

# **Lokakarya SuperKilat:** **Pengenalan Ringkas** *Machine Learning* untuk Sains **Material**

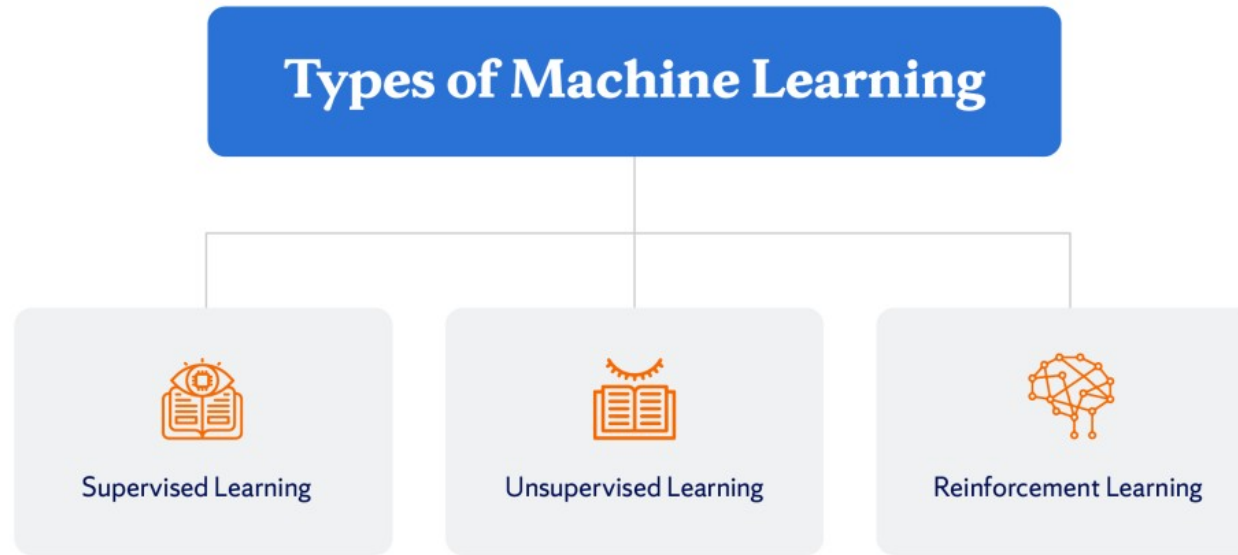
4 Desember 2025

Miftah Hadi Syahputra Anfa

# Machine Learning = Mesin Belajar?

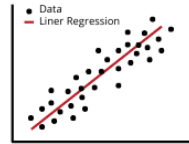
- Masalah yang ingin diselesaikan dengan **ML**:
  - **Regression**
  - **Classification**
  - *Clustering*

