

Lokakarya SuperKilat:

Pengenalan Ringkas

Machine Learning untuk Sains Material

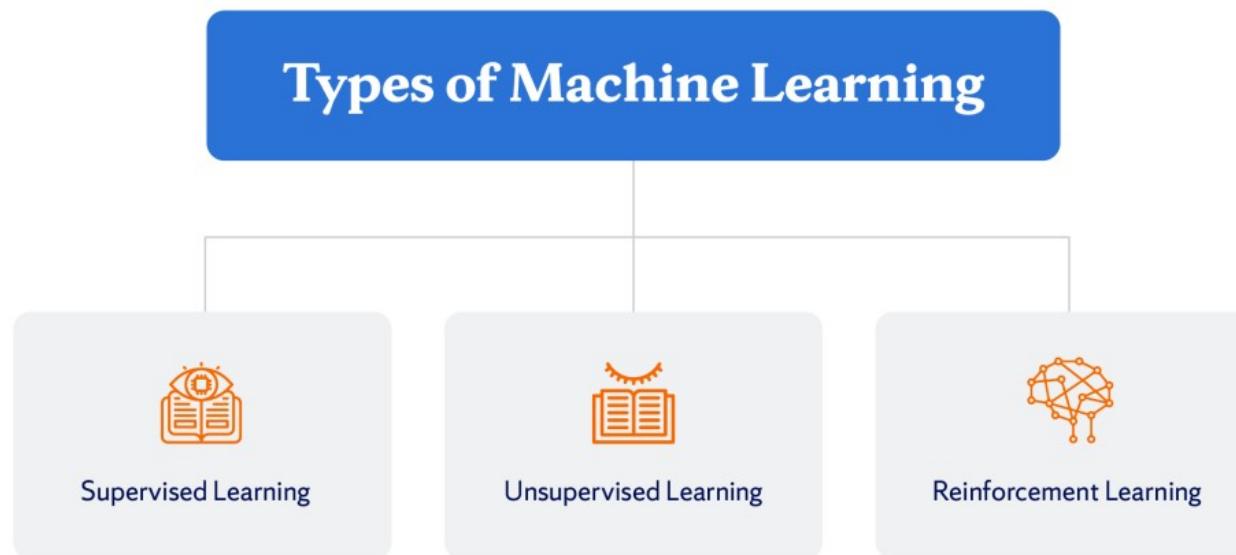
4 Desember 2025

Miftah Hadi Syahputra Anfa

Machine Learning = Mesin Belajar?

- Masalah yang ingin diselesaikan dengan **ML**:
 - **Regression**
 - **Classification**
 - ***Clustering***

