

# Data Science

Created by Siti Lujeng Luthfiyah



# Table Of Content

01 - Introduction

02- Study Case

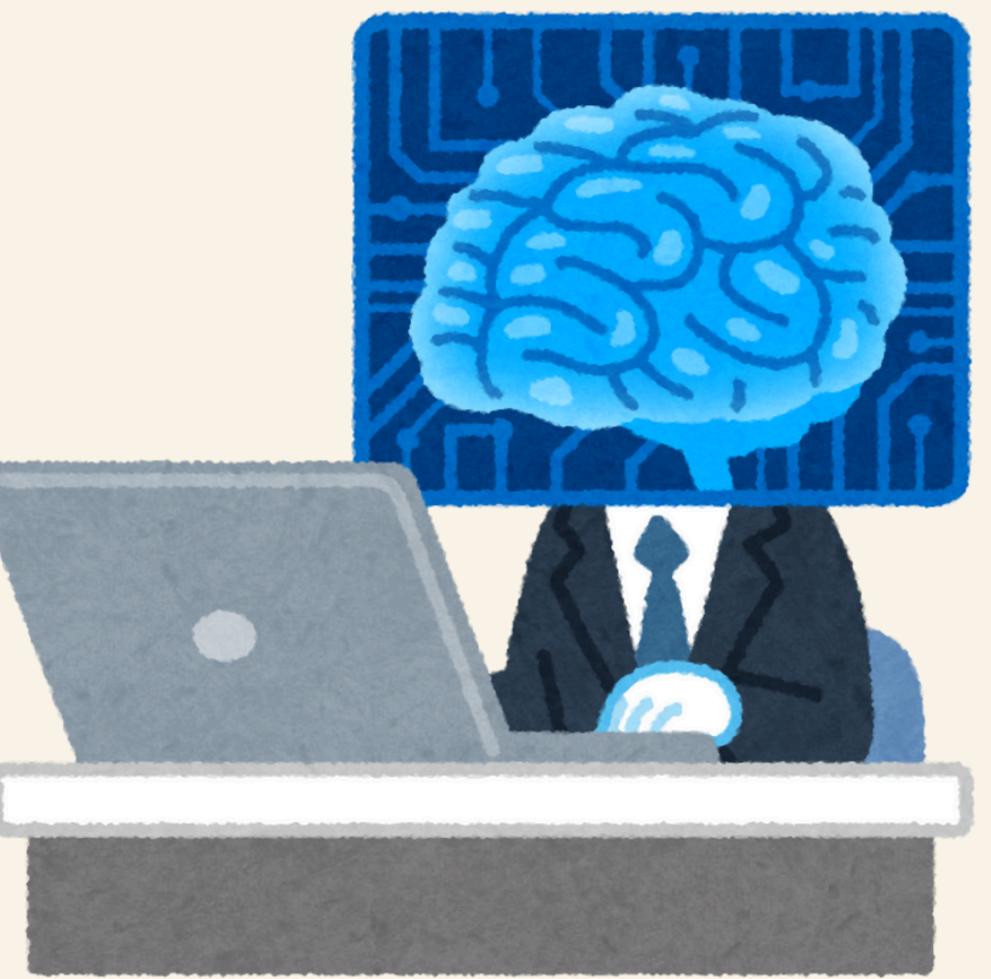
03 - The Dataset

04 - Machine Learning  
Algorithms

05 - Summary

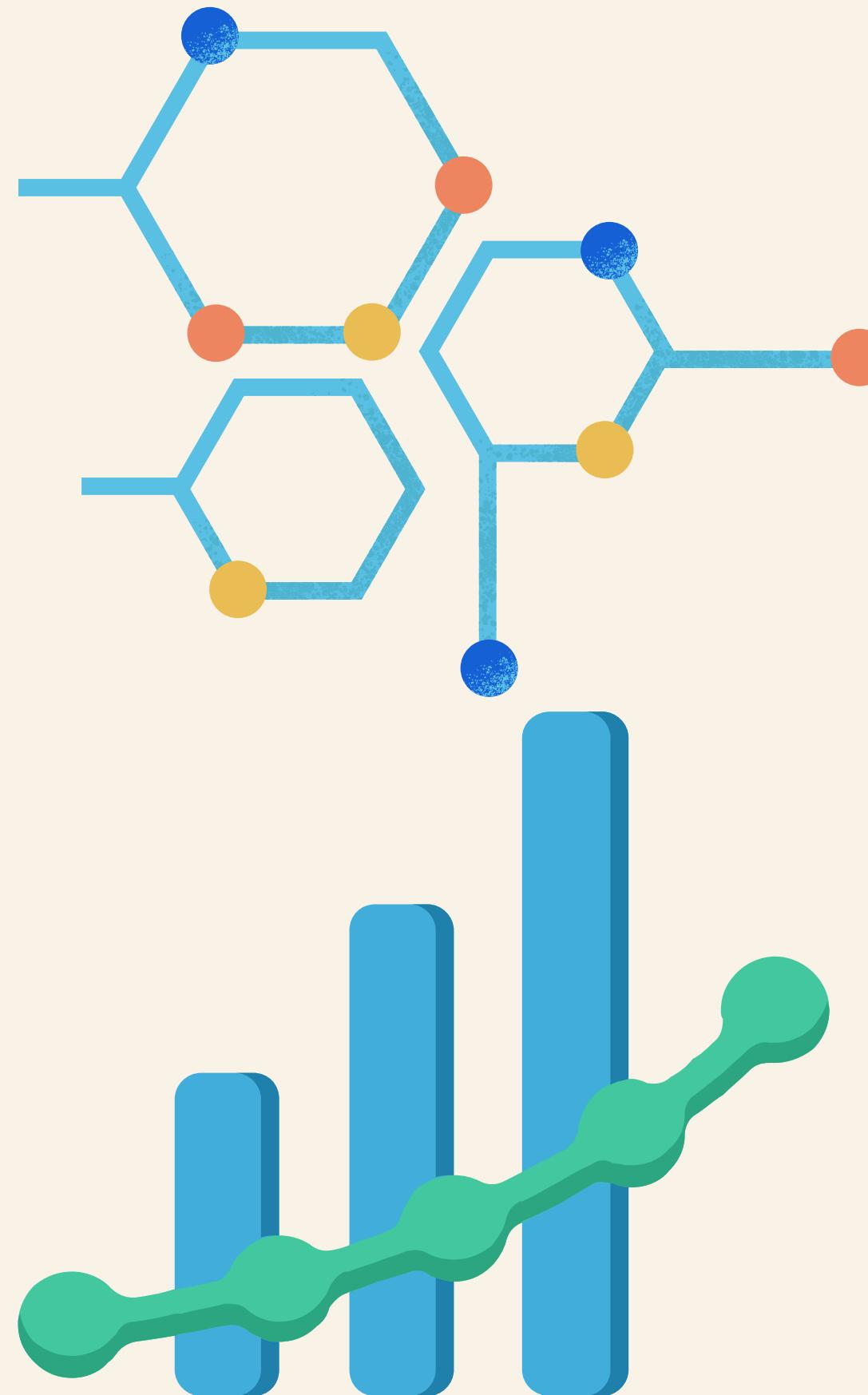


# 01 - Introduction



# 01 - Introduction

Data science is a multidisciplinary field that focuses on the identification of patterns and other insights in big, raw, or structured data collections to extract useful information. The field primarily looks for solutions using algorithms and scientific processes. In actual fact, data science analytics looks for patterns in vast volumes of raw data to help answer more specific queries. It also looks for links or correlations between different datasets to help solve problems that haven't been considered.



source

<https://alison.com/topic/learn/146578/introduction-to-data-sciences-and-data-analytics>

## 02 - Case Study



## 02 - Case Study

### Excercise

Buatlah sebuah Progam Machine Learning (Classification) sederhana menggunakan dataset yang sudah disediakan oleh scikit-learn.