# Macam Algoritma ML

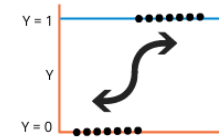


# Top 8 Machine Learning Algorithms explained in less than 1 minute each

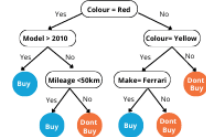
## Linear Regression



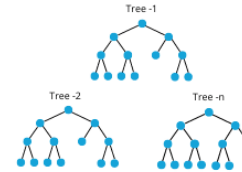
## Logistic Regression



## Decision Trees



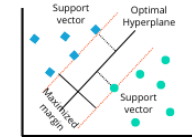
## Random Forest



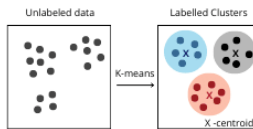
## K-Nearest Neighbor



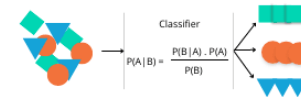
## Support Vector Machine

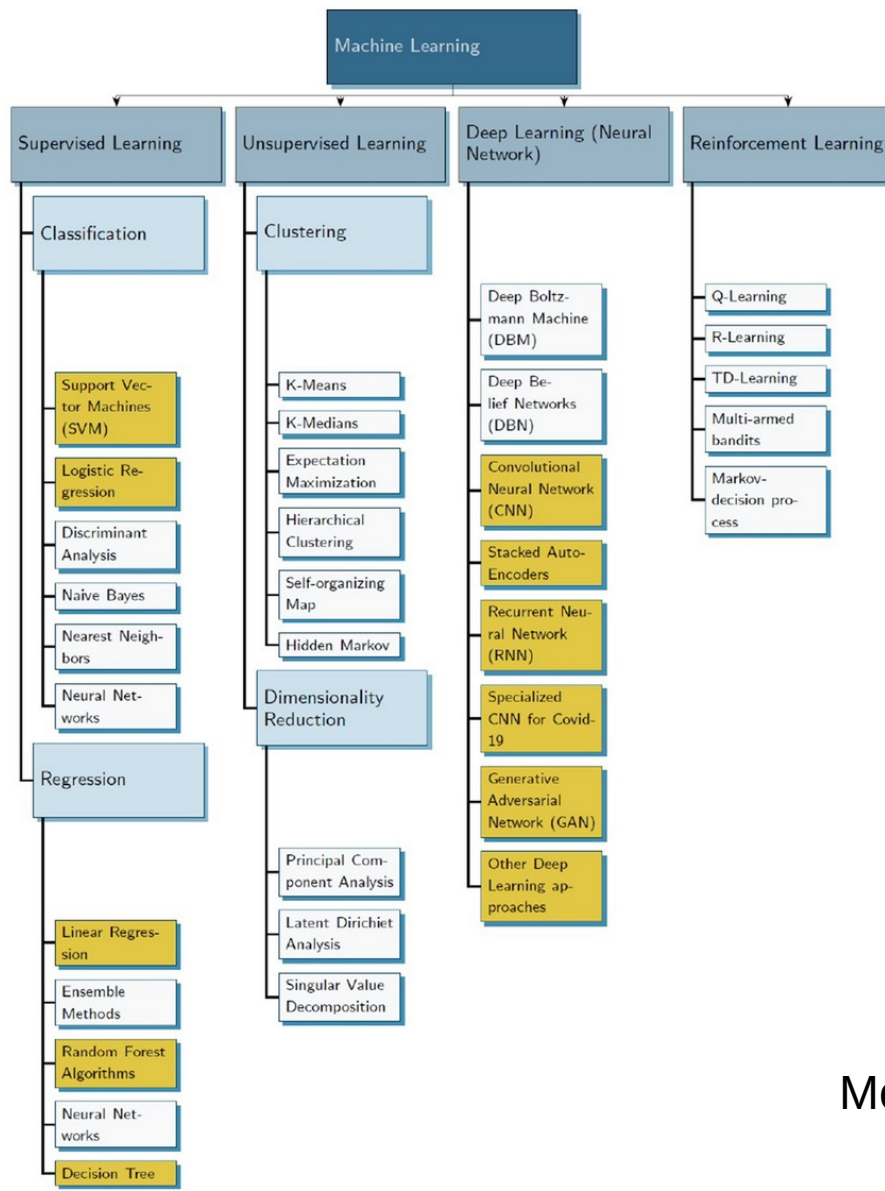


## K-Means Clustering



## Naïve Bayes





# 100+ MACHINE LEARNING ALGORITHMS

By @TheInsaneApp

## CLUSTERING

- k Means
- k Medians
- BIRCH
- Fuzzy C-Means
- Fuzzy K-Modes
- Mini Batch K-Means
- DBSCAN
- Fuzzy clustering
- Mean-shift
- Optics algorithm
- Expectation Maximization
- Hierarchical Clustering
- Minimum spanning tree

## OTHER ML ALGORITHMS (Part 1)

- ALOPEX
- CN2 algorithm
- FastICA
- Feature selection Algorithms
- Linde Buzo Gray Algorithm
- Forward Backward algorithm
- Algorithm Accuracy Evaluation
- Performance Measures
- Optimization Algorithm
- Dynamic Time Warping
- Local Outlier Factor
- Logic Learning Machine
- Markov Chain Monte Carlo
- T Distributed Stochastic neighbor Embedding
- Wake Sleep Algorithm
- Prefrontal Cortex Basal Ganglia Working Memory

## DIMENSIONALITY REDUCTION

- PCA Algorithm
- PCR Algorithm
- PLSR Algorithm
- Sammon Mapping
- MDS Algorithm
- Projection Pursuit
- LDA Algorithm
- ICA Algorithm
- NMF Algorithm
- RDA Algorithm
- MDA Algorithm
- PLSDA Algorithm
- QDA Algorithm
- CCA Algorithm
- Diffusion Map

## RULE SYSTEM

- Cubist
- OneR
- ZeroR
- Ripper

## OTHER ML ALGORITHMS (Part 2)

- LogitBoost
- Sparse PCA
- Structured kNN
- WMA Algorithm
- GeneRec
- Leabra
- RProp

## REGRESSION

- Ordinary Least Squares Regression
- Linear Regression
- Logistic Regression
- Stepwise Regression
- MARS Algorithm
- Locally Estimated Scatterplot Smoothing

## DEEP LEARNING

- CNN
- RNNs
- LSTMs
- Stacked Auto Encoders
- Deep Boltzmann Machine
- Deep Belief Networks

## DECISION TREE

- Conditional Decision Trees
- Decision Stump
- C4.5 and C5.0
- CHAID
- CART
- ID3
- M5

## REGULARIZATION

- Ridge Regression
- Elastic Net
- LASSO
- LARS

## ENSEMBLE

- GBDT
- GBRT
- Boosting
- Bagging
- AdaBoost
- Random Forest
- Blending Algorithm
- Stacked Generalization
- Gradient Boosting Machines

## INSTANCE BASED

- Learning Vector Quantization
- K Nearest Neighbor
- Self-Organizing Map
- Locally Weighted Learning
- Support Vector Machines

## BAYESIAN

- AODE
- Naive Bayes
- Gaussian Naive Bayes
- Multinomial Naive Bayes
- Bayesian Belief Network
- Bayesian Network