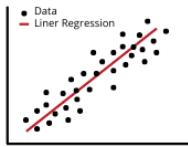
Macam Algoritma ML



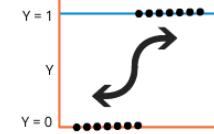
Top 8 Machine Learning Algorithms

explained in less than 1 minute each

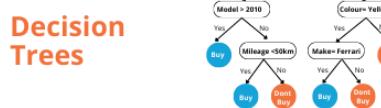
Linear Regression



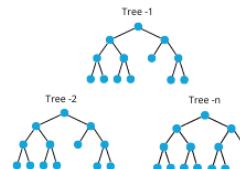
Logistic Regression



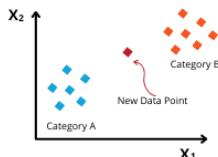
Decision Trees



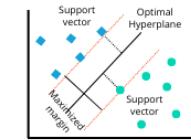
Random Forest



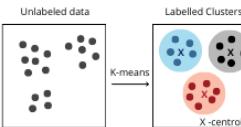
K-Nearest Neighbor



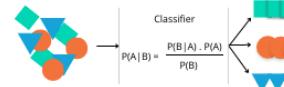
Support Vector Machine



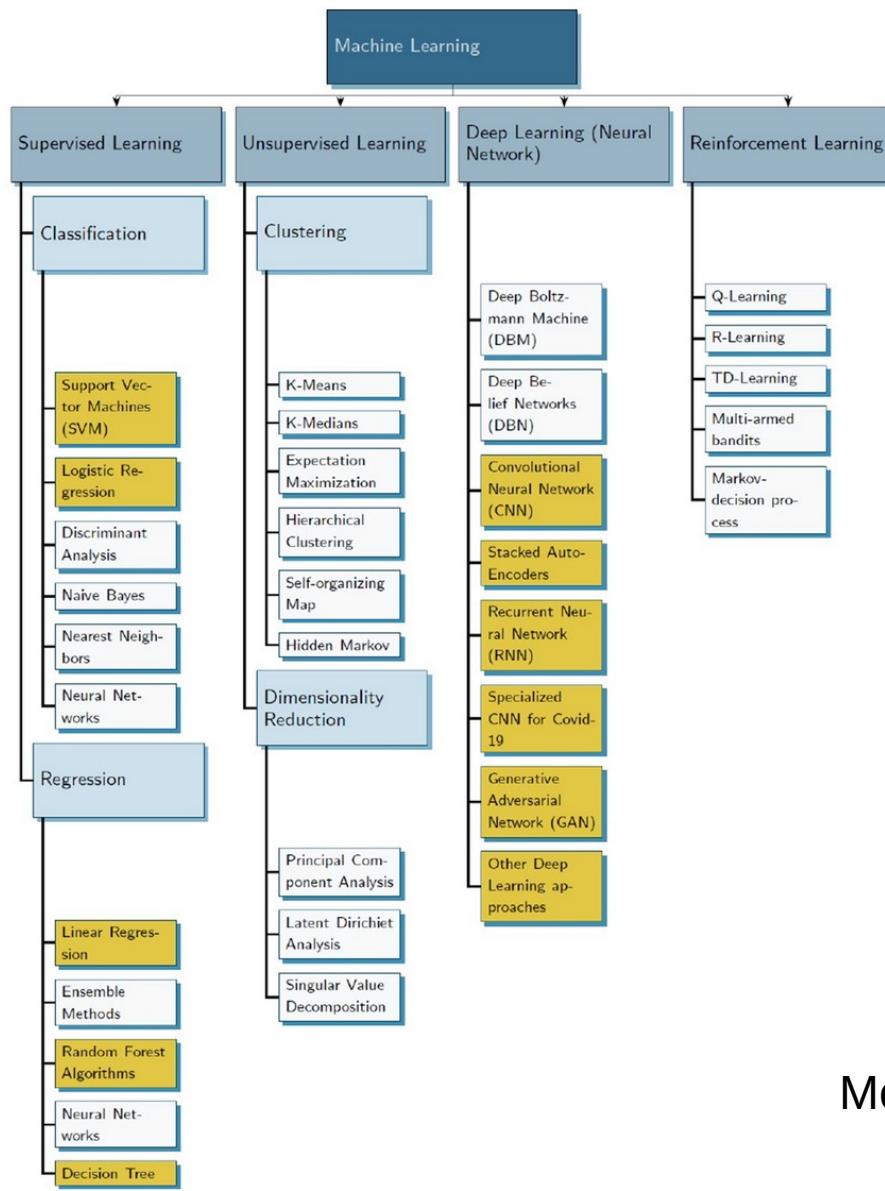
K-Means Clustering



Naïve Bayes

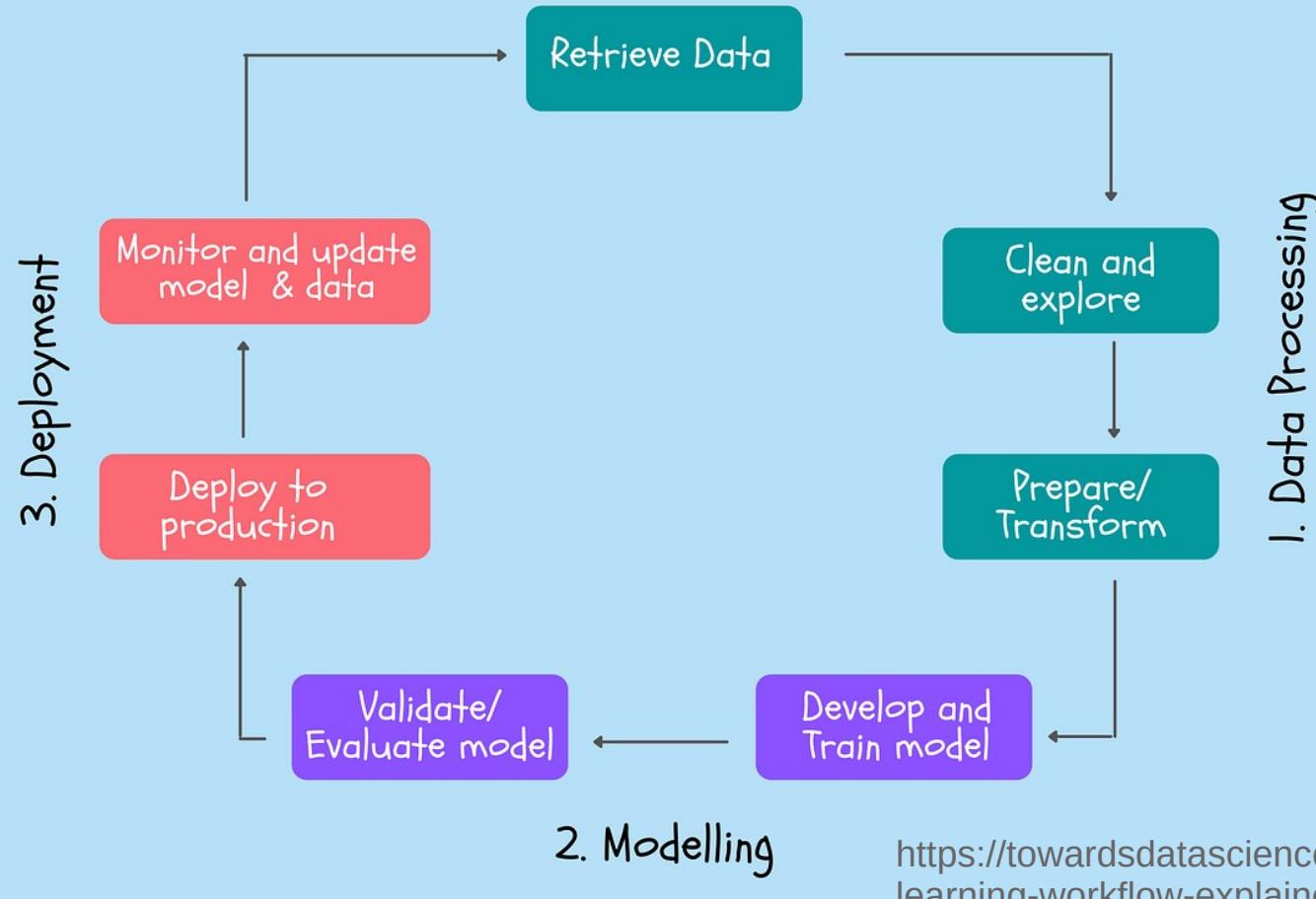


<https://datasciencedojo.com/blog/machine-learning-algorithms-explanation/>





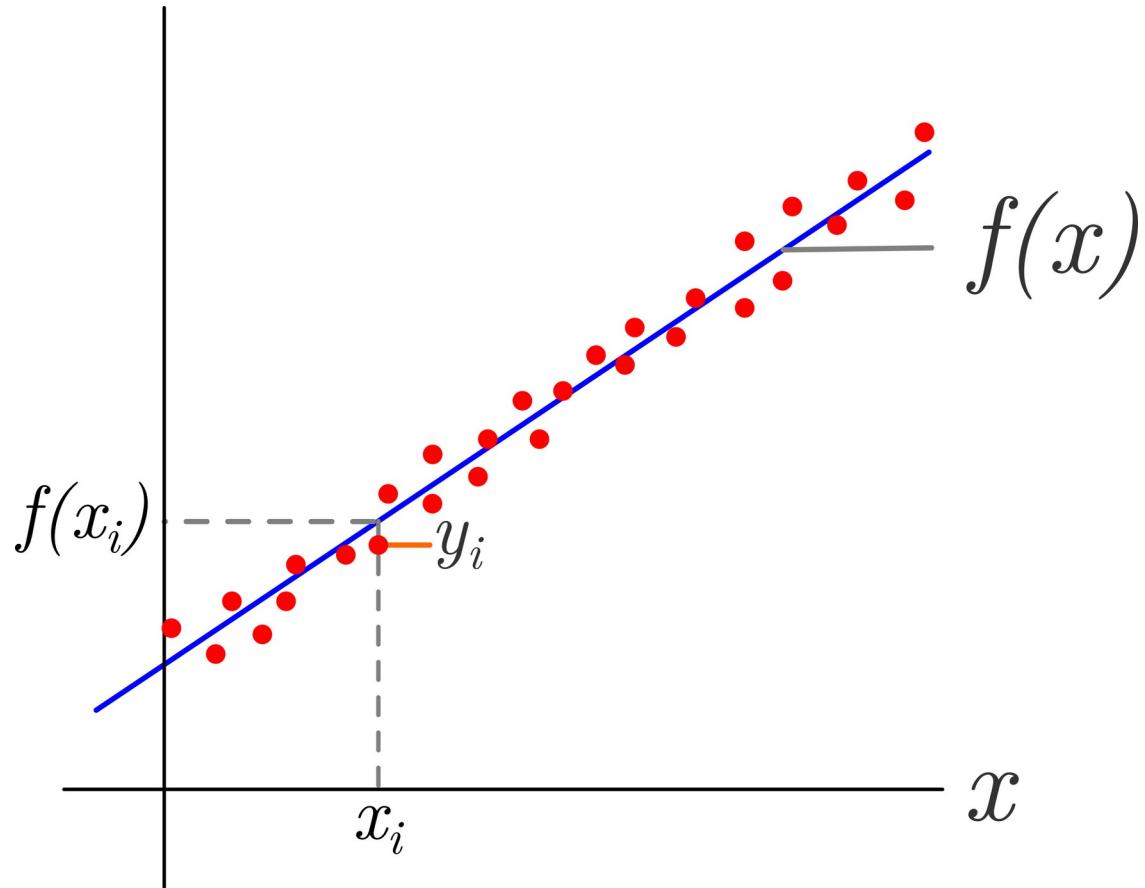
Machine Learning Workflow



2. Modelling

<https://towardsdatascience.com/the-machine-learning-workflow-explained-557abf882079/>

