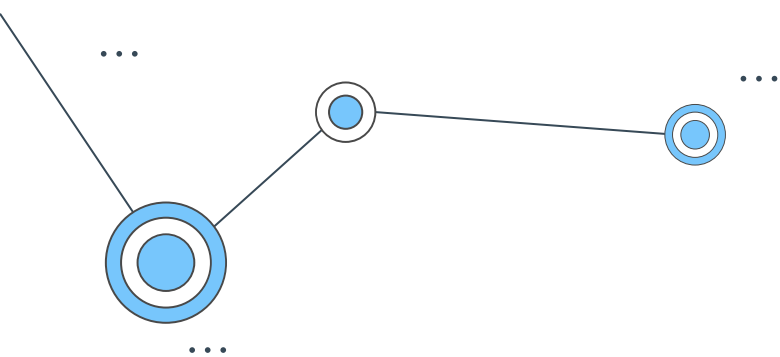
- Games AI, robot navigation



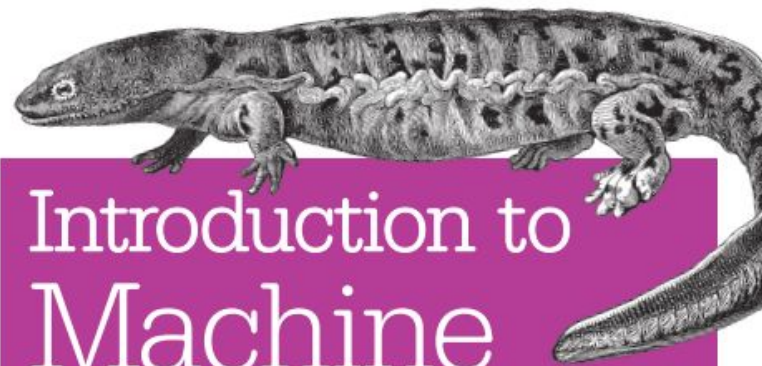
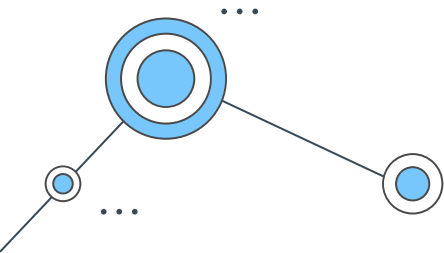
Top 9 Algorithms Machine Learning should know

1. Naive Bayes Classifier Algorithm
2. K- Means Clustering Algorithm
3. Support Vector Machines Algorithm
4. Linear Regression Algorithm
5. Logistics Regression Algorithm
6. Decision Tree Algorithm
7. Random Forest Algorithm
8. K-Nearest Neighbor Algorithm
9. Artificial Neural Network Algorithm





Machine Learning with Python

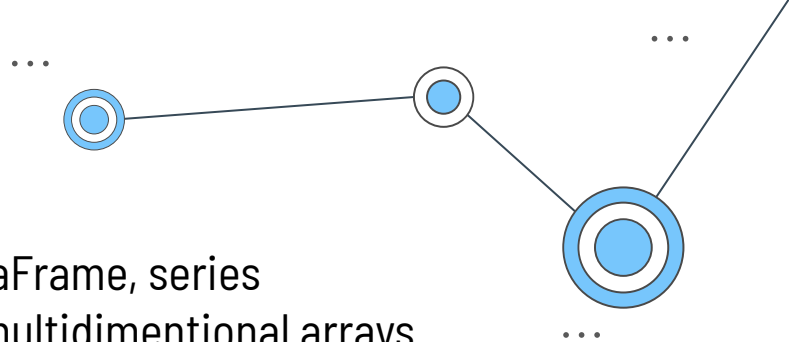


Introduction to Machine Learning with Python

A GUIDE FOR DATA SCIENTISTS



Python Tools for Machine Learning



1. **Pandas** : for data wrangling and analysis. Ex. DataFrame, series
2. **numPy** : Packages for scientific computing. Ex. multidimensional arrays
3. **Matplotlib** : for data visualize. Ex. histogram, scatterplot, boxplot
4. **Scikit-learn** : package for machine learning