Anda bisa engaksesnya pada link berikut :

[https://scikit-learn.org/1.5/datasets/toy\\_dataset.html](https://scikit-learn.org/1.5/datasets/toy_dataset.html)



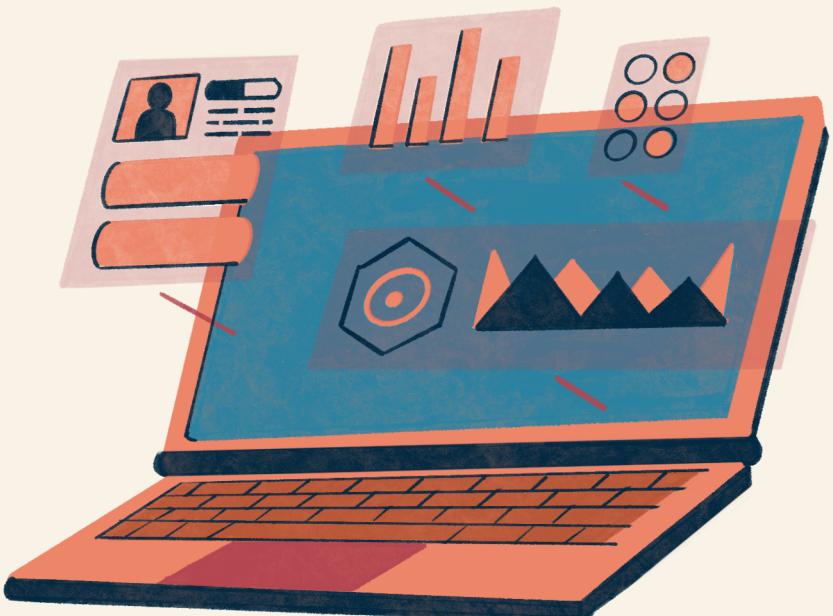
## 03 - The Dataset



# 03 - The Dataset

[https://scikit-learn.org/1.5/modules/generated/sklearn.datasets.load\\_iris.html#sklearn.datasets.load\\_iris](https://scikit-learn.org/1.5/modules/generated/sklearn.datasets.load_iris.html#sklearn.datasets.load_iris)

[https://en.wikipedia.org/wiki/Iris\\_flower\\_data\\_set#Use\\_of\\_the\\_data\\_set](https://en.wikipedia.org/wiki/Iris_flower_data_set#Use_of_the_data_set)



```
# Deskripsi dataset
print(iris.DESCR)
```

:Number of Instances: 150 (50 in each of three classes)  
:Number of Attributes: 4 numeric, predictive attributes and the class  
:Attribute Information:  
- sepal length in cm  
- sepal width in cm  
- petal length in cm  
- petal width in cm  
- class:  
- Iris-Setosa  
- Iris-Versicolour  
- Iris-Virginica

:Summary Statistics:

