# **Table of Contents**

- 1. ### Baseline
  - 1.1 Read the file
  - 1.2 Explore Data
  - 1.2 Visualize Data
  - 1.4 Engineer Features
  - 1.5 Handle Outliers
  - 1.6 Heatmap Visualization
  - 1.7 Feature Selection Initial Exploration
  - 1.8 Basic Model Comparison
  - 1.9 Does clipping CreditScore and Age give better results?
  - 1.10 Model Comparison With Smote

#### Hyper Parameter Optimization

- 1.11 Grid Search Gradient Boost
- 1.12 Grid Search Random Forest
- 1.13 Grid Search Support Vector Classifier
- 1.14 Grid Search Multi-Layer Perceptrons
- 1.15 Final Baseline
- 2. ### Basic Experimentation
  - 2.1 Are there any outliers detected by Random Forest?
  - 2.2 Feature Selection, PCA, SVC
  - 2.3 Feature Selection, Random Fores
  - 2.4 Compare Feature Selection Methods
  - 2.5 Check if MinMaxScaling makes a difference
  - 2.6 Add PCA to the mix
- 3. ### Research Section
  - 3.1 Experimentation with Weighted-Features in Random Forest
  - 3.2 Grid Search different weights (no SMOTE)
  - 3.3 GridSearch BalancedRandomForestClassifier from Imblearn
  - 3.4 Random Forest with Adjusted Cutoff
  - 3.5 Classification with Other Samplers (over-samplers and combine-samplers)
  - 3.6 Classifiers Samplers and Feature-Selectors compared
  - 3.7 RUSBoost (Boost + UnderSampling at each stage)

In [ ]:	
In [235]:	%matplotlib inline

```
In [86]: import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
         import seaborn as sns
         from sklearn.preprocessing import OneHotEncoder, StandardScaler
         from sklearn import metrics, model_selection
         from statistics import mean, median
         from time import perf counter
         from sklearn.pipeline import Pipeline
         from imblearn.over_sampling import SMOTE
         from sklearn.model selection import GridSearchCV, StratifiedKFold, cross val p
         redict
         from sklearn.metrics import make_scorer
         from sklearn.ensemble import IsolationForest
         from sklearn.feature selection import SelectPercentile, f classif
         from sklearn.neighbors import KNeighborsClassifier
         from sklearn.tree import DecisionTreeClassifier
         from sklearn.naive bayes import GaussianNB
         from sklearn.svm import SVC
         from sklearn.ensemble import RandomForestClassifier, AdaBoostClassifier, Gradi
         entBoostingClassifier, BaggingClassifier
         from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
         from sklearn.linear model import LogisticRegression, PassiveAggressiveClassifi
         er, Perceptron, RidgeClassifier, SGDClassifier
         from sklearn.neural network import MLPClassifier
         from sklearn.discriminant analysis import QuadraticDiscriminantAnalysis
         from sklearn.gaussian process import GaussianProcessClassifier
         from sklearn.neighbors import NearestCentroid
         import imblearn
         from imblearn.over sampling import SMOTE, ADASYN, BorderlineSMOTE, SMOTENC, KM
         eansSMOTE, SVMSMOTE
         from imblearn.combine import SMOTEENN, SMOTETomek
         from sklearn.metrics import make scorer
         from sklearn.ensemble import RandomForestClassifier
         import random
         from imblearn.ensemble import BalancedBaggingClassifier, BalancedRandomForestC
         from sklearn.preprocessing import MinMaxScaler
         from sklearn.decomposition import PCA
         from sklearn.tree import DecisionTreeClassifier
         from sklearn.svm import SVC
         from sklearn import model selection
         from sklearn import linear model
         from sklearn.feature selection import RFECV
         from sklearn.feature selection import SelectFromModel
         from imblearn.under sampling import RandomUnderSampler
         from imblearn.over_sampling import RandomOverSampler
         from sklearn.svm import LinearSVC
         from sklearn.feature selection import SelectFwe, chi2, f classif
         from sklearn.feature selection import VarianceThreshold
         from sklearn.ensemble import ExtraTreesClassifier
```

#### Read the file

```
In [216]: def read_file()->pd.DataFrame:
    df = pd.read_csv('Churn_Modelling.csv')
    df.describe()
    return df
df = read_file()
```

### **Baseline**

## **Explore the data**

13

Exited

memory usage: 1.1+ MB

```
In [217]:
           df.head()
Out[217]:
               RowNumber
                          Customerld Surname CreditScore Geography
                                                                      Gender Age
                                                                                  Tenure
                                                                                           Balance
            0
                        1
                            15634602
                                      Hargrave
                                                      619
                                                                      Female
                                                                               42
                                                                                       2
                                                                                              0.00
                                                              France
            1
                        2
                             15647311
                                           Hill
                                                      608
                                                               Spain
                                                                      Female
                                                                               41
                                                                                           83807.86
            2
                        3
                            15619304
                                          Onio
                                                      502
                                                              France
                                                                      Female
                                                                               42
                                                                                          159660.80
                        4
                                                      699
            3
                            15701354
                                          Boni
                                                              France
                                                                      Female
                                                                               39
                                                                                              0.00
                        5
                            15737888
                                       Mitchell
                                                      850
                                                               Spain
                                                                      Female
                                                                               43
                                                                                       2 125510.82
In [218]:
           df.info()
           <class 'pandas.core.frame.DataFrame'>
           RangeIndex: 10000 entries, 0 to 9999
           Data columns (total 14 columns):
            #
                 Column
                                   Non-Null Count
                                                     Dtype
                 ____
                                   _____
                                                     _ _ _ _ _
            0
                 RowNumber
                                   10000 non-null
                                                     int64
            1
                CustomerId
                                   10000 non-null
                                                     int64
            2
                 Surname
                                   10000 non-null
                                                     object
            3
                CreditScore
                                   10000 non-null
                                                    int64
            4
                 Geography
                                   10000 non-null
                                                     object
            5
                 Gender
                                   10000 non-null
                                                     object
            6
                                                     int64
                 Age
                                   10000 non-null
            7
                Tenure
                                   10000 non-null
                                                     int64
            8
                 Balance
                                   10000 non-null
                                                    float64
            9
                 NumOfProducts
                                   10000 non-null
                                                     int64
            10
                                                     int64
                HasCrCard
                                   10000 non-null
                                                     int64
            11
                IsActiveMember
                                   10000 non-null
            12
                EstimatedSalary
                                   10000 non-null
                                                     float64
```

10000 non-null

dtypes: float64(2), int64(9), object(3)

int64

# In [219]: df.describe().T

# Out[219]:

	count	mean	std	min	25%	50%	
RowNumber	10000.0	5.000500e+03	2886.895680	1.00	2500.75	5.000500e+03	٠
CustomerId	10000.0	1.569094e+07	71936.186123	15565701.00	15628528.25	1.569074e+07	
CreditScore	10000.0	6.505288e+02	96.653299	350.00	584.00	6.520000e+02	
Age	10000.0	3.892180e+01	10.487806	18.00	32.00	3.700000e+01	
Tenure	10000.0	5.012800e+00	2.892174	0.00	3.00	5.000000e+00	
Balance	10000.0	7.648589e+04	62397.405202	0.00	0.00	9.719854e+04	
NumOfProducts	10000.0	1.530200e+00	0.581654	1.00	1.00	1.000000e+00	
HasCrCard	10000.0	7.055000e-01	0.455840	0.00	0.00	1.000000e+00	
IsActiveMember	10000.0	5.151000e-01	0.499797	0.00	0.00	1.000000e+00	
EstimatedSalary	10000.0	1.000902e+05	57510.492818	11.58	51002.11	1.001939e+05	
Exited	10000.0	2.037000e-01	0.402769	0.00	0.00	0.000000e+00	
4						<b>&gt;</b>	

# Check if there are any missing values

In [220]:	df.isnull().sum(a	is=0)	
Out[220]:	RowNumber	0	
	CustomerId	0	
	Surname	0	
	CreditScore	0	
	Geography	0	
	Gender	0	
	Age	0	
	Tenure	0	
	Balance	0	
	NumOfProducts	0	
	HasCrCard	0	
	IsActiveMember	0	
	EstimatedSalary	0	
	Exited	0	
	dtype: int64		

There are no missing values that we need to deal with here

## Visualize the data

In [221]: sns.pairplot(df) plt.show()

# **Feature Engineering**

#### Drop columns we know from experience that are irrelevant, One Hot Encoding, Standardization

We can straight away drop Surname, Customerld, RowNumber without checking anything because we know that these don't have any bearing on the churn rate. Furthermore, including Surname might actually add discrimination based on race etc.

The next step is to one-hot encode Gender and Geography.

The last step is to normalize some of the fields which have numbers.

