SMOTE Tek

```
Best performing model based on AUC: XGBoost with AUC: 0.8210

Model Performance Comparison (Cross-validation):
                     Accuracy  Precision    Recall  F1-Score
Naive Bayes          0.601746   0.619116  0.528887  0.570442
Random Forest        0.986630   0.987839  0.985391  0.986613
Gradient Boosting    0.825390   0.814093  0.843437  0.828492
XGBoost              0.960388   0.963003  0.957563  0.960275
Logistic Regression  0.624353   0.623489  0.627907  0.625687

Model Performance Comparison (Test Set):
                     Accuracy  Precision    Recall  F1-Score
Naive Bayes          0.669643   0.033841  0.534335  0.063650
Random Forest        0.969381   0.278586  0.287554  0.282999
Gradient Boosting    0.802128   0.056536  0.536481  0.102291
XGBoost              0.950262   0.185587  0.403433  0.254226
Logistic Regression  0.619904   0.033509  0.613734  0.063548
```

SMOTEENN with RF class weight, subsample/loss for GBT, scale_pos_weight XGBoost

```
Best performing model based on AUC: XGBoost with AUC: 0.8251

Model Performance Comparison (Cross-validation):
                     Accuracy  Precision    Recall  F1-Score
Naive Bayes          0.611527   0.633789  0.559129  0.594118
Random Forest        0.995733   0.998240  0.993360  0.995794
Gradient Boosting    0.835286   0.819728  0.866698  0.842554
XGBoost              0.875593   0.804246  0.998444  0.890872
Logistic Regression  0.637898   0.638898  0.662171  0.650325

Model Performance Comparison (Test Set):
                     Accuracy  Precision    Recall  F1-Score
Naive Bayes          0.649170   0.033903  0.570815  0.064004
Random Forest        0.956935   0.222474  0.420601  0.291017
Gradient Boosting    0.770022   0.052357  0.581545  0.096065
XGBoost              0.718525   0.056919  0.796137  0.106243
Logistic Regression  0.595283   0.032421  0.633047  0.061683
```

Random Undersampling X_train_rus, y_train_rus, Default param

```
Best performing model based on AUC: Random Forest with AUC: 0.8309

Model Performance Comparison (Cross-validation):
                     Accuracy  Precision    Recall  F1-Score
Naive Bayes          0.596969   0.626322  0.482670  0.544684
Random Forest        0.741065   0.732148  0.760580  0.745963
Gradient Boosting    0.688524   0.679025  0.716698  0.697227
XGBoost              0.744044   0.734459  0.764915  0.749260
Logistic Regression  0.631918   0.630988  0.636528  0.633698

Model Performance Comparison (Test Set):
                     Accuracy  Precision    Recall  F1-Score
Naive Bayes          0.693994   0.035021  0.510730  0.065547
Random Forest        0.743912   0.060974  0.776824  0.113072
Gradient Boosting    0.673701   0.044905  0.716738  0.084514
XGBoost              0.734352   0.056709  0.744635  0.105391
Logistic Regression  0.617379   0.033833  0.624464  0.064189
```

Random Oversampling X_train_ros, y_train_ros, Default param

```
Best performing model based on AUC: XGBoost with AUC: 0.8488

Model Performance Comparison (Cross-validation):
                     Accuracy  Precision    Recall  F1-Score
Naive Bayes          0.598623   0.620295  0.508543  0.558883
Random Forest        0.991641   0.983558  1.000000  0.991711
Gradient Boosting    0.730065   0.717292  0.759487  0.737779
XGBoost              0.933315   0.906889  0.965794  0.935410
Logistic Regression  0.624724   0.623660  0.629041  0.626335

Model Performance Comparison (Test Set):
                     Accuracy  Precision    Recall  F1-Score
Naive Bayes          0.685561   0.034749  0.521459  0.065156
Random Forest        0.969652   0.269488  0.259657  0.264481
Gradient Boosting    0.703012   0.047872  0.695279  0.089577
XGBoost              0.892767   0.119730  0.645923  0.202013
Logistic Regression  0.620220   0.033755  0.618026  0.064014
```