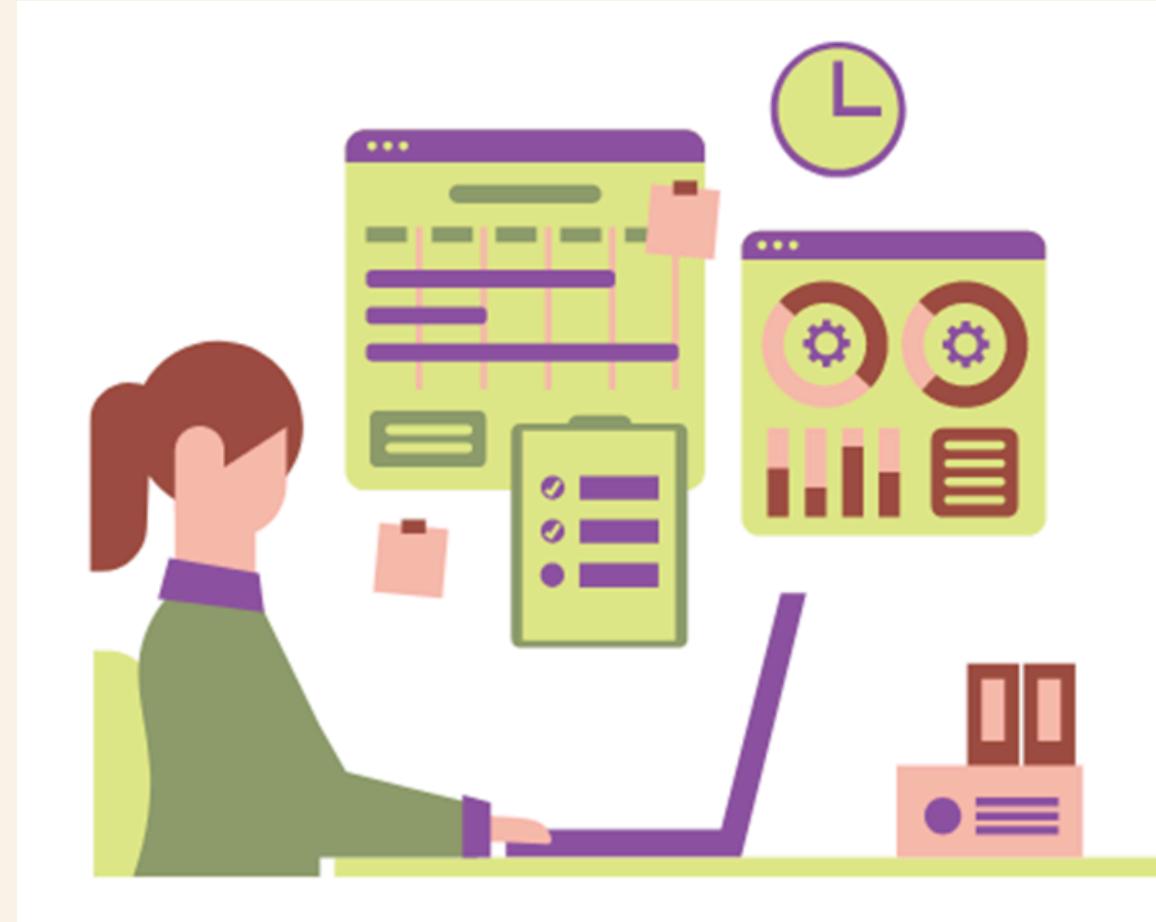
	Min	Max	Mean	SD	Class Correlation
sepal length:	4.3	7.9	5.84	0.83	0.7826
sepal width:	2.0	4.4	3.05	0.43	-0.4194
petal length:	1.0	6.9	3.76	1.76	0.9490 (high!)
petal width:	0.1	2.5	1.20	0.76	0.9565 (high!)

:Missing Attribute Values: None  
:Class Distribution: 33.3% for each of 3 classes.  
:Creator: R.A. Fisher  
:Donor: Michael Marshall ([MARSHALL%PLU@io.arc.nasa.gov](mailto:MARSHALL%PLU@io.arc.nasa.gov))  
:Date: July, 1988