Scaling is done before outlier handling because once scaling is done, it will allow us to plot the box plots at one go in a single graph. Scaling doesn't affect the frequency distribution and outlier detection.

```
In [222]: def engineer_features(df: pd.DataFrame, donot_scale=False)->tuple:
              ohdf = df.drop(columns=['RowNumber', 'Surname', 'CustomerId'])
              ohdf.describe()
              X, y = ohdf[ohdf.columns[:-1]], ohdf[df.columns[-1]]
              X = pd.get dummies(X, columns=['Gender', 'Geography'], dummy na=False)
              scaled X = X.copy()
              columns_to_transform = ['CreditScore', 'Age', 'Tenure', 'Balance', 'NumOfP
          roducts', 'EstimatedSalary']
              if not donot scale:
                  scaled X[columns to transform] = StandardScaler().fit transform(scaled
          X[columns to transform])
              return X, scaled X, y
          X, scaled_X, y = engineer_features(df)
In [223]: def engineer features2(df: pd.DataFrame)->tuple:
              ohdf = df.drop(columns=['RowNumber', 'Surname', 'CustomerId'])
              X = pd.get dummies(ohdf, columns=['Gender', 'Geography'], dummy_na=False)
              return X
```

#### **Handle Outliers**

If any outliers are present, they must be removed. The way to check for outliers is to

```
ax = sns.boxplot(data=scaled_X[['CreditScore', 'Age', 'Tenure', 'Balance', 'Nu
In [224]:
         mOfProducts', 'EstimatedSalary']])
         ax.set_xticklabels(ax.get_xticklabels(),rotation=45)
Text(2, 0, 'Tenure'),
          Text(3, 0, 'Balance'),
          Text(4, 0, 'NumOfProducts'),
          Text(5, 0, 'EstimatedSalary')]
           3
           2
           1
           0
          ^{-1}
          -2
          -3
                             pae
```

# Clip NumOfProducts to [mean - 1.5 IQR, mean + 1.5 IQR]

NumOfProducts is has outliers which are beyond the whiskers

```
In [225]: | q1 = scaled_X['NumOfProducts'].quantile(0.25)
          q3 = scaled_X['NumOfProducts'].quantile(0.75)
          avg = scaled_X['NumOfProducts'].mean()
          iqr = q3 - q1
          themin, themax = avg - 1.5 * iqr, avg + 1.5 * iqr
          scaled_X[['NumOfProducts']] = scaled_X[['NumOfProducts']].clip(themin, themax)
          ax = sns.boxplot(data=scaled_X[['CreditScore', 'Age', 'Tenure', 'Balance', 'Nu
          mOfProducts', 'EstimatedSalary']])
          ax.set_xticklabels(ax.get_xticklabels(),rotation=45)
Out[225]: [Text(0, 0, 'CreditScore'),
           Text(1, 0, 'Age'),
           Text(2, 0, 'Tenure'),
           Text(3, 0, 'Balance'),
           Text(4, 0, 'NumOfProducts'),
           Text(5, 0, 'EstimatedSalary')]
            3
            2
            1
            0
           ^{-1}
           -2
           -3
```

While there are some outliers here in CreditScore and Age, we will choose to let them remain as all these values form a continuum and touch the whiskers.

Visualize the Data as a heatmap

```
In [226]: sns.heatmap(pd.concat([scaled_X, y], axis=1), robust=True)
Out[226]: <AxesSubplot:>
                                    0
233
466
699
932
11398
1631
12997
23363
22796
3029
3495
3728
3495
3495
3495
3495
3495
55592
56058
66524
66524
66524
66524
67223
7456
6990
72923
7456
88521
8858
8854
99586
                                                                                                                                                                                                                                                                               - 1.5
                                                                                                                                                                                                                                                                               - 1.0
                                                                                                                                                                                                                                                                                - 0.5
                                                                                                                                                                                                                                                                                - 0.0
                                                                                                                                                                                                                                                                                  -0.5
                                                                                                                                                                                                                                                                                - -1.0
                                                                                                              NumOfProducts
                                                                                                                             HasCrCard
                                                                                                                                                          EstimatedSalary
                                                                                                                                                                        Gender_Female
                                                                                                                                                                                       Gender_Male
                                                                                                                                                                                                     Geography_France
                                                                                                                                                                                                                   Geography_Germany
                                                                                                                                                                                                                                 Geography_Spain
                                                                                                                                           SActiveMember
```

As we can see the Exited field is unbalanced Print the number of instances of each class

# **Feature Selection Initial Exploration**

#### Check univariate feature selection first

We know that this will be biased against categorical values, but we'll see this anyway

```
In [228]: def feature selection with select percentile():
            selector = SelectPercentile(f classif, percentile=25)
            selector.fit(scaled X, y)
            print('-' * 120)
            print("%40s ----- %10s" % ('Feature', 'Score'))
            print('-' * 120)
            sorted_indices = np.argsort(selector.scores_)[::-1]
            for i in sorted indices:
                print("%40s ----- %10.5f" % (scaled X.columns[i], selector.scores [i
         ]))
            print('-' * 120)
            return selector
         selector = feature_selection_with_select_percentile()
                                      Feature ----- Score
           -----
                                         Age ----- 886.06327
                             Geography_Germany ----- 310.25838
                                IsActiveMember ----- 249.80079
                                      Balance ----- 142.47383
                                 Gender_Female ----- 114.72799
                                  Gender Male ----- 114.72799
                              Geography France ----- 111.36052
                                 NumOfProducts ----- 49.28413
                               Geography Spain ----- 27.80947
                                  CreditScore ----- 7.34452
                                       Tenure ----- 1.96016
                               EstimatedSalary ----- 1.46326
                                    HasCrCard ----- 0.50940
```

Also check the Tree based feature selector to see which are better

```
def feature_selection_with_random_forest():
    forest = RandomForestClassifier(n estimators=250, random state=0)
    forest.fit(scaled_X, y)
    sorted indices = np.argsort(forest.feature importances )[::-1]
    print('-' * 120)
    print("%40s ----- %10s" % ('Feature', 'Importance'))
    print('-' * 120)
    for i in sorted indices:
       print("%40s ----- %10.5f" % (scaled_X.columns[i], forest.feature_imp
ortances_[i]))
    print('-' * 120)
    return forest
forest = feature_selection_with_random_forest()
                               Feature ----- Importance
                                  Age -----
                                                0.24002
                       EstimatedSalary -----
                                                0.14694
                           CreditScore ----- 0.14569
                               Balance ----- 0.14112
                         NumOfProducts ----- 0.12829
                               Tenure ----- 0.08314
                        IsActiveMember ----- 0.04043
                     Geography_Germany ----- 0.02082
                             HasCrCard ----- 0.01836
                      Geography France ----- 0.00993
                       Geography_Spain ----- 0.00855
                         Gender_Female ----- 0.00854
                           Gender Male ----- 0.00818
```

Also see Greedy feature selection to see what it gives us

```
In [16]:
        def feature_selection_with_svc():
             svm = SVC()
             scores = model_selection.cross_val_score(svm, scaled_X, y, cv=10)
             print('Initial Results', scores.mean())
             estimator = linear_model.LogisticRegression(multi_class='auto', solver='lb
         fgs')
             rfecv = RFECV(estimator, cv=10)
             rfecv.fit(scaled X, y)
             sorted_indices = np.argsort(rfecv.ranking_)
             print(rfecv.n features )
             print('-' * 120)
             print("%40s ----- %10s" % ('Feature', 'Ranking'))
             print('-' * 120)
             for i in sorted indices:
                 print("%40s ----- %10.5f" % (scaled_X.columns[i], rfecv.ranking_[i
         ]))
             print('-' * 120)
             return rfecv
         rfecv = feature selection with svc()
         Initial Results 0.858
         4
                                         Feature -----
                                                           Ranking
                                            Age -----
                                                           1.00000
                                  IsActiveMember -----
                                                           1.00000
                                     Gender_Male -----
                                                           1.00000
                               Geography_Germany -----
                                                           1.00000
                                Geography_France -----
                                                           2.00000
                                 Geography Spain -----
                                                           3.00000
                                   Gender Female -----
                                                          4.00000
                                         Balance -----
                                                           5.00000
                                   NumOfProducts -----
                                                           6.00000
                                     CreditScore -----
                                                          7.00000
                                         Tenure -----
                                                          8.00000
                                       HasCrCard -----
                                                          9.00000
                                 EstimatedSalary ----- 10.00000
```

```
In [17]: def get_normalized(a: list):
             Normalizes a list to [0,1]
             themin = min(a)
             themax = max(a)
             ret = []
             print(a, type(a))
             for i in a:
                 ret.append((i - themin) / (themax - themin))
             return ret
         print('-' * 120)
         print("%40s ----- %10s %10s %10s" % ('Feature', 'Score', 'Importa
         nce', 'Ranking'))
         print('-' * 120)
         for i, j, k, l in zip(scaled_X.columns, selector.scores_, forest.feature_impor
         tances_, rfecv.ranking_):
             print("%40s ----- %10.1f %10.3f %10d" % (i, j, k, l))
         print('-' * 120)
         fig, ax = plt.subplots()
         ax.plot(get_normalized(selector.scores_), label='Univariate')
         ax.plot(get_normalized(forest.feature_importances_), label='RandomForest')
         rf_ranking_inv = [max(rfecv.ranking_) - i for i in rfecv.ranking_]
         ax.plot(get_normalized(rf_ranking_inv), label='Greedy')
         plt.xticks(np.arange(0, len(scaled X.columns), 1.0))
         ax.set xticklabels(scaled X.columns, rotation=270)
         ax.legend()
```

ance	Ranking	Feature	 
0.146	7	CreditScore	 7.3
		Age	 886.1
0.240	1	Tenure	 2.0
0.083	8		
0.141	5	Balance	 142.5
		NumOfProducts	 49.3
0.128	6	HasCrCard	 0.5
0.018	9		
0.040	1	IsActiveMember	 249.8
0.040	<b>-</b>	EstimatedSalary	 1.5
0.147	10	Gender Female	 114.7
0.009	4		114.7
0.008	1	Gender_Male	 114.7
0.008	1	Geography_France	 111.4
0.010	2	Goognaphy Gonmany	 310.3
0.021	1		
0.009	3	Geography_Spain	 27.8

------

Out[17]: <matplotlib.legend.Legend at 0x1ec48994400>

<sup>[7.34452216</sup>e+00 8.86063275e+02 1.96016363e+00 1.42473832e+02

<sup>4.92841330</sup>e+01 5.09401035e-01 2.49800794e+02 1.46326192e+00

<sup>1.14727989</sup>e+02 1.14727989e+02 1.11360517e+02 3.10258384e+02

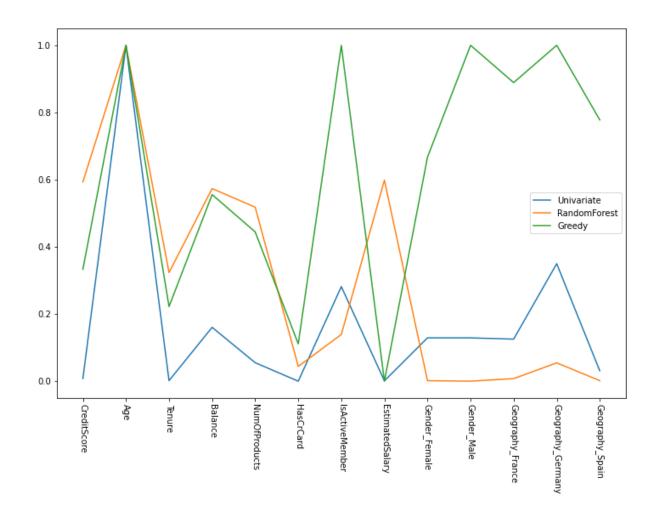
<sup>2.78094682</sup>e+01] <class 'numpy.ndarray'>

<sup>[0.14568742 0.24001543 0.08314233 0.14111972 0.12828553 0.01836008</sup> 

<sup>0.04043115 0.14693745 0.00854344 0.00817559 0.00993053 0.02081914</sup> 

<sup>0.00855218] &</sup>lt;class 'numpy.ndarray'>

<sup>[3, 9, 2, 5, 4, 1, 9, 0, 6, 9, 8, 9, 7] &</sup>lt;class 'list'>



With all the features, using the three different methods to see which features are noisy, we don't get any clear indication about which features are important and which are not.