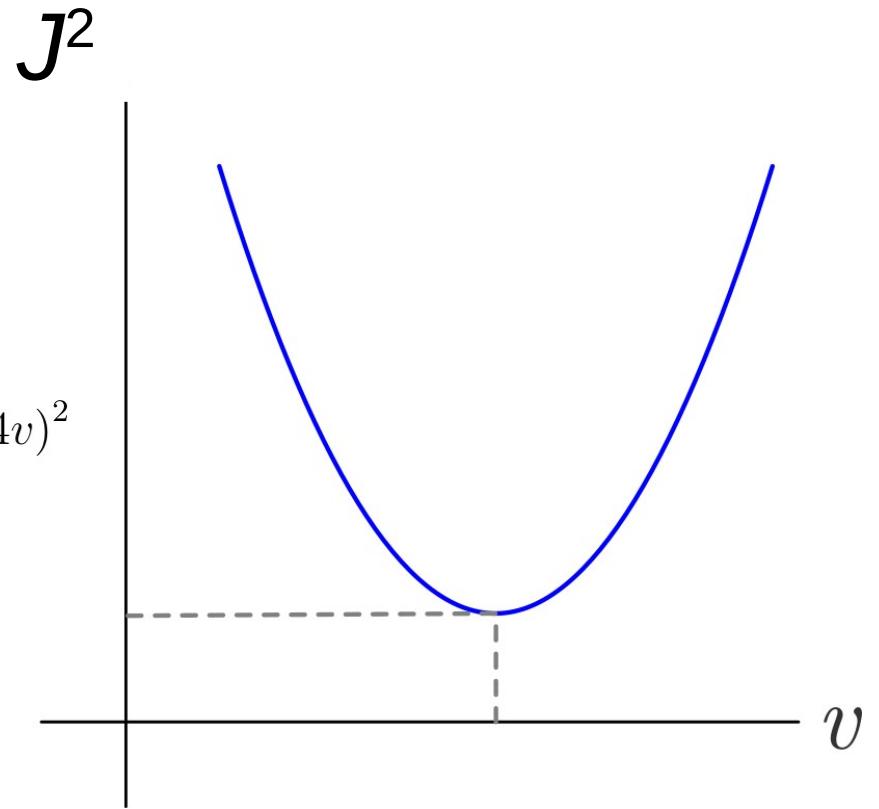
Regresi Linier



Regresi Linier

$$J^2 = \sum_{i=1}^N [y_i - f(x_i)]^2$$

$$\begin{aligned} J^2 &= (4.7 - v)^2 + (9.2 - 2v)^2 + (13.7 - 3v)^2 + (18.1 - 4v)^2 \\ &\quad + (22.7 - 5v)^2 + (27.2 - 6v)^2, \end{aligned}$$



