**RANDOM FOREST**

Random Forest can be used for both classification and regression task.Random Forest is a model made up of many decision trees rather than simply averaging the prediction of trees.

The model uses two key concepts:-

a) Random Sampling of training data points when building trees.

b) Random Subsets of features considered when splitting nodes.

It can also be used for Feature Engineering.

Random forest pseudocode:

1. Randomly select “k” features from total “m” features.

Where, k << m

1. Among the “k” features,calculate the node “d” using best split point.
2. Split the node into daughter nodes using best split.
3. Repeat 1 to 3 steps until “1” number of nodes has been reached.
4. Build forest by repeating steps 1 to 4 for “n” number times to create “n” number of trees.

**Data Used :**

1. Company dataset - for knowing the attribute that causes high sale using random forest.
2. Fraud dataset - for treating those who have taxable\_income <= 30000 as "Risky" and others are "Good" using random forest.

**Programming:-** Python

**The Codes regarding Random Forest with Two different business problems company dataset ,fraud dataset are present in this Repository in detail.**