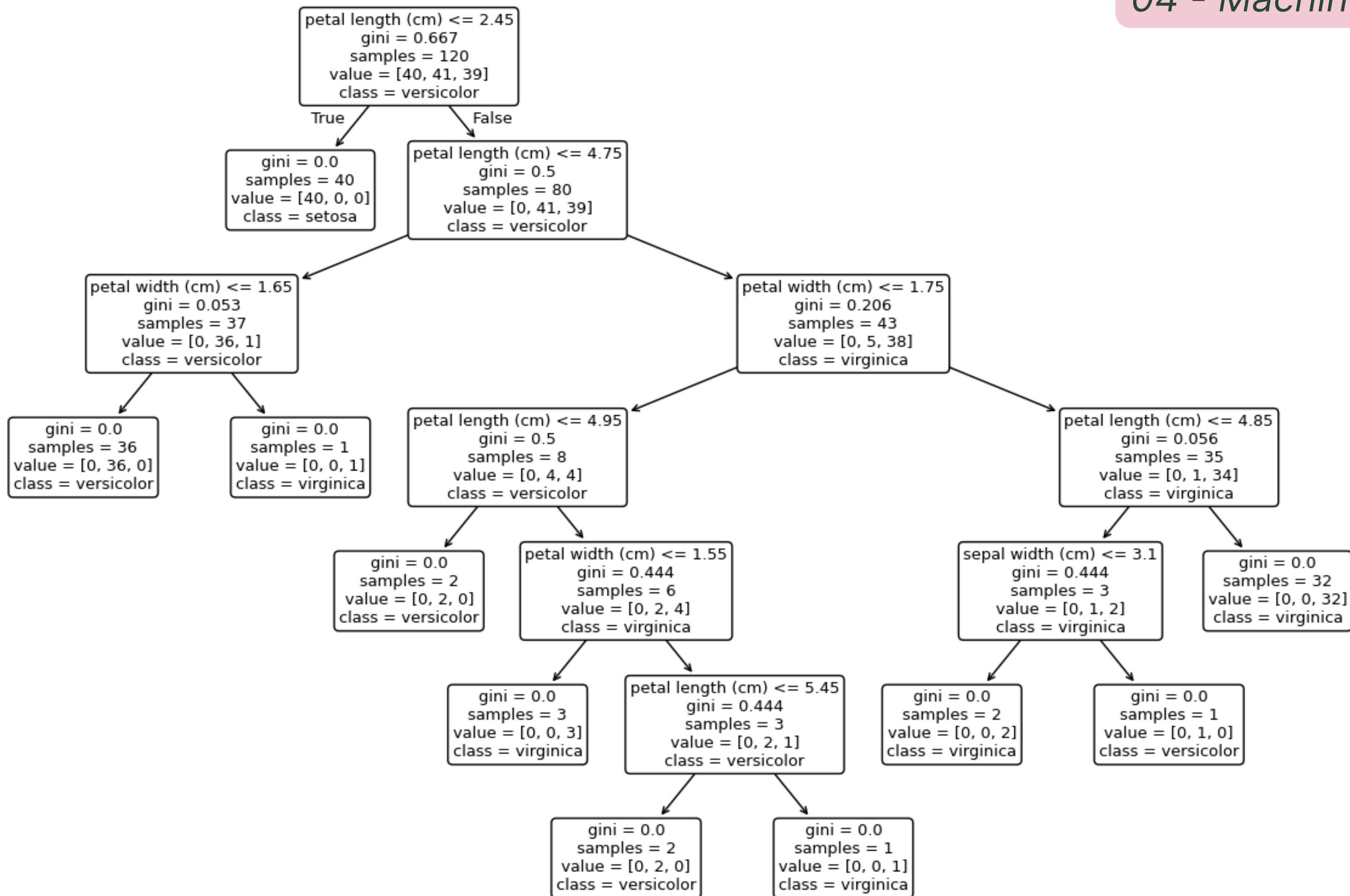
The famous Iris database, first used by Sir R.A. Fisher. The dataset is taken from Fisher's paper. Note that it's the same as in R, but not as in the UCI Machine Learning Repository, which has two wrong data points.

This is perhaps the best known database to be found in the pattern recognition literature. Fisher's paper is a classic in the field and is referenced frequently to this day. (See Duda & Hart, for example.) The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. One class is linearly separable from the other 2; the latter are NOT linearly separable from each other.

