# Decision Tree and random forest algorithm in Fraud-Check dataset:

* Business Problem: To make a model to classify a person having taxable income less than 30000 as fraud.
* Data Collection: Fraud\_check.csv
* EDA:

1. We can see that Taxable.income is integer variable, we will convert it into two categories. Taxable.income> 30000 as good and <30000 as risky.
2. There are no missing values.
3. We can find there is imbalance in the dataset, so we will first make a model using imbalanced dataset and then use sampling technique to make it balanced and then make another model using balanced dataset.

* Data Mining:

1. We will split the dataset into training and test data using imbalanced dataset and look at the prediction on test dataset. After going through the code we can see that the model is predicting everyone as good. But our are of interest is looking at the fraud type. So we will apply sampling technique and make another model.
2. We applied sampling technique and find that our model is predicting better than the previous model.
3. Now we will make another model on the sampled dataset using random forest algorithm and look at the accuracy of the model.
4. We can see that random forest model is producing better prediction that the decision tree model.