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn import linear_model
```

	mean radius	mean texture	mean perimeter	mean area	mean smoothness	mean compactness
0	17.99	10.38	122.80	1001.0	0.11840	0.27760
1	20.57	17.77	132.90	1326.0	0.08474	0.07864
2	19.69	21.25	130.00	1203.0	0.10960	0.15990
3	11.42	20.38	77.58	386.1	0.14250	0.28390
4	20.29	14.34	135.10	1297.0	0.10030	0.13280

```
[1400, 1600, 1700, 1875, 1100, 1550, 2350, 2450, 1425, 1700]
[[1400]
 [1600]
 [1700]
 [1875]
 [1100]
 [1550]
 [2350]
 [2450]
 [1425]
 [1700]]
```



Linear Regression Demo

```
#import sklearn from scikit-learn
from sklearn import linear_model
```

```
#load dataset
```

```
house_price = [245,312,279,308,199,219,405,324,319,255]
size = [1400,1600,1700,1875,1100,1550,2350,2450,1425,1700]
```

```
#reshape the input regression
```

```
size2= np.array(size).reshape(-1,1)
print(size)
print(size2)
```

```
[7] size_new = 1550
    price = (size_new*regr.coef_)+regr.intercept_
    print(price)
    print(regr.predict([[size_new]]))
```

```
[268.38832326]
[268.38832326]
```

```
from sklearn.metrics import r2_score
from sklearn.metrics import mean_squared_error
score_r2 = r2_score(house_price, regr.predict(size2))
print('R square :',score_r2*100)
print('MSE :', mean_squared_error(house_price,regr.predict(size2)))
```

R square : 58.08173118722721

MSE : 1366.5565224307989

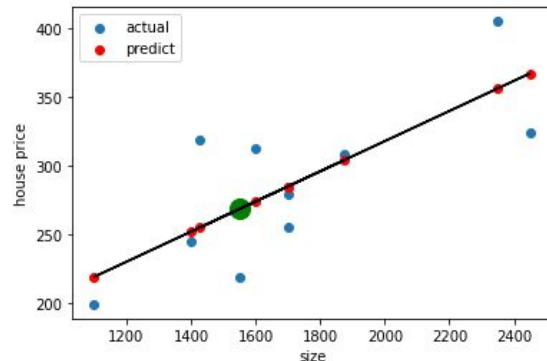
```
[6] #by using module in linear_model to fit the data
    regr = linear_model.LinearRegression()
    regr.fit(size2,house_price)
    print('Coefficient: \n', regr.coef_)
    print('intercept: \n',regr.intercept_)
```