## REINFORCEMENT LEARNING

- Q Learning
- SARSA Algorithm
- Deep Q-network
- Learning Automata
- DDPG Algorithm
- NAF Algorithm
- A3C Algorithm
- TRPO Algorithm
- PPO Algorithm
- Constructing Skill Trees

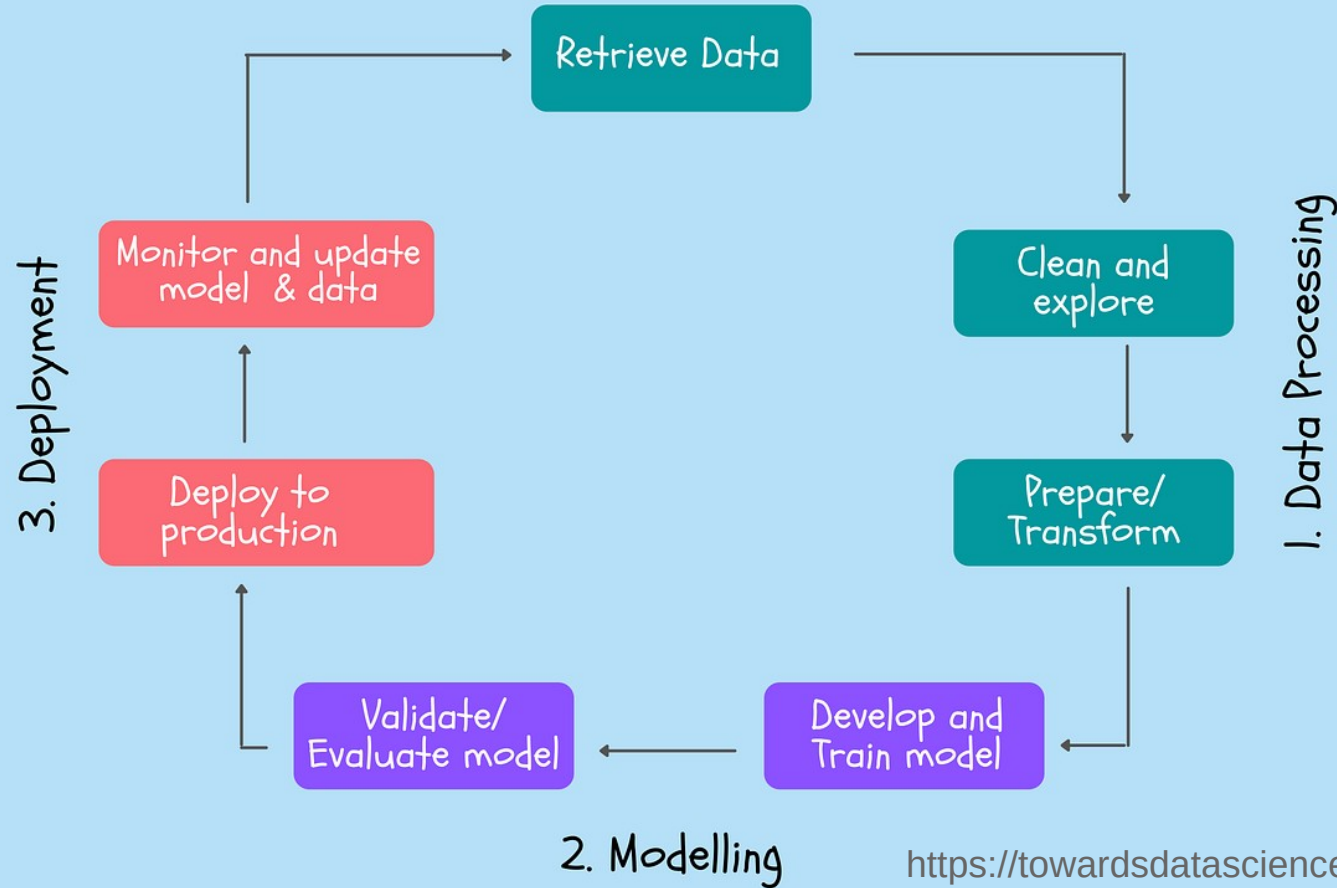
## ANNs

- Perceptron
- MLP Algorithm
- Back-Propagation
- Stochastic Gradient Descent
- Hopfield Network
- RBFN Algorithm