HasCrCard is probably not important, but even that is not very clear, so we'll let things be.

On the other hand, we know that Age, Balance, and IsActiveMember are probably very important.

## **Compare Basic Models**

```
In [ ]: import warnings; warnings.simplefilter('ignore')
        def compare basic models(run fn, test models=None, no sort=False)->None:
            Run models against the data, compare them according to F1 and Accuracy
            basic models = [
                 ('DecisionTreeClassifier', lambda: DecisionTreeClassifier(random state
        =1)),
                 ('KNeighborsClassifier', lambda: KNeighborsClassifier()),
                 ('NearestCentroid', lambda: NearestCentroid()),
                 ('NaiveBayes', lambda: GaussianNB()),
                #('Gaussian Process', lambda: GaussianProcessClassifier(random state=
        1)),
                 ('Support Vector Machines', lambda: SVC(random_state=1)),
                 ('Random Forest', lambda: RandomForestClassifier(random_state=1)),
                 ('Ada Boost', lambda: AdaBoostClassifier(random_state=1)),
                 ('Gradient Boost', lambda: GradientBoostingClassifier()),
                 ('Linear Discriminant Analysis', lambda: LinearDiscriminantAnalysis
         ()),
                 ('Quadriatic Discriminant Analysis', lambda: QuadraticDiscriminantAnal
        ysis()),
                 ('Logistic Regression', lambda: LogisticRegression(random_state=1)),
                 ('Ridge Classifier', lambda: RidgeClassifier(random_state=1)),
                 ('Bagging Classifier', lambda: BaggingClassifier(random_state=1)),
                 ('SGD Classifier', lambda: SGDClassifier(random state=1)),
                 ('Passive Aggressive Classifier', lambda: PassiveAggressiveClassifier(
        random state=1)),
                 ('Perceptron', lambda: Perceptron(random_state=1)),
                 ('Multi-Layer Perceptron', lambda: MLPClassifier(random_state=1))
            if None != test models:
                basic_models = test_models
            accuracy_scores = []
            f1 scores = []
            models = []
            run_times = []
            for name, fn in basic models:
                print(name, fn)
                t1 = perf_counter()
                acc, f1 = run_fn(scaled_X, y, fn)
                run_times.append(perf_counter() - t1)
                 accuracy scores.append(acc)
                f1 scores.append(f1)
                models.append(name)
            fig, ax = plt.subplots()
            min times = min(run times)
            max times = max(run times)
            if len(run times) > 1:
                for i in range(len(run times)):
                     run_times[i] = (run_times[i] - min_times) / (max_times - min_times
        )
            sorted indices = np.argsort(f1 scores) # Sort by f1 score
            if no sort:
                 sorted indices = [i for i in range(len(f1 scores))]
```

```
ax.plot(range(len(run_times)), [run_times[i] for i in sorted_indices], lab
        el="run_time")
            ax.plot(range(len(accuracy_scores)), [accuracy_scores[i] for i in sorted_i
        ndices], label="accuracy")
            ax.plot(range(len(f1 scores)), [f1 scores[i] for i in sorted indices], la
        bel="f1 score")
            plt.xticks(np.arange(0, len(accuracy scores), 1.0))
            ax.set_xticklabels([models[i] for i in sorted_indices], rotation=270)
            fig.legend()
            print('-' * 120)
            print("%40s ---- %10s %10s %10s" % ('Model', 'Accuracy', 'F1 Score', 'R
        un Time'))
            print('-' * 120)
            for i in sorted_indices:
                print("%40s ----
                                    %10.3f %10.3f %10.3f" % (models[i], accuracy_score
        s[i], f1_scores[i], run_times[i]))
            print('-' * 120)
            return None
In [ ]: def run basic model(X:pd.DataFrame, y:pd.DataFrame, model gen fn):
            kf = model selection.StratifiedKFold(n splits=6, shuffle=True, random stat
        e=5)
            accuracy_scores = []
            f1_scores = []
            for train_index, val_index in kf.split(X, y):
                clf = model_gen_fn()
                clf.fit(X.to numpy()[train index], y.to numpy()[train index])
                y_pred = clf.predict(X.to_numpy()[val_index])
                y_true = y.to_numpy()[val_index]
                accuracy_scores.append(metrics.accuracy_score(y_true, y_pred))
                f1_scores.append(metrics.f1_score(y_true, y_pred))
```

#print(median(accuracy\_scores), median(f1\_scores))
return mean(accuracy scores), mean(f1 scores)

In [230]: compare\_basic\_models(run\_basic\_model)

DecisionTreeClassifier <function compare\_basic\_models.<locals>.<lambda> at 0x
000001D921C8B1F0>

KNeighborsClassifier <function compare\_basic\_models.<locals>.<lambda> at 0x00
0001D921C8B3A0>

NearestCentroid <function compare\_basic\_models.<locals>.<lambda> at 0x000001D 85A6FC9D0>

NaiveBayes <function compare\_basic\_models.<locals>.<lambda> at 0x000001D85C11 3B80>

Support Vector Machines <function compare\_basic\_models.<locals>.<lambda> at 0
x000001D85C0DA940>

Random Forest <function compare\_basic\_models.<locals>.<lambda> at 0x000001D85
C0DAA60>

Ada Boost <function compare\_basic\_models.<locals>.<lambda> at 0x000001D85FCC9 040>

Gradient Boost <function compare\_basic\_models.<locals>.<lambda> at 0x000001D9
21DB81F0>

Linear Discriminant Analysis <function compare\_basic\_models.<locals>.<lambda>
at 0x000001D921DB8040>

Quadriatic Discriminant Analysis <function compare\_basic\_models.<locals>.<lam bda> at 0x000001D921DB80D0>

Logistic Regression <function compare\_basic\_models.<locals>.<lambda> at 0x000
001D921DB8160>

Ridge Classifier <function compare\_basic\_models.<locals>.<lambda> at 0x000001
D921DB83A0>

Bagging Classifier <function compare\_basic\_models.<locals>.<lambda> at 0x0000
01D921DB8430>

SGD Classifier <function compare\_basic\_models.<locals>.<lambda> at 0x000001D9 21DB84C0>

Passive Aggressive Classifier <function compare\_basic\_models.<locals>.<lambda
> at 0x000001D921DB8550>

Perceptron compare\_basic\_models.<locals>.<lambda> at 0x000001D921DB
85F0>

Multi-Layer Perceptron <function compare\_basic\_models.<locals>.<lambda> at 0x
000001D921DB8670>

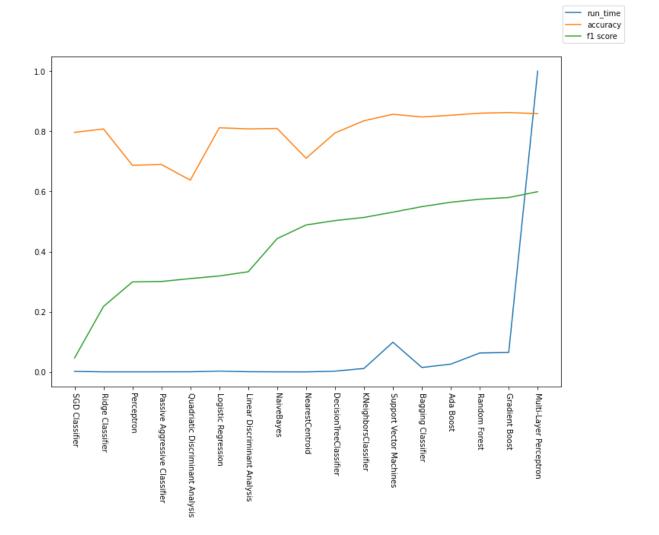
[0.7946989833725933, 0.8352004269014144, 0.7105023773156534, 0.80950130502110 86, 0.8568007887018034, 0.8603999488217603, 0.8534005683857226, 0.86249994898 97939, 0.8081003247129686, 0.637607952599156, 0.8119003450210318, 0.808000224 6849549, 0.8476987075574081, 0.796500483816802, 0.6897167985370513, 0.6870214 792575698, 0.8589005488218083]

[0.5031704409586167, 0.5137007304064225, 0.4886215058787005, 0.44326044155491 995, 0.5311815256147012, 0.5742653346426003, 0.5641624471239775, 0.5798795450 56633, 0.33304487433477115, 0.3102288048496446, 0.3190099885747713, 0.2175284 9103995742, 0.5496146844524826, 0.046178375883008924, 0.3006348728021189, 0.2 992211692260719, 0.5993378896652194]

