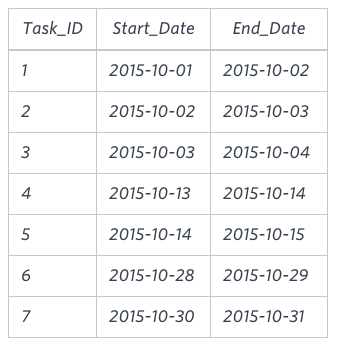
**1) Self Join:**

1. A self join is a regular join, but the tables joined with itself.
2. We make 2 different aliases for the same table.
3. It is like we make a copy of the same table and join them after putting the conditions to get the desired results.

**Example:**



Here,

Let’s perform a Self Join where we can extract Start\_Date who is also an End\_Date.

**Query:**

select A.Start\_Date from Project as A, Project as B where A.Start\_Date=B.End\_Date;

**Output:**

2015-10-02

2015-10-03

2015-10-14

**2) Decision Tree:**

* Decision Tree is a Supervised Machine Learning Algorithm.
* It can be used for solving regression as well as classification problems.



* It is important to select the right feature as a root node for splitting the tree and for this we can use **Entropy technique**.

**Entropy technique:**

* Entropy measures “Purity of Splitting”.
* We should aim to get the leaf node quickly and with good purity.



* Entropy=

Where,

P(+) : Percentage of +ve class

P(-) : Percentage of -ve class.

If we have to calculate the entropy of feature 1 (3yes/2No),

= 0.78 bits

* Entropy ranges from 0 to 1. 1 will be the worst split however 0 will be the best.

**Information Gain :**

* In the Decision Tree we have to select the features and **Information gain** helps to get the best pattern to split adjusting the entropy.
* Information gain causes reduction in entropy and it is calculated by comparing entropy before and after transformation.

Constructing a decision tree is all about finding the right attribute which returns the highest information Gain.

**Gini Index (Gini Impurity) :**

**Gini index** or **Gini** impurity measures the degree or probability of a particular variable being wrongly classified when it is randomly chosen.

Gini Impurity =

I.e =

Let’s consider the split with 1 Entropy and it can be (3yes/3No)

Gini impurity = [ =0.5

* When the probability of a positive class increases then your entropy also increases.
* If the entropy is 1 the Gini index will be 0.5
* The algorithm considers the gini index because it is computationally easy.