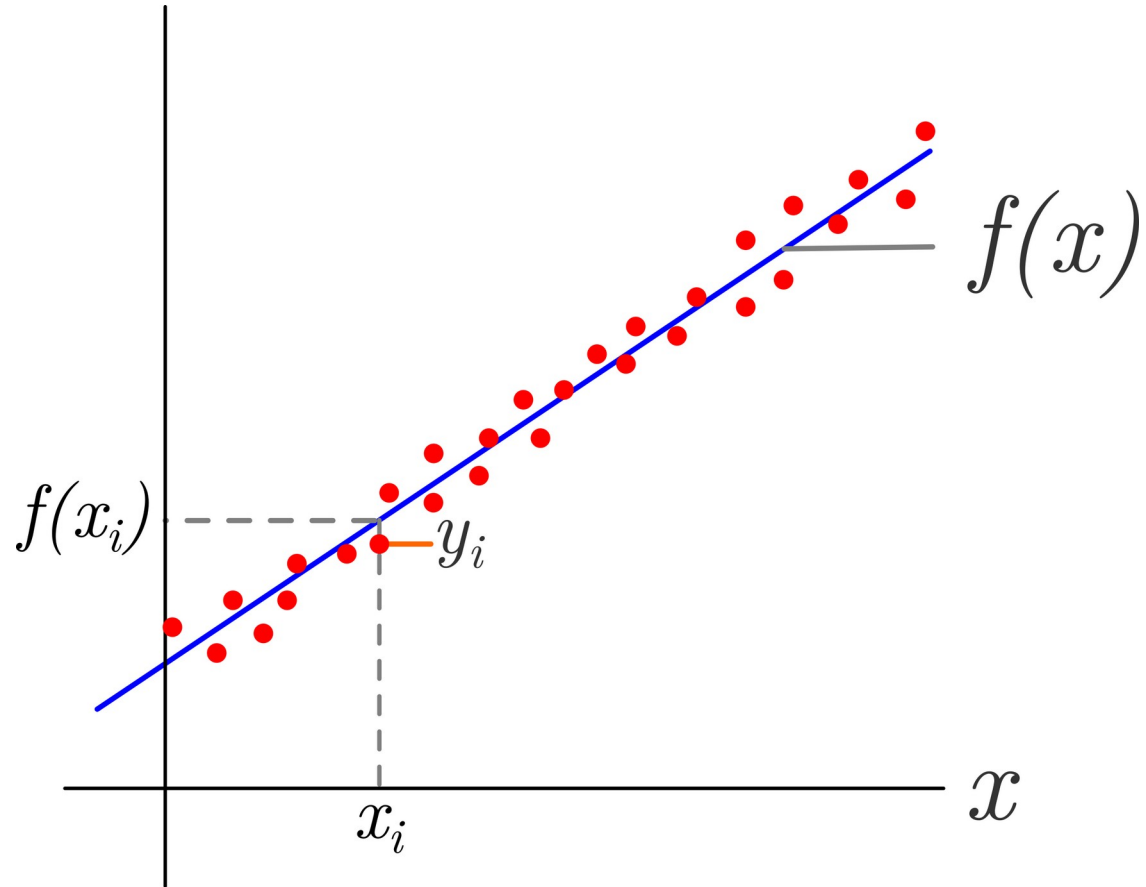
## ARL

- Apriori Algorithm
- Eclat Algorithm
- FP Growth Algorithm

# Machine Learning Workflow



# Regresi Linier

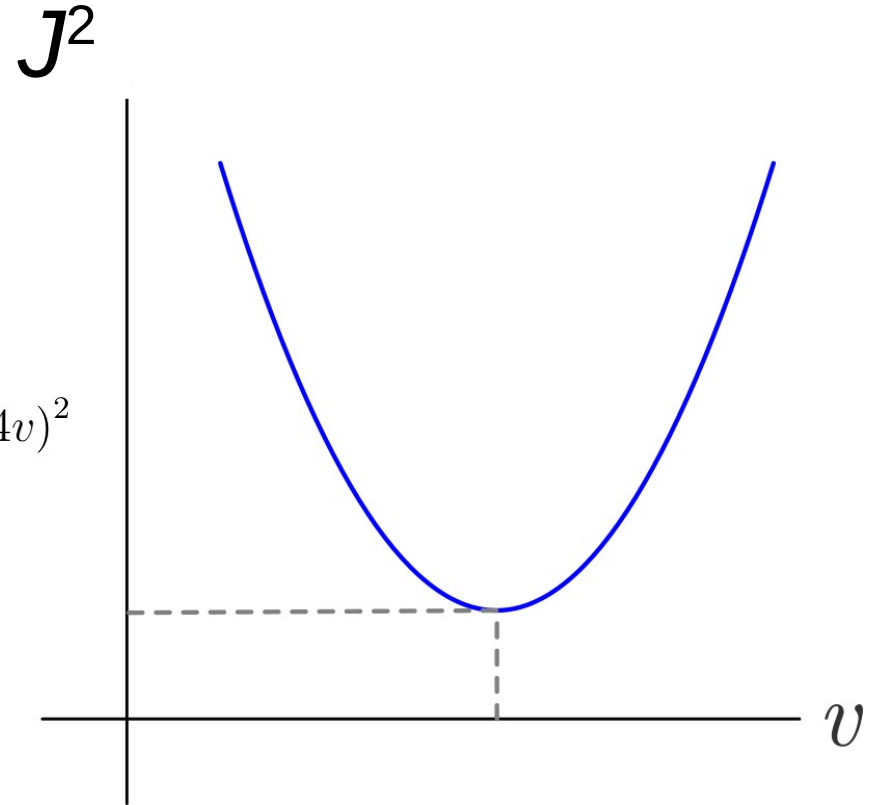




# Regresi Linier

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

$$J^2 = (4.7 - v)^2 + (9.2 - 2v)^2 + (13.7 - 3v)^2 + (18.1 - 4v)^2 \\ + (22.7 - 5v)^2 + (27.2 - 6v)^2,$$