Time	Model	. <b></b>	Accuracy	F1 Score	Run
	SGD Classifier		0.797	0.046	
0.002	Ridge Classifier		0.808	0.218	

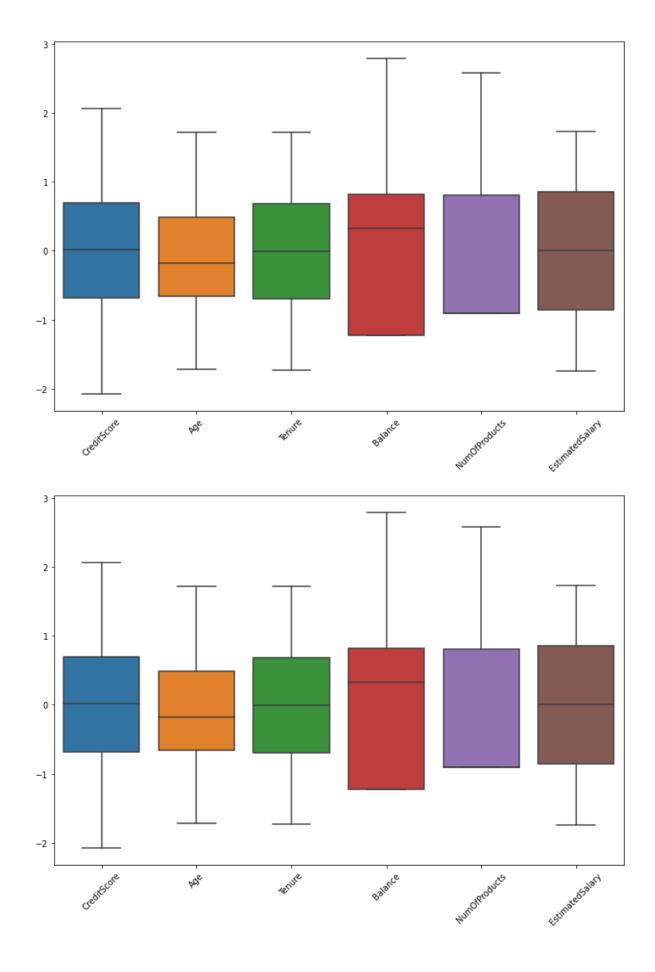
	Perceptron	 0.687	0.299
0.000	Passive Aggressive Classifier	 0.690	0.301
0.000	Quadriatic Discriminant Analysis	 0.638	0.310
0.000	Logistic Regression	 0.812	0.319
0.003	Linear Discriminant Analysis	 0.808	0.333
0.001	NaiveBayes	 0.810	0.443
0.000	NearestCentroid	0.711	0.489
0.000	DecisionTreeClassifier	0.795	0.503
0.003	KNeighborsClassifier	0.835	0.514
0.011	_		
0.099	Support Vector Machines	0.857	0.531
0.015	Bagging Classifier	0.848	0.550
0.026	Ada Boost	0.853	0.564
0.063	Random Forest	 0.860	0.574
0.065	Gradient Boost	 0.862	0.580
1.000	Multi-Layer Perceptron	 0.859	0.599

\_\_\_\_\_



Does clipping CreditScore and Age give better results?

```
In [231]: | ax = sns.boxplot(data=temp scaled X[['CreditScore', 'Age', 'Tenure', 'Balance'
          , 'NumOfProducts', 'EstimatedSalary']])
          ax.set_xticklabels(ax.get_xticklabels(),rotation=45)
          plt.show()
          q1 = scaled_X['Age'].quantile(0.25)
          q3 = scaled_X['Age'].quantile(0.75)
          avg = scaled_X['Age'].mean()
          iqr = q3 - q1
          themin, themax = avg - 1.5 * iqr, avg + 1.5 * iqr
          temp_scaled_X = scaled_X.copy()
          temp_scaled_X[['Age']] = scaled_X[['Age']].clip(themin, themax)
          q1 = scaled_X['CreditScore'].quantile(0.25)
          q3 = scaled_X['CreditScore'].quantile(0.75)
          avg = scaled X['CreditScore'].mean()
          iqr = q3 - q1
          themin, themax = avg - 1.5 * iqr, avg + 1.5 * iqr
          temp scaled X[['CreditScore']] = scaled X[['CreditScore']].clip(themin, themax
          )
          ax = sns.boxplot(data=temp_scaled_X[['CreditScore', 'Age', 'Tenure', 'Balance'
          , 'NumOfProducts', 'EstimatedSalary']])
          ax.set xticklabels(ax.get xticklabels(),rotation=45)
          plt.show()
          ax = sns.boxplot(data=scaled X[['CreditScore', 'Age', 'Tenure', 'Balance', 'Nu
          mOfProducts', 'EstimatedSalary']])
          ax.set_xticklabels(ax.get_xticklabels(),rotation=45)
          saved scaled X = scaled X
          scaled_X = temp_scaled_X
          compare basic models(run basic model)
          scaled X = saved scaled X
```



DecisionTreeClassifier <function compare\_basic\_models.<locals>.<lambda> at 0x
000001D921ED7550>

KNeighborsClassifier <function compare\_basic\_models.<locals>.<lambda> at 0x00
0001D85C0DAA60>

NearestCentroid <function compare\_basic\_models.<locals>.<lambda> at 0x000001D
92285CEE0>

NaiveBayes <function compare\_basic\_models.<locals>.<lambda> at 0x000001D92285 CE50>

Support Vector Machines <function compare\_basic\_models.<locals>.<lambda> at 0
x000001D92285CF70>

Random Forest <function compare\_basic\_models.<locals>.<lambda> at 0x000001D92
6084160>

Ada Boost <function compare\_basic\_models.<locals>.<lambda> at 0x000001D926084 1F0>

Gradient Boost <function compare\_basic\_models.<locals>.<lambda> at 0x000001D9
26084280>

Linear Discriminant Analysis <function compare\_basic\_models.<locals>.<lambda>
at 0x000001D926084310>

Quadriatic Discriminant Analysis <function compare\_basic\_models.<locals>.<lam bda> at 0x000001D9260843A0>

Logistic Regression <function compare\_basic\_models.<locals>.<lambda> at 0x000
001D926084430>

Ridge Classifier <function compare\_basic\_models.<locals>.<lambda> at 0x000001
D9260844C0>

Bagging Classifier <function compare\_basic\_models.<locals>.<lambda> at 0x0000
01D926084550>

SGD Classifier <function compare\_basic\_models.<locals>.<lambda> at 0x000001D9 260845E0>

Passive Aggressive Classifier <function compare\_basic\_models.<locals>.<lambda
> at 0x000001D926084670>

Perceptron compare\_basic\_models.<locals>.<lambda> at 0x000001D92608

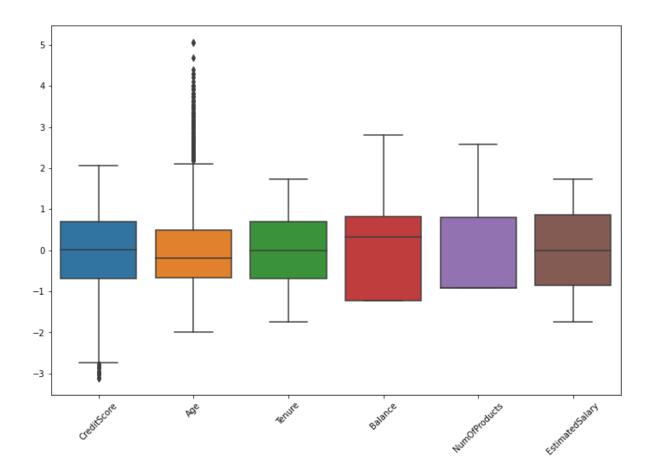
Multi-Layer Perceptron <function compare\_basic\_models.<locals>.<lambda> at 0x 000001D926084790>

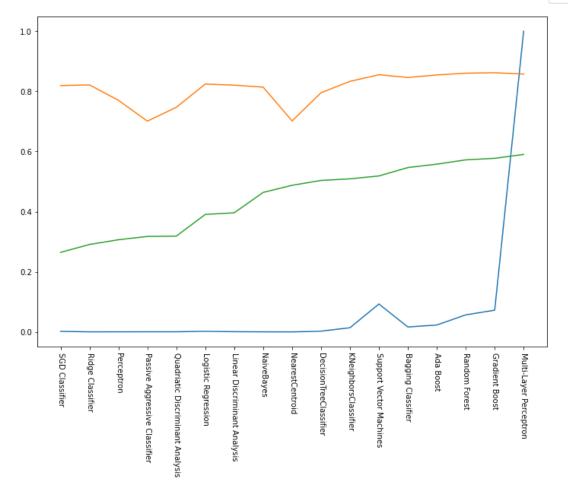
[0.7954008237968253, 0.8332002867133657, 0.701602236575494, 0.81400142540519 5, 0.855100828573781, 0.8604002488817963, 0.8544001283776859, 0.8619000689657 987, 0.8205006057612008, 0.7469090215570331, 0.8242003460052287, 0.8214003057 731791, 0.8461982273413264, 0.8190986052489382, 0.700916419357185, 0.77020796 32092789, 0.857499508501661]

[0.5037048797982363, 0.5089879804640534, 0.48752384058352977, 0.4638115851373 148, 0.5186137509635543, 0.5720896277616712, 0.5578547353621756, 0.5772606442 819923, 0.3960868197866976, 0.3185031625042545, 0.3906023043257644, 0.2906881 081709625, 0.5464874727535429, 0.2644552407389587, 0.3174896123028212, 0.3063 8731279514375, 0.590189636921225]

Time	Model	 Accuracy		Run
0.002	SGD Classifier	 0.819	0.264	
0.000	Ridge Classifier	 0.821	0.291	

	Perceptron	 0.770	0.306
0.000	Passive Aggressive Classifier	 0.701	0.317
0.000	Quadriatic Discriminant Analysis	 0.747	0.319
0.000	•		
0.002	Logistic Regression	0.824	0.391
0.001	Linear Discriminant Analysis	 0.821	0.396
0.000	NaiveBayes	 0.814	0.464
	NearestCentroid	 0.702	0.488
0.000	DecisionTreeClassifier	 0.795	0.504
0.002	KNeighborsClassifier	 0.833	0.509
0.014	Support Vector Machines	 0.855	0.519
0.093			
0.016	Bagging Classifier	0.846	0.546
0.023	Ada Boost	 0.854	0.558
0.057	Random Forest	 0.860	0.572
	Gradient Boost	 0.862	0.577
0.072	Multi-Layer Perceptron	 0.857	0.590
1.000		 	





No, Clipping these column produces worse results.