```
Coefficient:
[0.10976774]
intercept:
98.24832962138078
```

```
#plotting the prediction
```

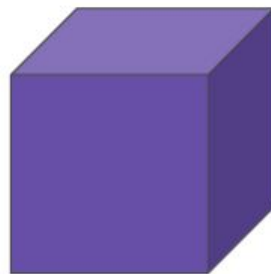
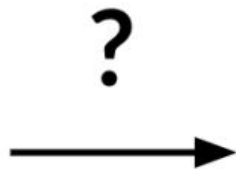
```
plt.scatter(x=size2,y=house_price,label='actual')
plt.scatter(x=size2,y=regr.predict(size2),label='predict',color='red')
plt.scatter(x=size_new,y=price,s=200,color='green')
plt.legend(loc="upper left")
plt.xlabel('size')
plt.ylabel('house price')
x=size2
plt.plot(x, (x*regr.coef_)+regr.intercept_,color='black')
```

[<matplotlib.lines.Line2D at 0x7fe696334240>]

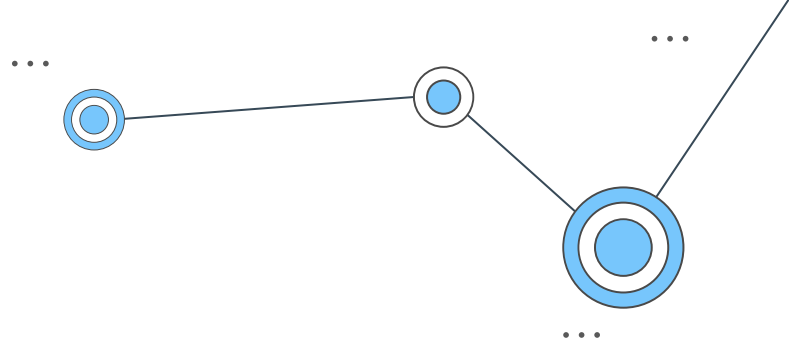




Historical data



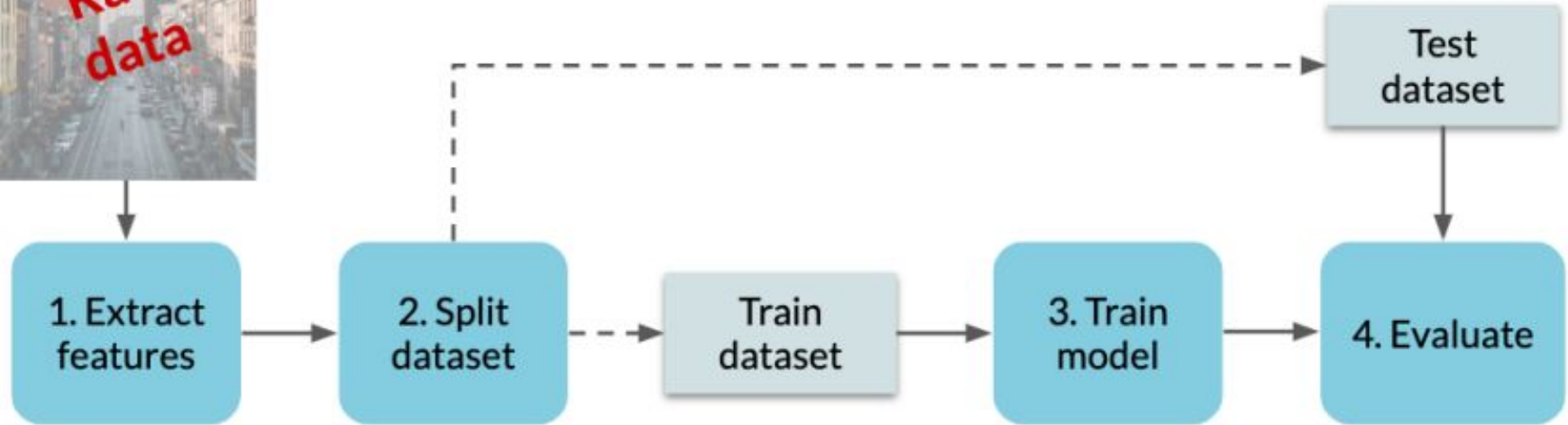
Machine learning
model



Machine Learning Workflow

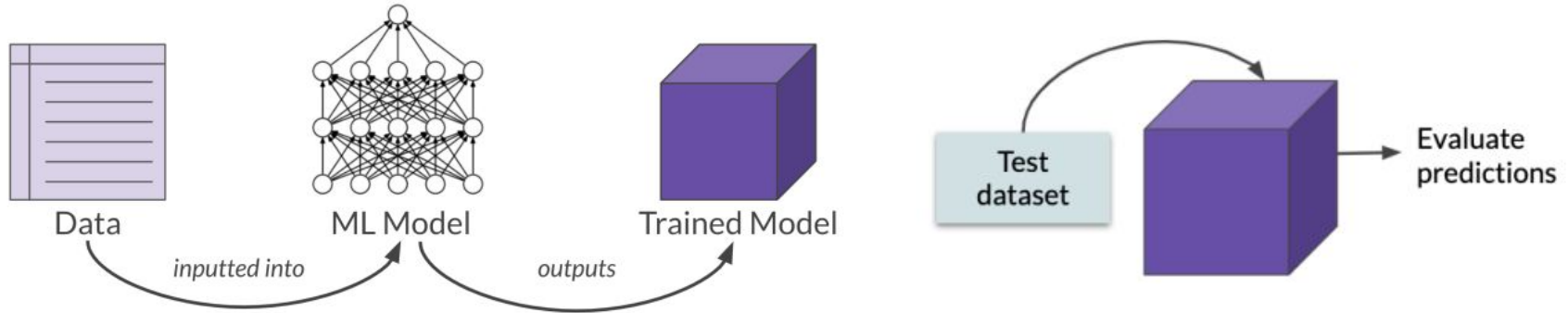
How the beginners to
start machine
learning model?

Machine Learning Workflow



Step Train Model → Evaluate