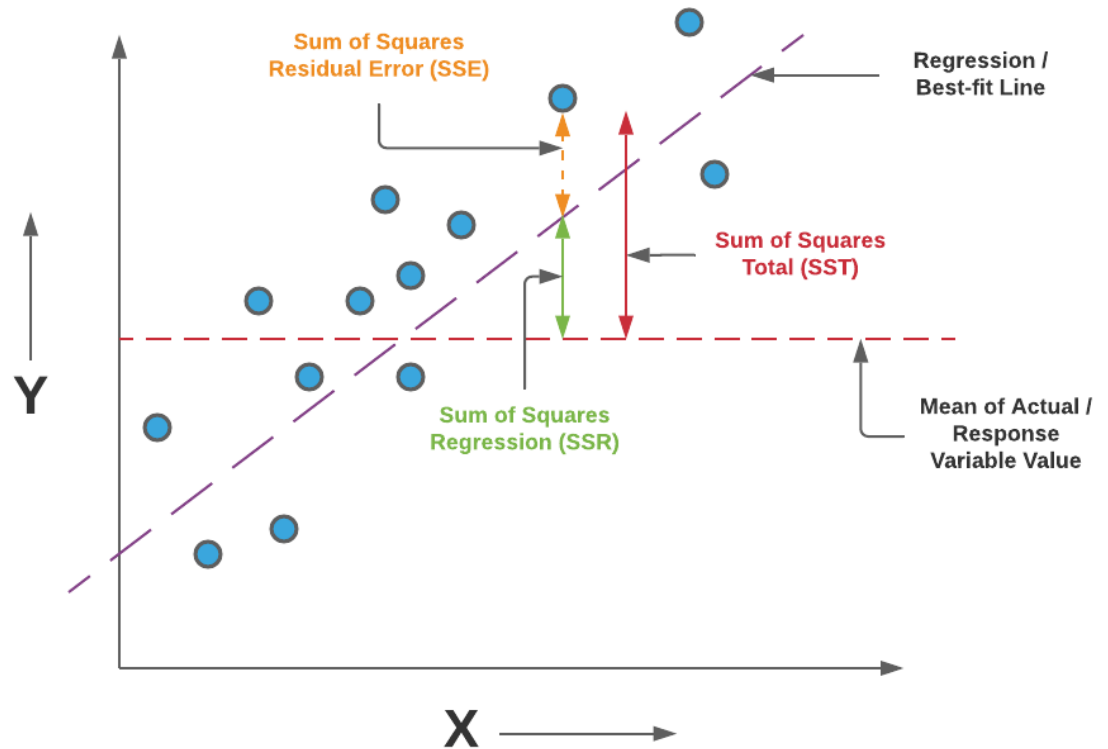
# Dataset untuk Model ML

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

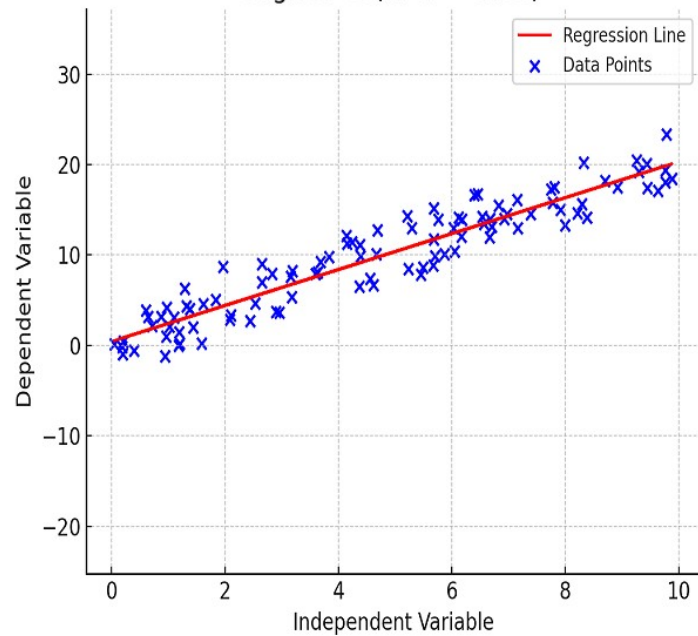
# Performa Model

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

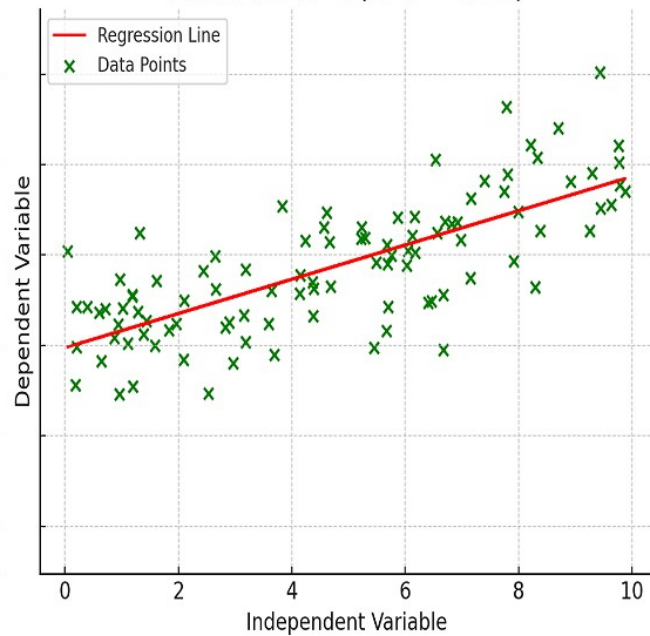