### **Evaluate the same models with SMOTE**

```
In [ ]:
        def run_basic_model_with_smote(X, y, model_gen_fn):
            accuracy_scores = []
            f1_scores = []
            kf = model_selection.StratifiedKFold(n_splits=6, shuffle=True, random_stat
        e=2)
            for train_index, val_index in kf.split(scaled_X, y):
                model = model_gen_fn()
                smote = SMOTE(random_state=1)
                p = imblearn.pipeline.Pipeline([('smote', smote), ('model', model)])
                p.fit(X.to_numpy()[train_index], y.to_numpy()[train_index])
                y_pred = p.predict(X.to_numpy()[val_index])
                y_true = y.to_numpy()[val_index]
                accuracy_scores.append(metrics.accuracy_score(y_true, y_pred))
                f1_scores.append(metrics.f1_score(y_true, y_pred))
            return mean(accuracy_scores), mean(f1_scores)
```

In [20]: compare\_basic\_models(run\_basic\_model\_with\_smote)

DecisionTreeClassifier <function compare\_basic\_models.<locals>.<lambda> at 0x
000001EC48AF5E50>

KNeighborsClassifier <function compare\_basic\_models.<locals>.<lambda> at 0x00
0001EC48AF5C10>

NearestCentroid <function compare\_basic\_models.<locals>.<lambda> at 0x000001E C48AF5B80>

NaiveBayes <function compare\_basic\_models.<locals>.<lambda> at 0x000001EC48AF 5F70>

Support Vector Machines <function compare\_basic\_models.<locals>.<lambda> at 0
x000001EC48AF5DC0>

Random Forest <function compare\_basic\_models.<locals>.<lambda> at 0x000001EC4
8AF5CA0>

Ada Boost <function compare\_basic\_models.<locals>.<lambda> at 0x000001EC48AF5 AF0>

Gradient Boost <function compare\_basic\_models.<locals>.<lambda> at 0x000001EC
48AF5A60>

Linear Discriminant Analysis <function compare\_basic\_models.<locals>.<lambda>
at 0x000001EC48AF58B0>

Quadriatic Discriminant Analysis <function compare\_basic\_models.<locals>.<lam bda> at 0x000001EC48E015E0>

Logistic Regression <function compare\_basic\_models.<locals>.<lambda> at 0x000
001EC48E01550>

Ridge Classifier <function compare\_basic\_models.<locals>.<lambda> at 0x000001
EC48E014C0>

Bagging Classifier <function compare\_basic\_models.<locals>.<lambda> at 0x0000
01EC48E01040>

SGD Classifier <function compare\_basic\_models.<locals>.<lambda> at 0x000001EC 48E010D0>

Passive Aggressive Classifier <function compare\_basic\_models.<locals>.<lambda
> at 0x000001EC48E011F0>

Perceptron compare\_basic\_models.<locals>.<lambda> at 0x000001EC48E0
1280>

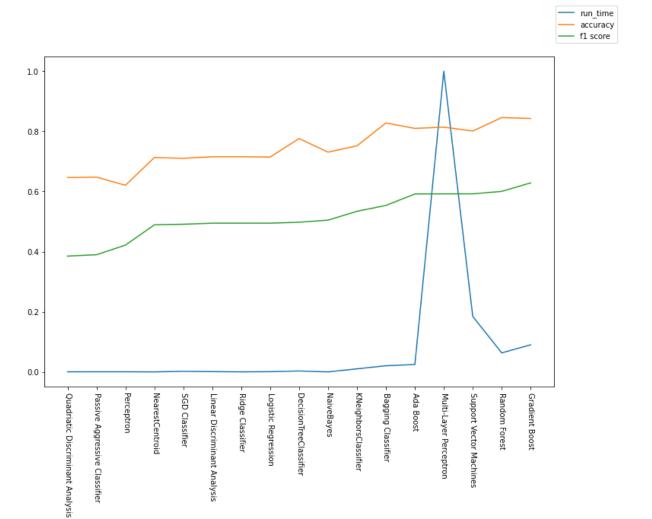
Multi-Layer Perceptron <function compare\_basic\_models.<locals>.<lambda> at 0x 000001EC48E013A0>

[0.7760009822765339, 0.7517983198078272, 0.7129967163830139, 0.73080035853573 59, 0.8009976636125355, 0.8456989874534097, 0.8097986645167965, 0.84259942729 34129, 0.715598536955274, 0.6467847486925185, 0.7144985768752612, 0.715598536 955274, 0.8279003262972856, 0.7103015171275469, 0.6477943907016915, 0.6205905 877648, 0.8139980647327918]

[0.49773909368697317, 0.5340905579876735, 0.4890381273390903, 0.5046504045580 438, 0.5921478562351585, 0.6002998630496208, 0.5916159095475583, 0.6285986183 281258, 0.4944645757888678, 0.38513165717897785, 0.4945295530508275, 0.494464 5757888678, 0.5536137536279351, 0.49110946471286265, 0.3896596518307815, 0.42 158509394340804, 0.5920796851762115]

Run								

	Perceptron	 0.621	0.422
0.000	NearestCentroid	 0.713	0.489
0.000	SGD Classifier	 0.710	0.491
0.002	Linear Discriminant Analysis	 0.716	0.494
0.001			
0.000	Ridge Classifier	0.716	0.494
0.001	Logistic Regression	 0.714	0.495
0.003	DecisionTreeClassifier	 0.776	0.498
0.000	NaiveBayes	 0.731	0.505
0.010	KNeighborsClassifier	 0.752	0.534
	Bagging Classifier	 0.828	0.554
0.021	Ada Boost	 0.810	0.592
0.025	Multi-Layer Perceptron	 0.814	0.592
1.000	Support Vector Machines	 0.801	0.592
0.184	Random Forest	0.846	0.600
0.063	Gradient Boost		0.629
0.090		0.843	



# **Hyper Parameter Optimization**

Test GradientBoost with GridSearchCV

```
In [42]: | def grid_search_gradient_boosting(X, y):
             sm = SMOTE(random state=1)
             gbc = GradientBoostingClassifier(random state=1)
             p = imblearn.pipeline.Pipeline([('sm', sm), ('gbc', gbc)])
             param grid = {
                  'gbc__n_estimators': [10, 100, 1000],
                 #'gbc__loss': ['deviance', 'exponential'],
                  'gbc__max_depth': [3, 10],
                  'gbc__min_samples_split': [i for i in range(2, 210, 40)],
                  'gbc__max_features': [None, 'auto', 'sqrt', 'log2'],
             }
             scorer = make_scorer(metrics.f1_score)
             cv1 = StratifiedKFold(n_splits=6, random_state=10, shuffle=True)
             search = GridSearchCV(p, param_grid, n_jobs=7, cv=cv1, scoring=scorer)
             search.fit(X, y)
             print("GRADIENT BOOSTING")
             print(f"Best Params = {search.best params }")
             print(f"Best score = {search.best_score_}")
             print('-' * 120)
             return search.best_estimator_
         random.seed(22)
         gbc = grid_search_gradient_boosting(scaled_X.to_numpy(), y.to_numpy())
```

#### GRADIENT BOOSTING

```
Best Params = {'gbc__max_depth': 3, 'gbc__max_features': None, 'gbc__min_samp
les_split': 42, 'gbc__n_estimators': 100}
Best score = 0.6311883955640584
```

-----

```
In [176]: def grid search gradient boosting2(X, y):
              sm = SMOTE(random state=1)
              gbc = GradientBoostingClassifier(random state=1)
              p = imblearn.pipeline.Pipeline([('sm', sm), ('gbc', gbc)])
              param grid = {
                   'gbc__n_estimators': [10, 100, 1000, 10000],
                  #'gbc__loss': ['deviance', 'exponential'],
                  'gbc max depth': [3],
                   'gbc__min_samples_split': [42],
                   'gbc__max_features': [None],
                   'gbc learning rate': [0.01, 0.05, 0.1, 0.15, 0.2]
              }
              scorer = make_scorer(metrics.f1_score)
              cv1 = StratifiedKFold(n splits=6, random state=10, shuffle=True)
              search = GridSearchCV(p, param_grid, n_jobs=7, cv=cv1, scoring=scorer)
              search.fit(X, y)
              print("GRADIENT BOOSTING")
              print(f"Best Params = {search.best_params_}")
              print(f"Best score = {search.best_score_}")
              print('-' * 120)
              return search.best_estimator_
          random.seed(22)
          gbc = grid_search_gradient_boosting2(scaled_X.to_numpy(), y.to_numpy())
          GRADIENT BOOSTING
          Best Params = {'gbc_learning_rate': 0.1, 'gbc_max_depth': 3, 'gbc_max_feat
          ures': None, 'gbc__min_samples_split': 42, 'gbc__n_estimators': 100}
          Best score = 0.6311883955640584
```