Default X_train scaled and Y_train, No param (baseline)

```
Best performing model based on AUC: XGBoost with AUC: 0.8517

Model Performance Comparison (Cross-validation):
                     Accuracy  Precision    Recall  F1-Score
Naive Bayes          0.959054   0.057421  0.062837  0.059981
Random Forest        0.976055   0.351760  0.179298  0.237503
Gradient Boosting    0.979099   0.175000  0.003248  0.006375
XGBoost              0.979583   0.574204  0.073131  0.129556
Logistic Regression  0.979189   0.000000  0.000000  0.000000

Model Performance Comparison (Test Set):
                     Accuracy  Precision    Recall  F1-Score
Naive Bayes          0.959686   0.054167  0.055794  0.054968
Random Forest        0.974657   0.312500  0.171674  0.221607
Gradient Boosting    0.978941   0.000000  0.000000  0.000000
XGBoost              0.979843   0.633803  0.096567  0.167598
Logistic Regression  0.978986   0.000000  0.000000  0.000000
```

SMOTE without scale_pos_weight XGBoost, default all others

```
Best performing model based on AUC: XGBoost with AUC: 0.8231

Model Performance Comparison (Cross-validation):
                     Accuracy  Precision    Recall  F1-Score
Naive Bayes          0.601294   0.617877  0.530948  0.571116
Random Forest        0.982511   0.983861  0.981118  0.982487
Gradient Boosting    0.827853   0.817219  0.844639  0.830699
XGBoost              0.958103   0.960953  0.955017  0.957974
Logistic Regression  0.623929   0.622998  0.627729  0.625353

Model Performance Comparison (Test Set):
                     Accuracy  Precision    Recall  F1-Score
Naive Bayes          0.667974   0.033671  0.534335  0.063351
Random Forest        0.970328   0.285714  0.274678  0.280088
Gradient Boosting    0.804293   0.056344  0.527897  0.101821
XGBoost              0.951299   0.184805  0.386266  0.250000
Logistic Regression  0.620446   0.033556  0.613734  0.063633
```

Random Forest class_weight='balanced' + XGBoost scale_pos_weight + GBT subsample = 0.8 and loss = 'log_loss'

```
Best performing model based on AUC: XGBoost with AUC: 0.8492

Model Performance Comparison (Cross-validation):
                     Accuracy  Precision    Recall  F1-Score
Naive Bayes          0.959054   0.057421  0.062837  0.059981
Random Forest        0.975029   0.323579  0.184171  0.234627
Gradient Boosting    0.979099   0.298352  0.006497  0.012677
XGBoost              0.903294   0.127470  0.624064  0.211694
Logistic Regression  0.979189   0.000000  0.000000  0.000000

Model Performance Comparison (Test Set):
                     Accuracy  Precision    Recall  F1-Score
Naive Bayes          0.959686   0.054167  0.055794  0.054968
Random Forest        0.973575   0.295918  0.186695  0.228947
Gradient Boosting    0.978986   0.500000  0.006438  0.012712
XGBoost              0.897231   0.125258  0.650215  0.210052
Logistic Regression  0.978986   0.000000  0.000000  0.000000
```

SMOTE with scale_pos_weight XGBoost, subsample = 0.8 and loss = 'log_loss' and class sample Gradient Boosted Trees

```
Best performing model based on AUC: XGBoost with AUC: 0.8257

Model Performance Comparison (Cross-validation):
                     Accuracy  Precision    Recall  F1-Score
Naive Bayes          0.601294   0.617877  0.530948  0.571116
Random Forest        0.982511   0.983861  0.981118  0.982487
Gradient Boosting    0.824031   0.811749  0.843753  0.827434
XGBoost              0.845509   0.764773  0.998031  0.865963
Logistic Regression  0.623929   0.622998  0.627729  0.625353

Model Performance Comparison (Test Set):
                     Accuracy  Precision    Recall  F1-Score
Naive Bayes          0.667974   0.033671  0.534335  0.063351
Random Forest        0.970328   0.285714  0.274678  0.280088
Gradient Boosting    0.805916   0.057222  0.532189  0.103333
XGBoost              0.704591   0.053558  0.783262  0.100261
Logistic Regression  0.620446   0.033556  0.613734  0.063633
```