- Purpose of splitting datasets into train and test set to **estimate how well the learned model will generalize new data**



Is performance good enough?, if yes : this is good method
If no, change model's options, remove features, or random train dataset again

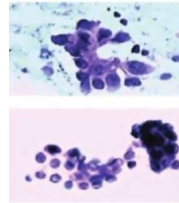
Let's Demo on google colab

DATASET IN MACHINE LEARNING TERMS

2. DATASET IN ML TERMS

INPUT: 30 FEATURES

TARGET CLASS: 2



RADIUS
TEXTURE
PERIMETER
AREA
SMOOTHNESS

CLASSIFIER

MALIGNANT

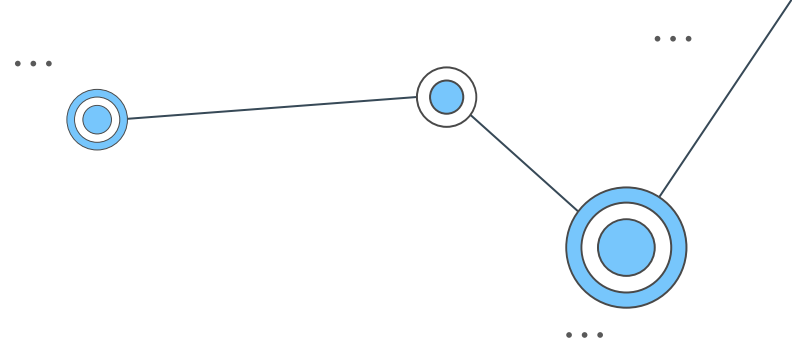
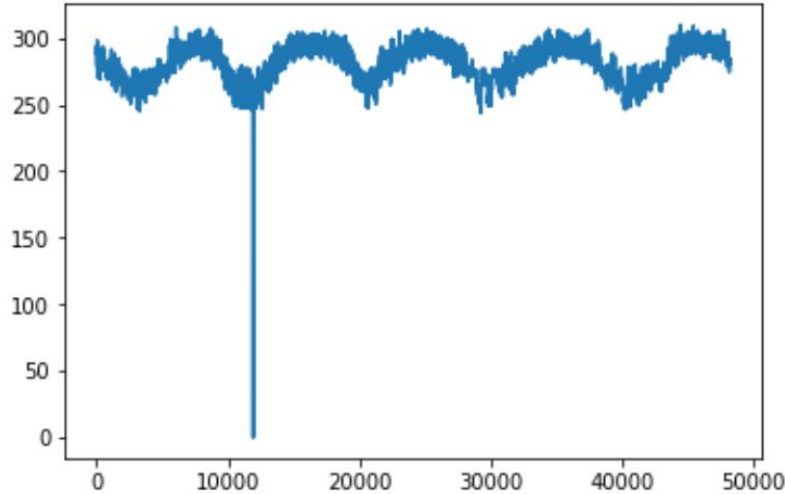
BENIGN

Dataset:

- Number of Instances: 569
- Class Distribution: 212 Malignant, 357 Benign

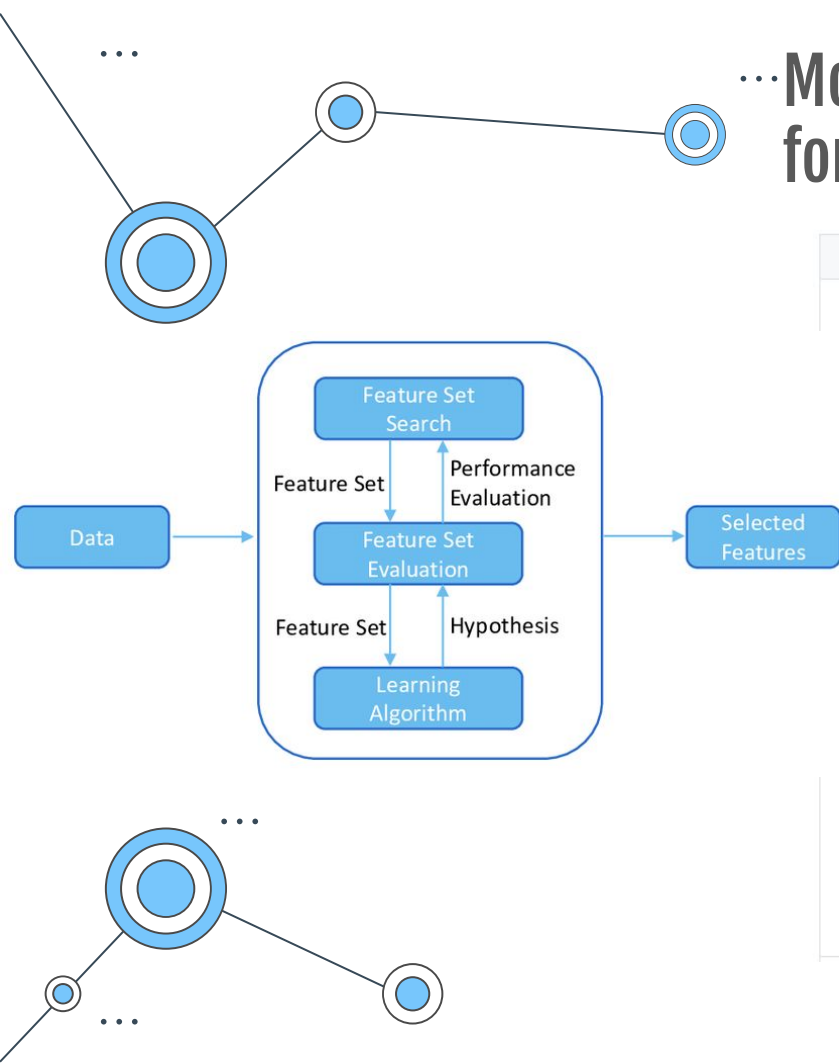
- Data source: [https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic))
- https://www.researchgate.net/publication/271907638_Breast_Cancer_Detection_with_Reduced_Feature_Set

- Preprocessing data before modeling