Test RandomForest with GridSearchCV

```
In [36]: def grid_search_random_forest(X, y):
             sm = SMOTE(random state=1)
             rfc = RandomForestClassifier(random state=1)
             p = imblearn.pipeline.Pipeline([('sm', sm), ('rfc', rfc)])
             param grid = {
                 'rfc__n_estimators': [10, 100, 1000],
                  'rfc__criterion': ['gini', 'entropy'],
                 'rfc__min_samples_split': [i for i in range(20, 200, 40)],
                 'rfc__max_depth': [None, 5],
                 'rfc__class_weight': ['balanced', 'balanced_subsample']
             #scoring = {'F1': make_scorer(metrics.f1_score), 'ACC': make_scorer(metric
         s.accuracy_score)}
             scorer = make scorer(metrics.f1 score)
             cv1 = StratifiedKFold(n splits=6, random state=10, shuffle=True)
             search = GridSearchCV(p, param_grid, n_jobs=7, cv=cv1, scoring=scorer)
             search.fit(X, y)
             print("RANDOM FOREST")
             print(f"Best Params = {search.best_params_}")
             print(f"Best score = {search.best score }")
             print('-' * 120)
             return search.best_estimator_
         random.seed(22)
         rfc = grid_search_random_forest(scaled_X.to_numpy(), y.to_numpy())
```

#### RANDOM FOREST

```
Best Params = {'rfc_class_weight': 'balanced_subsample', 'rfc_criterion':
'entropy', 'rfc_max_depth': None, 'rfc_min_samples_split': 100, 'rfc_n_est
imators': 1000}
Best score = 0.624375225286892
```

-----

```
In [172]: def grid search random forest2(X, y):
              sm = SMOTE(random state=1)
              rfc = RandomForestClassifier(random state=1)
              p = imblearn.pipeline.Pipeline([('sm', sm), ('rfc', rfc)])
              param grid = {
                  'rfc__n_estimators': [1000, 10000],
                   'rfc__criterion': ['entropy'],
                  'rfc min samples split': [42],
                  'rfc__max_depth': [None, 5],
                  'rfc__class_weight': ['balanced_subsample'],
                   'rfc min impurity decrease': [0, 0.00001, 0.0001, 0.001, 0.01, 0.1]
              }
              #scoring = { 'F1': make scorer(metrics.f1 score), 'ACC': make scorer(metric
          s.accuracy score)}
              scorer = make scorer(metrics.f1 score)
              cv1 = StratifiedKFold(n_splits=6, random_state=10, shuffle=True)
              search = GridSearchCV(p, param_grid, n_jobs=7, cv=cv1, scoring=scorer)
              search.fit(X, y)
              print("RANDOM FOREST")
              print(f"Best Params = {search.best params }")
              print(f"Best score = {search.best score }")
              print('-' * 120)
              return search.best estimator
          random.seed(22)
          rfc = grid search random forest2(scaled X.to numpy(), y.to numpy())
          RANDOM FOREST
          Best Params = {'rfc__class_weight': 'balanced_subsample', 'rfc__criterion':
          'entropy', 'rfc__max_depth': None, 'rfc__min_impurity_decrease': 0.0001, 'rfc
           _min_samples_split': 42, 'rfc__n_estimators': 1000}
          Best score = 0.6212218073500146
```

Test SVC with GridSearchCV

```
In [217]: def grid_search_svc(X, y):
              sm = SMOTE(random state=1)
              svc = SVC(random state=1)
              p = imblearn.pipeline.Pipeline([('sm', sm), ('svc', svc)])
              param grid = {
                   'svc_kernel': ['linear', 'poly', 'rbf', 'sigmoid', 'precomputed'],
                   'svc__class_weight': [None, 'balanced'],
                   'svc__gamma': ['scale', 'auto', 0.001],
                   'sm__k_neighbors': [5, 20],
              }
              scorer = make_scorer(metrics.f1_score)
              cv1 = StratifiedKFold(n_splits=6, random_state=10, shuffle=True)
              search = GridSearchCV(p, param_grid, n_jobs=7, cv=cv1, scoring=scorer)
              search.fit(X, y)
              print("SVC")
              print(f"Best Params = {search.best_params_}")
              print(f"Best score = {search.best_score_}")
              print('-' * 120)
              return search.best_estimator_
          random.seed(22)
          selector = []
          svc = grid_search_svc(scaled_X.to_numpy(), y.to_numpy())
          SVC
          Best Params = {'sm_k_neighbors': 20, 'svc__class_weight': None, 'svc__gamm
          a': 'auto', 'svc__kernel': 'rbf'}
          Best score = 0.5958117676964181
In [218]: | scaled_X.shape
Out[218]: (10000, 13)
```

Test MLP with GridSearchCV

```
In [48]: def grid_search_mlp(X, y):
             sm = SMOTE(random state=1)
             mlp = MLPClassifier(random state=1, n iter no change=100, max iter=10000)
             p = imblearn.pipeline.Pipeline([('sm', sm), ('mlp', mlp)])
             param grid = {
                 'mlp__hidden_layer_sizes': [(100,), (100, 50,),],
                 'mlp activation': ['relu', 'logistic'],
                 'mlp solver': ['adam'],
                 'mlp__learning_rate': ['constant', 'adaptive'],
                 #'mlp_alpha': [0.1, 0.01, 0.001, 0.0001]
             }
             scorer = make_scorer(metrics.f1_score)
             cv1 = StratifiedKFold(n_splits=6, random_state=10, shuffle=True)
             search = GridSearchCV(p, param grid, n jobs=7, cv=cv1, scoring=scorer)
             search.fit(X, y)
             print("MULTI LEVEL PERCEPTRON")
             print(f"Best Params = {search.best params }")
             print(f"Best score = {search.best score }")
             print('-' * 120)
             return search.best estimator
         random.seed(22)
         mlp = grid search mlp(scaled X.to numpy(), y.to numpy())
         MULTI LEVEL PERCEPTRON
         Best Params = {'mlp activation': 'relu', 'mlp hidden layer sizes': (100,),
         'mlp learning rate': 'constant', 'mlp solver': 'adam'}
         Best score = 0.5404128313643786
```

#### **Final Baseline**

```
GRADIENT BOOSTING Best Params = {'gbcmax_depth': 3, 'gbcmax_features': None, 'gbcmin_samples_split': 42, 'gbcn_estimators': 100} Best score = 0.6311883955640584

RANDOM FOREST Best Params = {'rfcclass_weight': 'balanced_subsample', 'rfccriterion': 'entropy', 'rfcmax_depth': None, 'rfcmin_samples_split': 100, 'rfc__n_estimators': 1000} Best score = 0.624375225286892

SVC Best Params = {'smk_neighbors': 20, 'svcclass_weight': None, 'svcgamma': 'auto', 'svckernel': 'rbf'} Best score = 0.5958117676964181

MULTI LEVEL PERCEPTRON Best Params = {'mlpactivation': 'relu', 'mlphidden_layer_sizes': (100,), 'mlplearning_rate': 'constant', 'mlpsolver': 'adam'} Best score = 0.5404128313643786
```

# **Basic Experimentation**

## Are there any outliers detected by Random Forest?

```
In [213]: tdf = read_file()
    tdf = engineer_features2(tdf)
    clf = IsolationForest(contamination=1.0e-2000).fit(StandardScaler().fit_transf
    orm(tdf))
    outliers = clf.predict(tdf)
    out = 0
    inl = 0
    for i in outliers:
        if 1 == i:
              inl += 1
        else:
              out += 1
    print(f"Outliers = {out} in = {inl}")
```

Outliers = 20 in = 9980

```
In [245]:
          # Check the number of items categorized as outliers as the random number is va
          ried
          for r in range(10):
              tdf = read file()
              tdf = engineer_features2(tdf)
               clf = IsolationForest(contamination=0.01, random_state=r, n_estimators=100
          0, n jobs=-1).fit(StandardScaler().fit transform(tdf))
               outliers = clf.predict(tdf)
               tdf = tdf[outliers == 1]
               print(r, tdf.shape)
          print('-' * 80)
          for r in range(10):
               tdf = read file()
              tdf = engineer features2(tdf)
               clf = IsolationForest(contamination=0.001, random_state=r, n_estimators=10
          00, n_jobs=-1).fit(StandardScaler().fit_transform(tdf))
               outliers = clf.predict(tdf)
               tdf = tdf[outliers == 1]
               print(r, tdf.shape)
          0 (155, 14)
          1 (231, 14)
          2 (386, 14)
          3 (540, 14)
          4 (323, 14)
          5 (435, 14)
          6 (147, 14)
          7 (166, 14)
          8 (320, 14)
          9 (272, 14)
          0 (1658, 14)
          1 (3905, 14)
          2 (4221, 14)
          3 (3715, 14)
          4 (2616, 14)
          5 (2760, 14)
          6 (2492, 14)
          7 (2296, 14)
          8 (2865, 14)
          9 (2856, 14)
```

We'll not use isolation forest because it is very sensitive to the random state, and with the same contamination it produces a vastly different output depending on the random value

```
In [272]: def perform_grid_search(X, y, params, param_grid, description):
    p = imblearn.pipeline.Pipeline(params)
    cv1 = StratifiedKFold(n_splits=6, random_state=10, shuffle=True)
    scorer = make_scorer(metrics.f1_score)
    search = GridSearchCV(p, param_grid, n_jobs=7, cv=cv1, scoring=scorer)
    search.fit(X, y)
    print(description)
    print(f"Best Params = {search.best_params_}")
    print(f"Best score = {search.best_score_}")
    print('-' * 120)
    return search.best_estimator_
```

