```
Imbalanced dataset Cluster approach

1. cluster the data until we have 80%+ data points in a cluster
target wont be present for clustering
2. cluster wise confidence for the classification
3. training for the data points on the respective clusters



init N = 2
while len(N) >= len(target)/3:
    cluster into N:
    for n in N:
        if n has 80%+ target points:
            Exit external loop

    N++
    fit Cluster(n_clus=N)
    Compute percentage of target present

    continue


let say
N=10
cluster(1,2,3) > 80% target
1 -> 83%
2->80%
3->81%


new data in either of cluster
x in cluster 2
probability will be 80%

if true or false:
    update the score



Cluster with Model approach
---------------------------

target to total ratio = 0.45%
cluster with 20% to 25% target -> 50X better
percent point can be replaced with 20% in that case

for n in N clusters:
    model(n).fit(xn, yn)

for prediction-> predict cluster number(n)
use model(n).predict
```