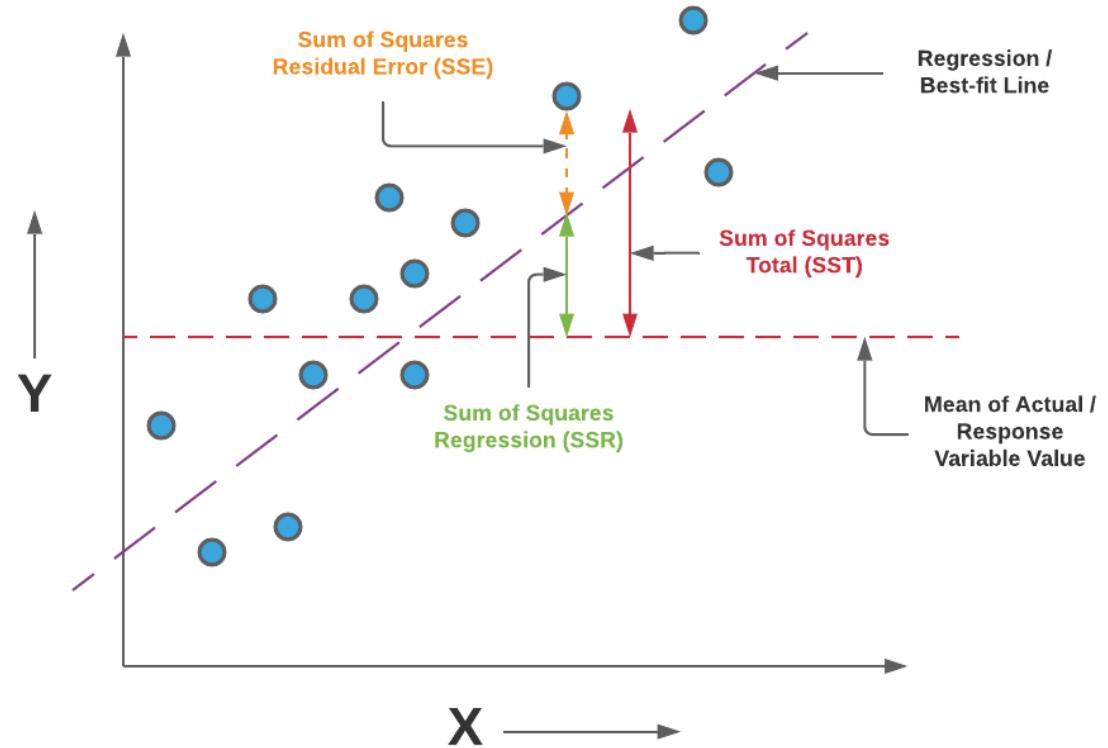
Dataset untuk Model ML

- Bagi data jadi 2: **training** dan **test**
- Pisahkan mana *feature* (X) mana *target* (y)

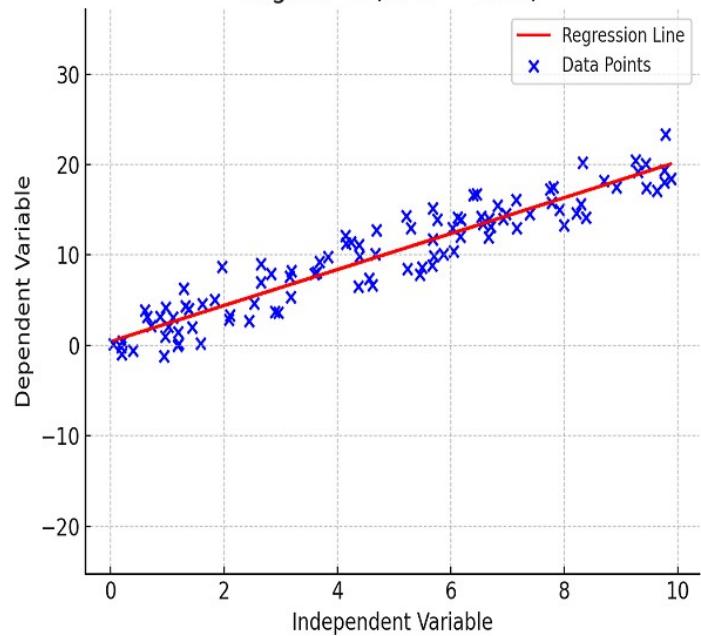
Performa Model

$$R^2 = \frac{SSR}{SST} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}$$

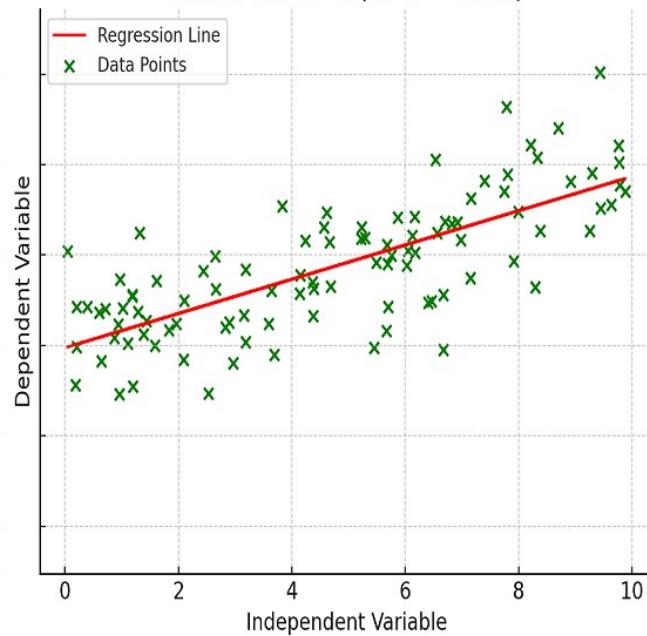
$$\text{RMSE}(\vec{y}, \hat{y}) = \sqrt{\frac{1}{m} \sum_{i=1}^m (\hat{y}_i - y_i)^2}$$



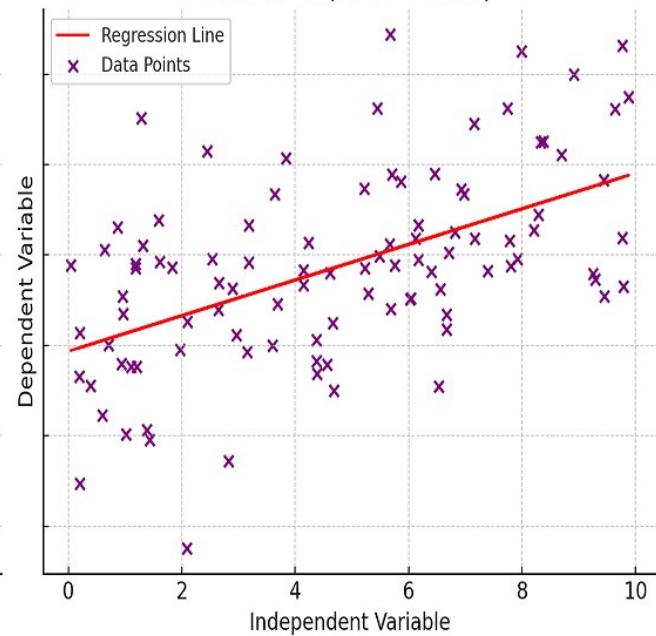
High R² (R² = 0.89)



