

### 4.3. Random Forest

The random forest method is one of the decision tree methods. A decision tree is a tree-like flowchart that has a root node used to collect data. A decision tree classifies a data sample that has no known class. A random forest is a combination of each good tree then combined into one model. The random forest also relies on a random vector value with the same distribution in all trees each decision tree has a maximum depth.

- Add Dataset

The first step is the same as AdaBoost by importing a CSV file using the widget in orange data mining. And data obtained from kaggle.com.

- Data Splitting

After adding a dataset in orange data mining that can divide data into training data and testing data. For example, 6000 data are using fixed 70% making the training data into 4200 training data and 1800 testing data.

- Inputting model/Algoritma

In this step use tools in the orange data mining application and the name of the tool used is a random forest by configuring the number of trees, training replications, and also balancing the class distribution. This step aims to ensure good accuracy and can produce good results.

- Prediction

Then from the use of orange data mining is to make predictions and the prediction output is MSE, RMSE, MAE, R2.

$$MSE = \frac{1}{n} \sum_{i=0}^n (target - prediction)^2 \quad (1)$$

$$RMSE = \sqrt{MSE} \quad (2)$$

$$MAE = \frac{1}{n} \sum_{i=0}^n |target - prediction| \quad (3)$$