$$\text{RMSE}(\vec{y}, \hat{y}) = \sqrt{\frac{1}{m} \sum_{i=1}^m (\hat{y}_i - y_i)^2}$$



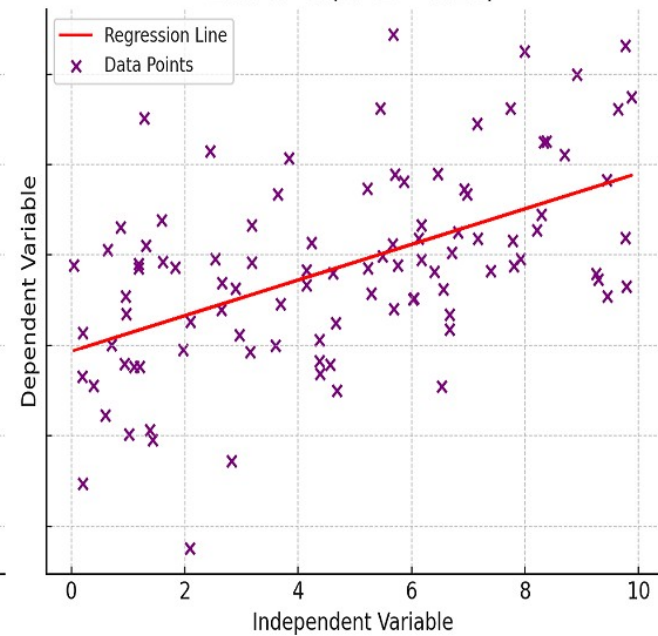
High  $R^2$  ( $R^2 = 0.89$ )



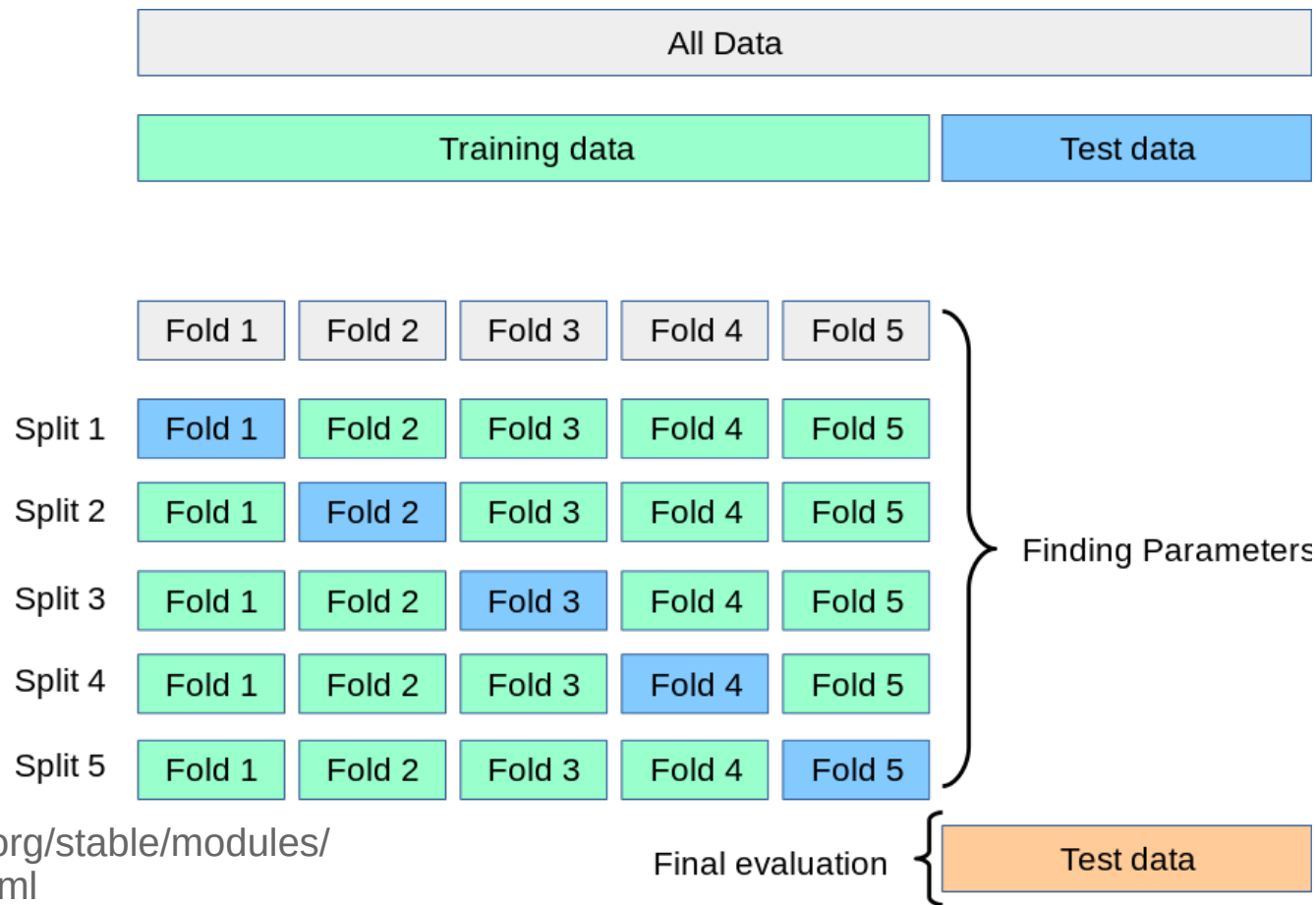
Moderate  $R^2$  ( $R^2 = 0.55$ )



