{'ccp\_alpha': 0.0,

'class\_weight': None,

'criterion': 'gini',

'max\_depth': None,

'max\_features': None,

'max\_leaf\_nodes': None,

'min\_impurity\_decrease': 0.0,

'min\_samples\_leaf': 1,

'min\_samples\_split': 2,

'min\_weight\_fraction\_leaf': 0.0,

'random\_state': None,

'splitter': 'best'}

Criterion has two possible cases: 1) Entropy 2) Gini

Max\_depth : 3,4,5,6,7

'min\_samples\_split' : minimum how many splitters a variables should have : 2,3,4

Random state: 42,1234

Dictionary: (Assumption)

{

'criterion':[ 'gini',’entropy’],

'max\_depth': [3,4,5,6,7,8,9,10],

'random\_state’: [42,1234,45,23]

}

Criterion has 2 parameters

Max depth has 5 parameters

Random state has 2 parameter

Model can develop by 2\*5\*2= 20

20 decision trees will develop

Chance1: {'criterion':[ 'gini',],'max\_depth': [3],'random\_state’: [42]} ========= acc pr re

Chance2: {'criterion':[ 'entropy',],'max\_depth': [3],'random\_state’: [42]} ====== acc pr rec

Chance3: {'criterion':[ 'gini',],'max\_depth': [4],'random\_state’: [42]}

Chance4: {'criterion':[ 'entropy',],'max\_depth': [4],'random\_state’: [42]}

Chance5: {'criterion':[ 'entropy',],'max\_depth': [4],'random\_state’: [1234]}

Model will give the best parameters

20 possible chances are there

There are two possible ways to develop a model

We can choose step by step each combination : GridSearchCV

Or

We can randomly pick any combination : RandomSerachCV

P1+p2+p3 : 750 consider as train data p4:250 consider as test data

P1+p2+p4 : Train data p3: Test data

P1+p3+p4: Train data p2: Test data

P2+p3+p4: Train data p1: Test data

Advnatage : Every tuple or observation consider as training also consider as testing

Total how many times model will develop

Case1: P1+P2+P3 vs P4

Hyper parameters 2\*5\*2 =20

=============================== > BEST

=============================== > Accuracy

Case2: P1+P3+P4 vs P2

Hyper parameters 2\*5\*2 =20

=============================== > BEST

=============================== > Accuracy

Case3: P1+P2+P4 vs P3

Hyper parameters 2\*5\*2 =20

=============================== > BEST

=============================== > Accuracy

Case3: P2+P3+P4 vs P1

Hyper parameters 2\*5\*2 =20

=============================== > BEST

=============================== > Accuracy

So totally 80 DT model

Best one it will pick

Accuracy = (c1+c2+c3+c4)/4

Step-1: we created a base model

With default parameters

Step-2: we created a parameters dictionary file

Step-3: we apply GridSearchCV

We use base model and parameter dictionary file

Inside the GrdSerachCV

Step-4: We need to pass the data (X,y) on GridSearchCV

Step-5: Then after so many combinations check

Best estimator [ This is our best Model]

Best score

Best parameters

Step-6 (Optinonal) cross\_val\_score

We are cross checking, if we pass the best estimator model for CV=5

The mean accuracy = Best score

====== Case-1:

We develop the model by using default parameters

* Accuracy
* Precision
* Recall
* F1cosre
* CMT
* Roc auc curve

======== Case-2:

We find the best parameters by using hyper parameter tuning

We got the best parameters also

========== Case-3:

Again we need develop the model by using best parameters

* Accuracy
* Precision
* Recall
* F1cosre
* CMT
* Roc auc curve