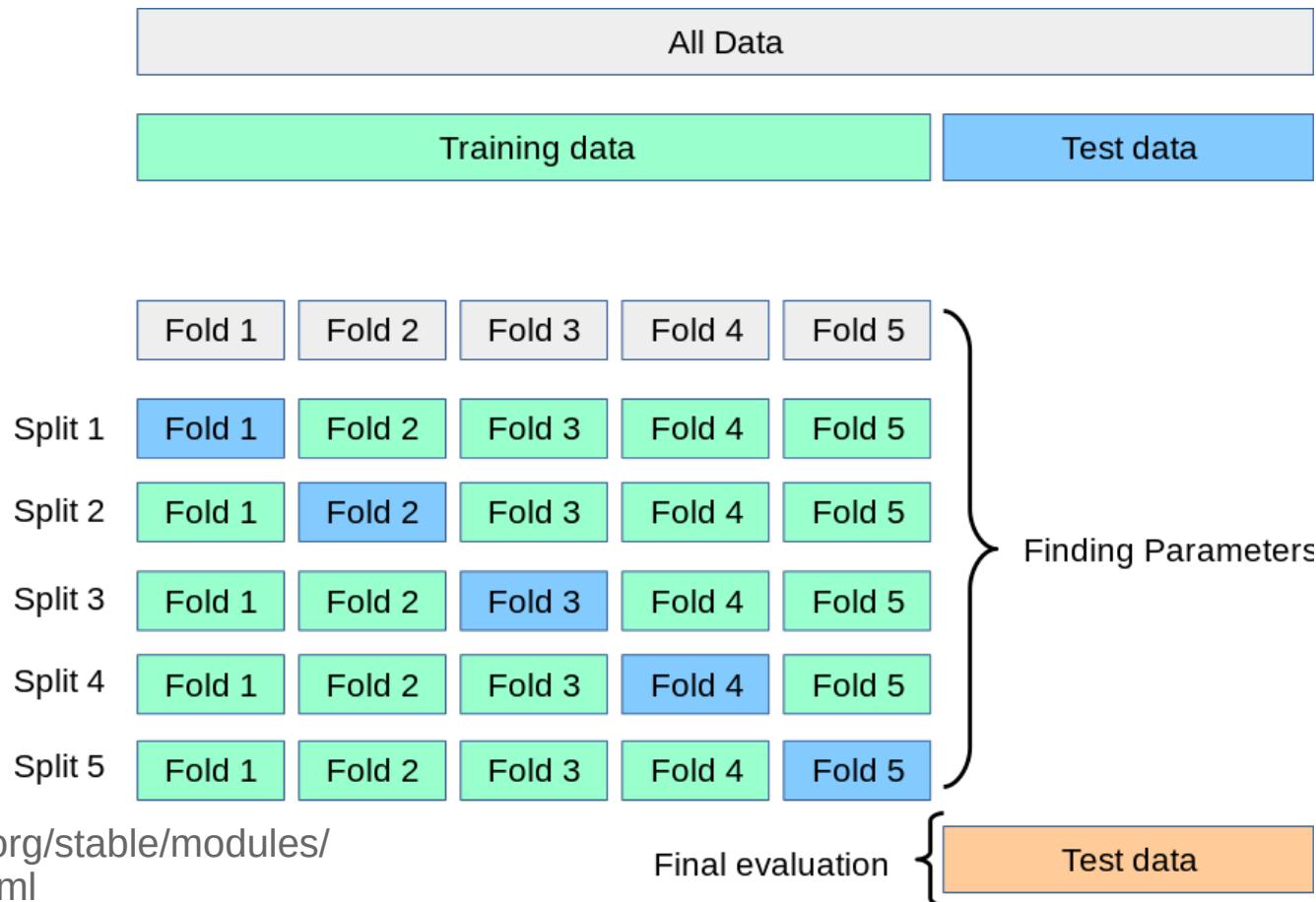
Moderate R² (R² = 0.55)