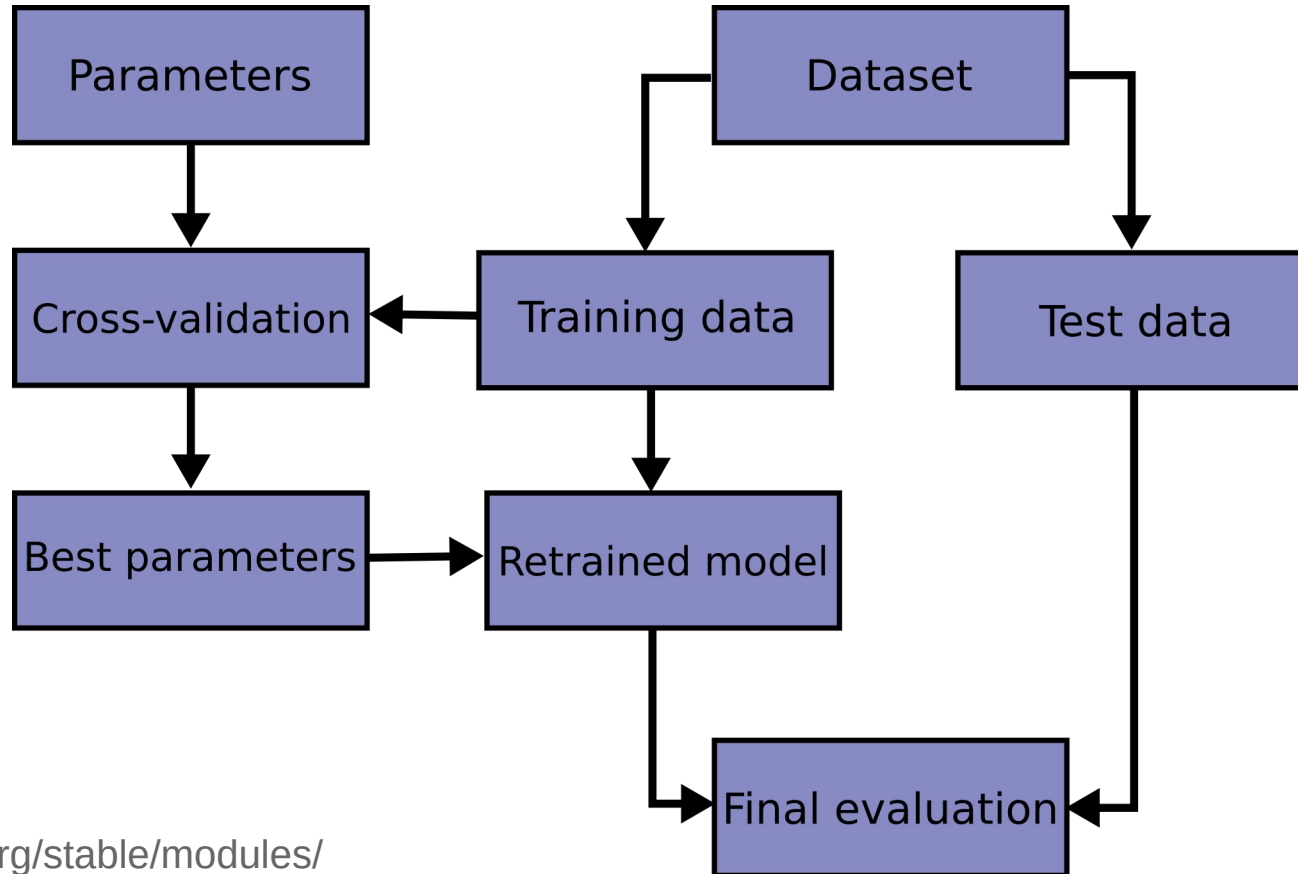
Low  $R^2$  ( $R^2 = 0.28$ )