#### Feature Selection, PCA, SVC

```
In [288]: | def grid_search_svc_1(X, y):
              sm = SMOTE(random state=1)
              svc = SVC(random_state=1)
              ms = MinMaxScaler()
              pca = PCA(random state=0)
              rfecv = RFECV(estimator=RandomForestClassifier())
              params = [('sm', sm), ('pca', pca), ('rfe', rfecv), ('svc', svc)]
              param_grid = {
                   'svc__kernel': ['rbf'],
                   'svc__class_weight': [None],
                   'svc gamma': ['auto'],
                   'sm__k_neighbors': [20],
                   'pca n components': range(6, X.shape[1]+1)
              return perform_grid_search(X, y, params, param_grid, description="PCA, RFE
          CV, SVC")
          random.seed(22)
          selector = []
          svc = grid_search_svc_1(scaled_X.to_numpy(), y.to_numpy())
          PCA, RFECV, SVC
          Best Params = {'pca__n_components': 10, 'sm__k_neighbors': 20, 'svc__class_we
          ight': None, 'svc__gamma': 'auto', 'svc__kernel': 'rbf'}
```

Best score = 0.5963218633254905

-----

Select the best parameters from above, and run them again, this time without feature selection

```
In [283]: def grid_search_svc_1(X, y):
              sm = SMOTE(random state=1)
              svc = SVC(random state=1)
              ms = MinMaxScaler()
              pca = PCA(random_state=0)
              rfecv = RFECV(estimator=RandomForestClassifier())
              params = [('sm', sm), ('pca', pca), ('svc', svc)]
              param_grid = {
                   'svc kernel': ['rbf'],
                   'svc__class_weight': [None],
                   'svc__gamma': ['auto'],
                   'sm k neighbors': [20],
                   'pca__n_components': range(10, X.shape[1]+1)
              return perform grid search(X, y, params, param grid, description="PCA, SV
          C")
          random.seed(22)
          selector = []
          svc = grid_search_svc_1(scaled_X.to_numpy(), y.to_numpy())
          PCA, RFECV, SVC
          Best Params = {'pca__n_components': 10, 'sm__k_neighbors': 20, 'svc__class_we
          ight': None, 'svc__gamma': 'auto', 'svc__kernel': 'rbf'}
          Best score = 0.5963218633254905
```

Do the same steps again, except use an SVC classifier for feature selection

```
In [286]: def grid search svc 1(X, y):
             sm = SMOTE(random state=1)
             svc = SVC(random state=1)
             ms = MinMaxScaler()
             pca = PCA(random state=0)
             rfecv = RFECV(estimator=DecisionTreeClassifier())
             params = [('sm', sm), ('pca', pca), ('rfe', rfecv), ('svc', svc)]
             param grid = {
                  'svc_kernel': ['rbf'],
                 'svc__class_weight': [None],
                 'svc gamma': ['auto'],
                  'sm__k_neighbors': [20],
                 'pca__n_components': range(10, X.shape[1]+1)
             return perform_grid_search(X, y, params, param_grid, description="PCA, RFE
          CV, SVC")
          random.seed(22)
          selector = []
          svc = grid search svc 1(scaled X.to numpy(), y.to numpy())
         PCA, RFECV, SVC
          Best Params = {'pca__n_components': 13, 'sm__k_neighbors': 20, 'svc__class_we
          ight': None, 'svc gamma': 'auto', 'svc kernel': 'rbf'}
          Best score = 0.5967162189237114
In [287]: print(svc.named_steps['rfe'].support_)
          for i, j in zip(svc.named steps['rfe'].support , scaled X.columns):
             print(j, i)
          Falsel
         CreditScore True
         Age True
          Tenure True
          Balance True
         NumOfProducts True
         HasCrCard True
          IsActiveMember True
          EstimatedSalary True
         Gender Female True
         Gender Male True
         Geography_France True
          Geography_Germany False
          Geography_Spain False
```

```
In [281]:
          def grid search random forest fs(X, y):
              sm = SMOTE(random state=1)
              rf = RandomForestClassifier(random state=1, criterion='entropy', n estimat
          ors=1000, class weight='balanced subsample')
              rfecv = RFECV(estimator=RandomForestClassifier())
              param_grid = {'rf__min_samples_split': [100, 150]}
              params = [('sm', sm), ('rf', rf)]
              return perform_grid_search(X, y, params, param_grid, description="Random F
          orest Classifier with Feature Selection")
          random.seed(22)
          rfcfs = grid_search_random_forest_fs(scaled_X.to_numpy(), y.to_numpy())
          Random Forest Classifier with Feature Selection
          Best Params = {'rf_min_samples_split': 100}
          Best score = 0.624375225286892
In [277]: def grid_search_random_forest_fs(X, y):
              sm = SMOTE(random state=1)
              rf = RandomForestClassifier(random state=1, criterion='entropy', n estimat
          ors=1000, class_weight='balanced_subsample')
              rfecv = RFECV(estimator=RandomForestClassifier())
              param grid = {'rf min samples split': [100, 150]}
              params = [('sm', sm), ('rfecv', rfecv), ('rf', rf)]
              return perform_grid_search(X, y, params, param_grid, description="Random F
          orest Classifier with Feature Selection")
          random.seed(22)
          rfcfs = grid search random forest fs(scaled X.to numpy(), y.to numpy())
          Random Forest Classifier with Feature Selection
          Best Params = {'rf_min_samples_split': 100}
          Best score = 0.6255208031561018
```

```
In [280]: print(svc.named_steps['rfe'].support_)
         for i, j in zip(rfcfs.named_steps['rfecv'].support_, scaled_X.columns):
             print(j, i)
         CreditScore True
         Age True
         Tenure True
         Balance True
         NumOfProducts True
         HasCrCard True
         IsActiveMember True
         EstimatedSalary True
         Gender_Female True
         Gender Male True
         Geography_France True
         Geography_Germany True
         Geography_Spain False
         Index(['CreditScore', 'Age', 'Tenure', 'Balance', 'NumOfProducts', 'HasCrCar
         d',
                'IsActiveMember', 'EstimatedSalary', 'Gender_Female', 'Gender_Male',
                'Geography_France', 'Geography_Germany', 'Geography_Spain'],
               dtype='object')
```

### **Compare Feature Selection Methods**

```
In [164]:
          def feature select score(X, y, selector, model gen fn, model desc, selector de
          sc,\
                                    min max scale=False, do pca=False, n pca=-1):
              p = []
              if min max scale:
                   p.append(('ms', MinMaxScaler(),))
              p.append(('sm', SMOTE(),))
              if None != selector:
                   p.append(('sel', selector(),))
              if do_pca:
                  if -1 == n_pca:
                       n_pca = 10
                   p.append(('pca', PCA(n_components=n_pca),))
              p.append(('model', model_gen_fn(),))
              p = imblearn.pipeline.Pipeline(p)
              y_pred = cross_val_predict(p, X, y, \
                                      cv=model_selection.StratifiedKFold(n_splits=6, shuf
          fle=True, random_state=2),\
                                      n jobs=7
              print(f"Selector: {selector_desc} Model: {model_desc}\n{metrics.classifica
          tion_report(y, y_pred)}")
              return metrics.accuracy_score(y, y_pred), metrics.f1_score(y, y_pred),\
                       metrics.classification_report(y, y_pred), metrics.confusion_matrix
          (y, y_pred)
          best models = [
               ('GradientBoosting', lambda: GradientBoostingClassifier(random_state=0, ma
          x depth=3, min samples split=42, n estimators=100)),
               ('RandomForest', lambda: RandomForestClassifier(class weight='balanced sub
          sample', criterion='entropy', max_depth=None, min_samples_split=100, n_estimat
          ors=1000, random state=0)),
               ('SVC', lambda: SVC(class weight=None, gamma='auto', kernel='rbf', random
          state=0)),
          1
          all feature selectors = [
              ('LinearSVC', lambda: SelectFromModel(LinearSVC())),
               ('LinearSVC L1', lambda: SelectFromModel(LinearSVC(penalty="l1"))),
               ('Variance Threshold 0.8', lambda: VarianceThreshold(threshold=(.8 * (1 -
           .8)))),
               ('RFECV RandomForest', lambda: RFECV(estimator=RandomForestClassifier())),
              ('FWE chi2', lambda: SelectFwe(chi2, alpha=0.01)),
               ('FWE f classif', lambda: SelectFwe(f classif, alpha=0.01)),
               ('Tree Based Feature Selection', lambda: SelectFromModel(ExtraTreesClassif
          ier(n_estimators=50))),
              ('NONE', None),
          ]
          def compare selectors with models(models, selectors,\)
                                            min_max_scale=False, do_pca=False, n_pca=-1):
              For each combination of model and feature selector, run cross validation a
          nd aggregate
              the accuracy and f1 scores
```

```
.....
   all_models = {}
   desc_string = ""
   for model_desc, model_gen in models:
       accuracy_scores = []
       f1_scores = []
       selector_descs = []
       for selector_desc, selector in selectors:
           acc = 0
           f1 = 0
           try:
              acc, f1, report, conf_mat = feature_select_score(scaled_X.to_n
umpy(), y.to_numpy(),\
                                            selector, model_gen, model_desc
, selector_desc,\
                                             min max scale, do pca, n pca)
              accuracy scores.append(acc)
              f1_scores.append(f1)
           except:
              print(f'====>>>> Failure in {model_desc}, {selector_desc}')
              accuracy_scores.append(acc)
              f1 scores.append(f1)
           selector_descs.append(selector_desc)
       sorted_indices = np.argsort(f1_scores)
       desc_string += "\nMODEL: %s\n-----\n" % model_desc
       for i in sorted_indices:
           desc_string += ("%40s %10.5f %10.5f\n" % (selector_descs[i], accur
acy_scores[i], f1_scores[i]))
       desc string += '-----
        ----\n'
       print(desc_string)
       all_models[desc_string] = {'accuracy': accuracy_scores, 'f1': f1_score
s, 'selectors': selector_descs}
   return all_models, desc_string
```

In [124]: all\_models, desc\_string = compare\_selectors\_with\_models(best\_models\_, all\_feat
 ure\_selectors\_)