- Variable Selection
- Overfitting or underfitting results

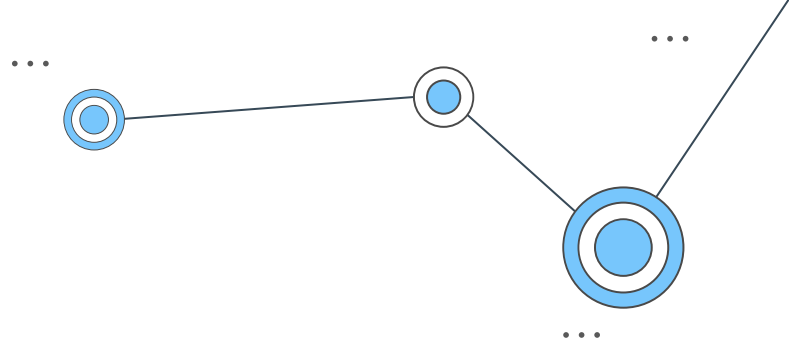


Modeling machine learning constraint for beginners

...Modeling machine learning constraint for beginners



	Underfitting	Just right	Overfitting
Symptoms	<ul style="list-style-type: none">• High training error• Training error close to test error• High bias	<ul style="list-style-type: none">• Training error slightly lower than test error	<ul style="list-style-type: none">• Very low training error• Training error much lower than test error• High variance
Regression illustration			
Classification illustration			



- Study statistics method
- Learn programming language
- Understand basic of machine learning
- Understand Exploratory Data Analysis

**Tips and Tricks to
learn machine
learning**

MACHINE LEARNING



Thankyou



<https://www.linkedin.com/in/regita-putri-permata/>