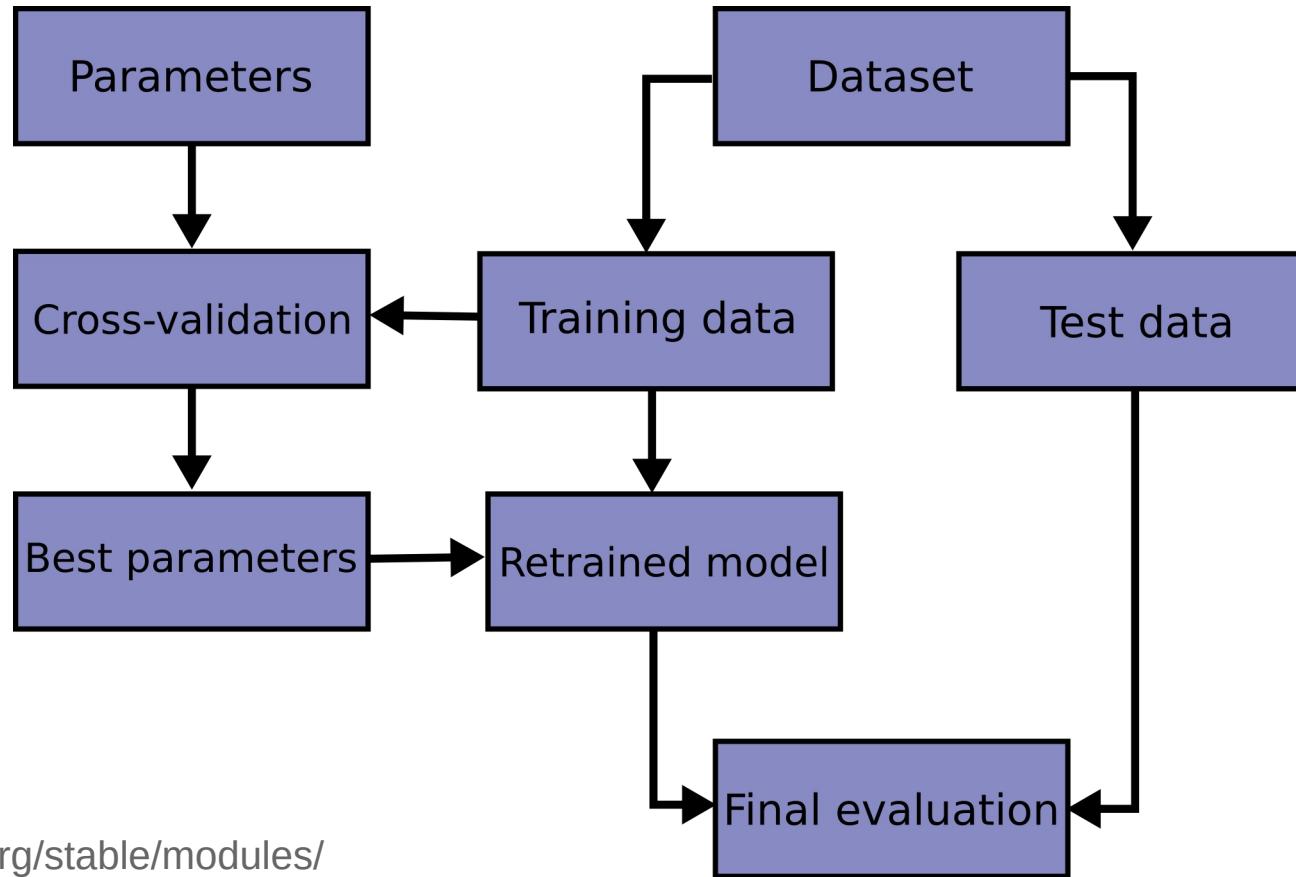
Low R² (R² = 0.28)



