Training

Model 1 : Logistic Regression

Full Data :

F1-Score 0.8121212121212122

average precision : 0.756154786324492

Both undersample and oversample give lower accuracy

UnderSample:

Using Factor: 4

F1-Score 0.7790697674418605

average precision : 0.6969089908886618

OverSample:

Using Factor 4:

F1-Score 0.8143712574850299

average precision : 0.7131188143481443

Cost-Sensitive:

F1-Score 0.8342857142857144

average precision : 0.7397958111094391

Model 2: RandomForestClassifier

Full Data :

Max\_depth = 10,n\_estimators = 20

F1-Score 0.8953488372093023

average precision : 0.8641799398695132

our best model

OverSample :

Using Factor 4:

Max\_depth = 10,n\_estimators = 20

F1-Score 0.8588235294117648

average precision : 0.85168614173277

Model 3:Neural Network

For layer (64,32,16,8), activation = relu,solver = adam

F1-Score 0.8092485549132947

average precision : 0.801244051135713

Model 3:Voting Classifier

Getting the best three models from random forest,logistic regression,neural network

F1-Score 0.8606060606060606

average precision : 0.8530045188712575

Testing

Using the best model , random forest : max depth = 10 , n\_estimators = 45

Confusion Matrix:

[[56842 21]

[ 13 84]]

Classification Report:

precision recall f1-score support

0 1.00 1.00 1.00 56863

1 0.80 0.87 0.83 97

accuracy 1.00 56960

macro avg 0.90 0.93 0.92 56960

weighted avg 1.00 1.00 1.00 56960

F1-Score: 0.8317

average\_precision: 0.8245941425472436