## *04 - Machine Learning Algorithms*

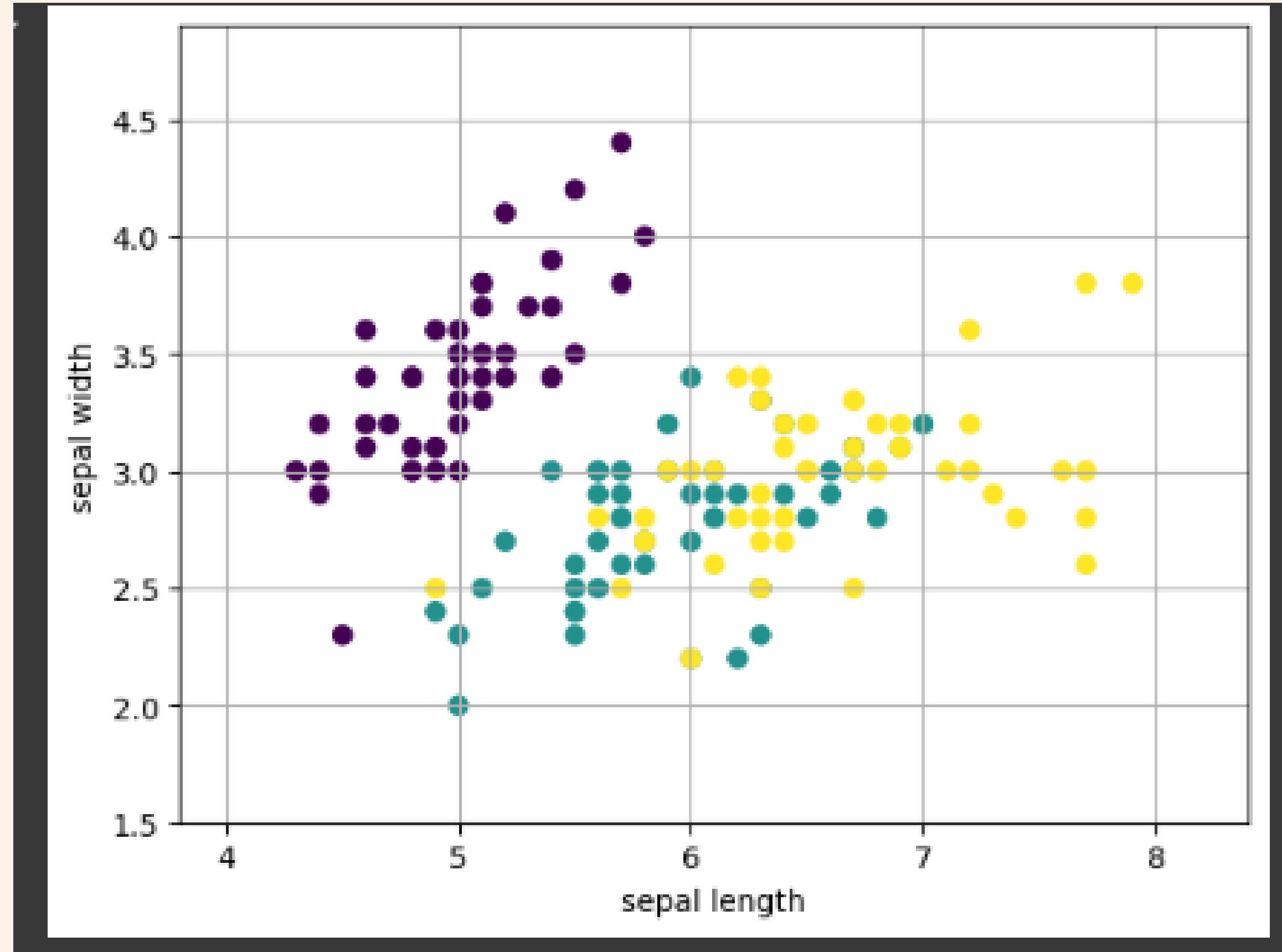


## Decision Tree Classifier



# 04 - Machine Learning Algorithms

## Logistic Regression



## 05 - Summary

Data science and machine learning use data to improve products, services, infrastructure, etc. Both are high-demand, high-paying careers. The relation between data science and machine learning is that of rectangles and squares, where both shapes relate and are yet still distinct. In our analogy, data science can be exemplified as a rectangle, while the square can be considered a separate entity. Both fields carry great importance for data scientists, and they are helping all industries adopt these frameworks efficiently.

- Data Science: It studies data and works on extracting meaningful insights from it.
- Machine Learning: Machine learning is a branch of Artificial Intelligence. It helps understand and build methods for utilizing data, improving performance, and making predictions.

Recently, data analytics, data sciences, and business intelligence have increased their reliance on machine learning and AI. Machine learning automates data analysis and data collection with the help of models and algorithms to make predictions.

# Thank You For Watching



Siti Lujeng Luthfiyah



@luthfiyahsl