# K-Fold Cross-Validation

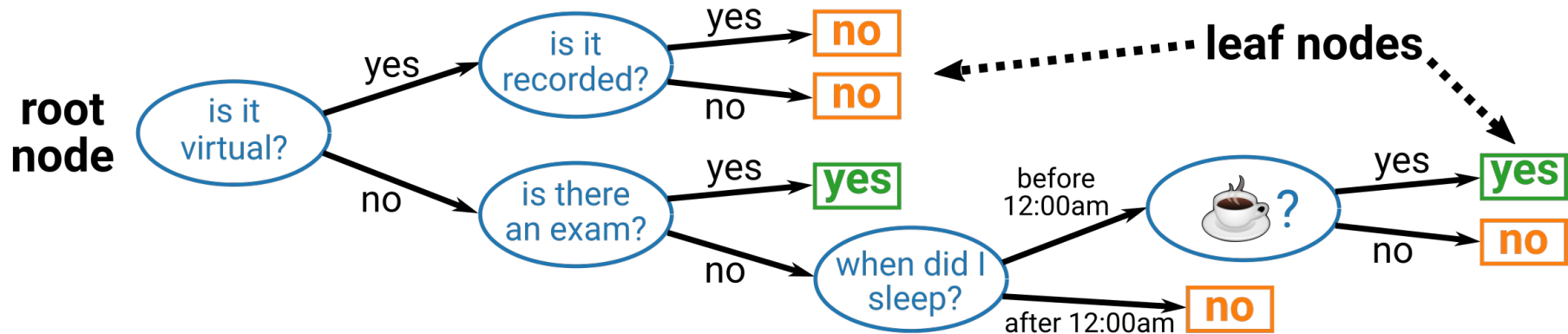


# K-Fold Cross-Validation

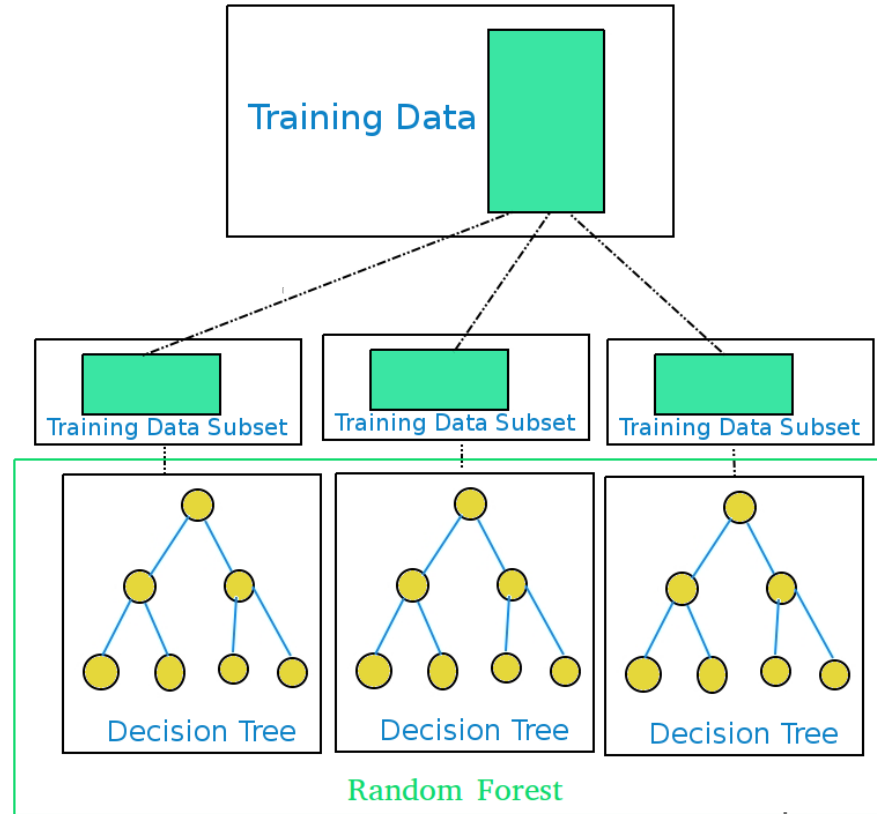


# Decision Tree

*Should I go to my 8:00am class?*



# Random Forest





# Perfoma Classifier

	Positive prediction	Negative prediction
Positive label:	True positive (TP)	False negative (FN)
Negative label:	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-glyW8iR2H9Ur>
  - [https://enze-chen.github.io/mi-book-2021/week\\_1/04/intro\\_ml\\_blank.html](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/cgnn?tab=readme-ov-file>
  - <https://github.com/sedaoturak/data-resources-for-materials-science>