



# Machine Learning

Spirit of learning.



# Intro to Machine Learning

By : Nurul Amelia

# Models



## Decision Tree

A decision tree is a support tool with a tree-like structure that models possible outcomes, resource costs, utilities, and possible consequences.

Decision tree provides a way to present algorithms with conditional control statements. They include branches that represent decision-making measures that can lead to favorable outcomes.

# Models



## Random Forest

Random forest is a combination of each good tree then combined into one model. Random Forest relies on a random vector value with the same distribution on all trees, each of which has maximum depth.



# Library Python

## **Pandas**

Pandas are commonly used to create tables, change data dimensions, check data, and so on.

## **Scikit-learn (Sklearn)**

Scikit-learn is easily the most popular library for modeling the types of data typically stored in DataFrames.



# Building Your Model

The steps to building and using a model are:

## **Define**

What type of model will it be? A decision tree? Some other type of model? Some other parameters of the model type are specified too.

## **Fit**

Capture patterns from provided data. This is the heart of modeling.

## **Predict**

Just what it sounds like.

## **Evaluate**

Determine how accurate the model's predictions are.



# Train and Testing Data

In conducting trains and data tests we use the nature of random sampling where the data will be randomized (must not take data from the top or bottom).

Proposed ones used usually :

1. Training Set (70% - 85%) - Used to build models.
2. Testing Set (the rest of the training set) - Used to test model performance.

**Example :**

If you want the Training Set is 80%, then for Test Testing is 20%.

**Noted :**

The results of training and testing should be almost the same.



# Overfitting and Underfitting

## **Overfitting**

Capturing spurious patterns that won't recur in the future, leading to less accurate predictions

## **Underfitting**

Failing to capture relevant patterns, again leading to less accurate predictions.





# Referensi

Accurate. 2022. Decision Tree: Pengertian, Cara Buat, Kelebihan dan Kekurangannya. <https://accurate.id/marketing-manajemen/decision-tree/> , accessed February 25, 2022.

Kaggle. 2022. Course : Intro to Machine Learning. [Learn Intro to Machine Learning Tutorials \(kaggle.com\)](https://www.kaggle.com/learn/intro-to-machine-learning) , accessed February 25, 2022.

Universitas Gajah Mada. 2018. Random Forest. <https://machinelearning.mipa.ugm.ac.id/2018/07/28/random-forest/> , accessed February 25, 2022.



# Thanks

Add your text