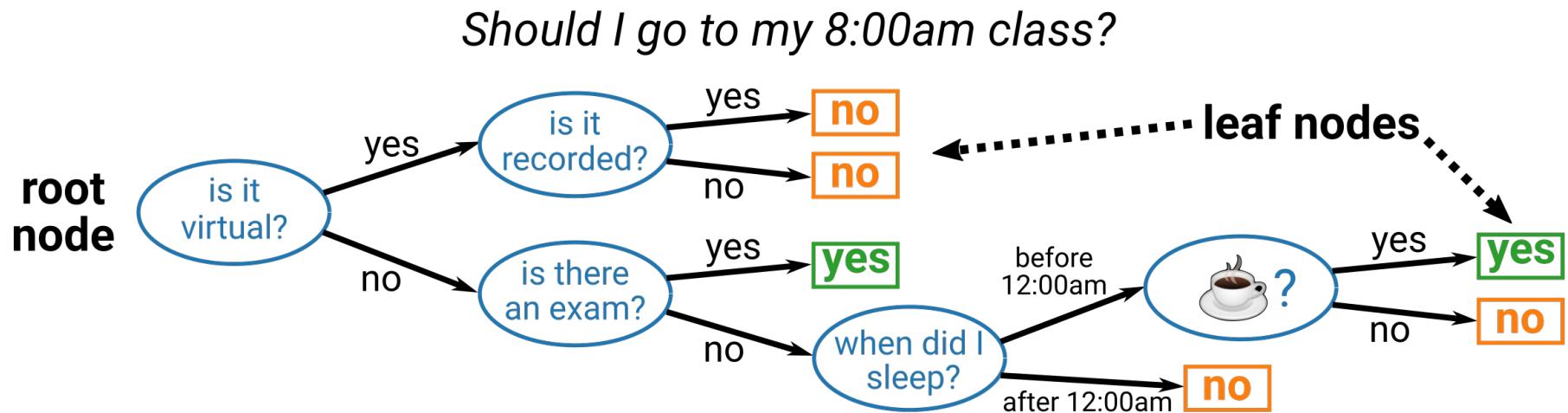
K-Fold Cross-Validation



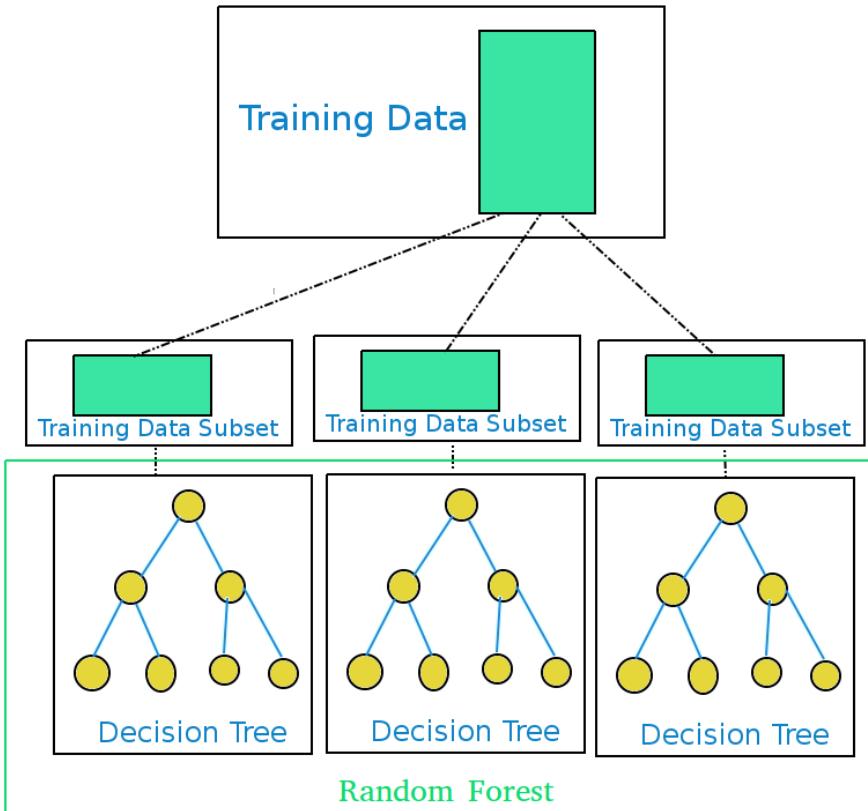
K-Fold Cross-Validation



Decision Tree



Random Forest



Performa Classifier

	Positive prediction	Negative prediction
--	---------------------	---------------------

<i>Positive label:</i>	True positive (TP)	False negative (FN)
------------------------	--------------------	---------------------

<i>Negative label:</i>	False positive (FP)	True negative (TN)
------------------------	---------------------	--------------------

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \in [0, 1]$$

Resources

- Belajar Machine Learning
 - <https://developers.google.com/machine-learning/crash-course>
 - <https://github.com/neonwatty/machine-learning-refined>
 - <https://slds-lmu.github.io/i2ml/>
 - <https://introml.mit.edu/notes/>
 - <https://cburdine.github.io/materials-ml-workshop>
 - <https://www.youtube.com/playlist?list=PLSnC4a32tFDptnjLbnOg-gIyW8iR2H9Ur>
 - https://enze-chen.github.io/mi-book-2021/week_1/04/intro_ml_blank.html

Resources

- Dataset
 - <https://www.kaggle.com/>
 - <https://materialsdata.nist.gov/>
 - <https://matbench-discovery.materialsproject.org/>
 - <https://github.com/IntelLabs/matsciml>
 - <https://pymatgen.org/>
 - <https://aflowlib.org/>
 - <https://www.materialsdatafacility.org/search>
 - <https://huggingface.co/datasets/Allanatrix/Materials>

Resources

- Misc
 - <https://github.com/anthony-wang/BestPractices>
 - <https://aronwalsh.github.io/MLforMaterials>
 - <https://huggingface.co/>
 - <https://enze-chen.github.io/resources/>
 - <https://github.com/jcwang587/cgcnn2>
 - <https://github.com/Tony-Y/cggn?tab=readme-ov-file>
 - <https://github.com/sedaoturak/data-resources-for-materials-science>