Selector: Lir	nearSVC Model:	Gradien	tBoosting		
	precision	recall	f1-score	support	
_					
0	0.88	0.86	0.87	7963	
1	0.50	0.56	0.53	2037	
accuracy			0.80	10000	
macro avg	0.69	0.71	0.70	10000	
weighted avg	0.81	0.80	0.80	10000	
0 0					
	lure in Gradie		_		
Selector: Var	riance Thresho				ng
	precision	recall	f1-score	support	
0	0.91	0.89	0.90	7963	
1	0.91 0.61	0.65	0.63	2037	
-	0.01	0.03	0.03	2037	
accuracy			0.84	10000	
macro avg	0.76	0.77	0.76	10000	
weighted avg	0.85	0.84	0.84	10000	
_					
Selector: RFE	CV RandomFore			_	
	precision	recall	f1-score	support	
0	0.91	0.89	0.90	7963	
1	0.61	0.65	0.63	2037	
_					
accuracy			0.84	10000	
macro avg	0.76	0.77	0.76	10000	
weighted avg	0.85	0.84	0.85	10000	
5-47			EUE -1-1		
	lure in Gradie Ef_classif Mo		_		
Selector, FWE	precision		f1-score	support	
	precision	rccarr	11 30010	зиррог с	
0	0.91	0.87	0.89	7963	
1	0.57	0.68	0.62	2037	
accuracy			0.83	10000	
macro avg	0.74	0.78	0.76	10000	
weighted avg	0.85	0.83	0.84	10000	
Selector: Tre	ee Based Featu	re Selec	tion Model:	Gradient	Roosting
SCICCOO. Tro	precision		f1-score	support	DOOSCING
	p. 001010		555. 5	ouppo. c	
0	0.90	0.89	0.89	7963	
1	0.58	0.61	0.60	2037	
accuracy			0.83	10000	
macro avg	0.74	0.75	0.74	10000	
weighted avg	0.83	0.83	0.83	10000	
Selector: NON	NE Model: Grad	lientBoos	ting		
2222237 1401	precision		f1-score	support	
	•	·		11.	
0	0.91	0.89	0.90	7963	
1	0.61	0.65	0.63	2037	

accuracy macro avg		0.77		10000		
weighted avg	0.85	0.84	0.85	10000		
MODEL: Gradie	ntBoosting					
T		FW Lin ature Sel FWE f_c e Thresho	ection lassif NONE ld 0.8		0.62396 0.62750 0.62822	
Selector: Lin						
	precision	recall	+1-score	support		
0			0.88			
1	0.53	0.50	0.52	2037		
accuracy			0.81	10000		
macro avg	0.70	0.69	0.70	10000		
weighted avg	0.80	0.81	0.81	10000		
====>>>> Fail						
Selector: Var						
	precision	recall	†1-score	support		
0			0.89			
1	0.59	0.66	0.62	2037		
accuracy			0.84	10000		
macro avg	0.75	0.77	0.76	10000		
weighted avg	0.84	0.84	0.84	10000		
Selector: RFE	CV RandomFore	est Model	: RandomFo	orest		
3616660111112	precision		f1-score	support		
0	0.91	0.88	0.90	7963		
1	0.59	0.66	0.62	2037		
20011201			0.84	10000		
accuracy macro avg	0.75	0.77	0.84	10000		
weighted avg	0.84	0.84	0.84	10000		
====>>>> Fail		-				
Selector: FWE	_			cunnont		
	precision	recall	f1-score	support		
0	0.91	0.87	0.89	7963		
1	0.58	0.68	0.63	2037		

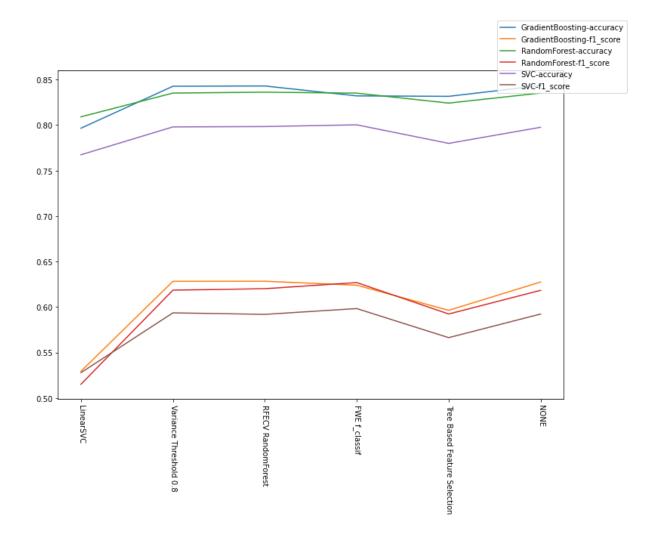
accura	асу			0.84	10000		
macro a	avg	0.75	0.78	0.76	10000		
weighted a	avg	0.85	0.84	0.84	10000		
Salactor	Troo	Based Featu	ıra Salac	tion Mode	l· RandomFo	nast	
Selector.		precision				1 63 6	
	ŀ	01 60131011	recarr	11-30016	зиррог с		
	0	0.90	0.87	0.89	7963		
	1		0.63				
accura	асу			0.82	10000		
macro a	avg	0.73	0.75	0.74	10000		
weighted a	avg	0.83	0.82	0.83	10000		
Solocton	NONE	Model: Rand	domEanast				
Selector.		orecision			support		
	ŀ	J. CC131011	, ccarr	11 30016	Suppor C		
	0	0.91	0.88	0.89	7963		
	1	0.59		0.62	2037		
accura	асу			0.84			
macro a	_	0.75					
weighted a	avg	0.84	0.84	0.84	10000		
			FW	SVC L1 E chi2	0.00000 0.00000 0.79660	0.00000 0.00000 0.52938	
		ee Based Fea	FW Lin ature Sel	E chi2 earSVC ection lassif	0.00000 0.79660 0.83160 0.83210	0.00000 0.52938 0.59636 0.62396	
		ee Based Fea	FW Lin ature Sel FWE f_c	E chi2 earSVC ection lassif NONE	0.00000 0.79660 0.83160 0.83210 0.84340	0.00000 0.52938 0.59636 0.62396 0.62750	
		ee Based Fea Variance	FW Lin sture Sel FWE f_c	E chi2 earSVC ection lassif NONE ld 0.8	0.00000 0.79660 0.83160 0.83210 0.84340 0.84270	0.00000 0.52938 0.59636 0.62396 0.62750 0.62822	
		ee Based Fea Variance	FW Lin ature Sel FWE f_c	E chi2 earSVC ection lassif NONE ld 0.8	0.00000 0.79660 0.83160 0.83210 0.84340	0.00000 0.52938 0.59636 0.62396 0.62750	
		ee Based Fea Variance	FW Lin sture Sel FWE f_c	E chi2 earSVC ection lassif NONE ld 0.8	0.00000 0.79660 0.83160 0.83210 0.84340 0.84270	0.00000 0.52938 0.59636 0.62396 0.62750 0.62822	
	Tre	ee Based Fea Variance RFEG	FW Lin sture Sel FWE f_c	E chi2 earSVC ection lassif NONE ld 0.8	0.00000 0.79660 0.83160 0.83210 0.84340 0.84270	0.00000 0.52938 0.59636 0.62396 0.62750 0.62822	
	Tre	ee Based Fea Variance RFEG	FW Lin sture Sel FWE f_c	E chi2 earSVC ection lassif NONE ld 0.8	0.00000 0.79660 0.83160 0.83210 0.84340 0.84270	0.00000 0.52938 0.59636 0.62396 0.62750 0.62822	
	Tre	ee Based Fea Variance RFEG	FW Lin ature Sel FWE f_c e Thresho CV Random	E chi2 earSVC ection lassif NONE ld 0.8	0.00000 0.79660 0.83160 0.83210 0.84340 0.84270	0.00000 0.52938 0.59636 0.62396 0.62750 0.62822	
	Tre	ee Based Fea Variance RFEG	FW Lin ature Sel FWE f_c Thresho CV Random	E chi2 earSVC ection lassif NONE ld 0.8 Forest	0.00000 0.79660 0.83160 0.83210 0.84340 0.84270 0.84300	0.00000 0.52938 0.59636 0.62396 0.62750 0.62822 0.62831	
	Tre	ee Based Fea Variance RFEG	FW Lin ature Sel FWE f_c Thresho CV Random Linear FW	E chi2 earSVC ection lassif NONE ld 0.8 Forest	0.00000 0.79660 0.83160 0.83210 0.84340 0.84270 0.84300	0.00000 0.52938 0.59636 0.62396 0.62750 0.62822 0.62831	
	Tre	ee Based Fea Variance RFEG	FW Lin ature Sel FWE f_c e Thresho CV Random  Linear FW Lin	E chi2 earSVC ection lassif NONE ld 0.8 Forest SVC L1 E chi2 earSVC	0.00000 0.79660 0.83160 0.83210 0.84340 0.84270 0.84300	0.00000 0.52938 0.59636 0.62396 0.62750 0.62822 0.62831	
	Tre	Variance RFEC	FW Lin ature Sel FWE f_c e Thresho CV Random  Linear FW Lin	E chi2 earSVC ection lassif NONE ld 0.8 Forest SVC L1 E chi2 earSVC	0.00000 0.79660 0.83160 0.83210 0.84340 0.84270 0.84300 	0.00000 0.52938 0.59636 0.62396 0.62750 0.62822 0.62831 	
	Tre	Variance RFEC	FW Lin ature Sel FWE f_c e Thresho CV Random  Linear FW Lin	E chi2 earSVC ection lassif NONE ld 0.8 Forest SVC L1 E chi2 earSVC ection NONE	0.00000 0.79660 0.83160 0.83210 0.84340 0.84270 0.84300 	0.00000 0.52938 0.59636 0.62396 0.62750 0.62822 0.62831 	
	Tre	Variance RFEC	FW Lin ature Sel FWE f_c  Thresho CV Random Linear FW Lin ature Sel	E chi2 earSVC ection lassif NONE ld 0.8 Forest SVC L1 E chi2 earSVC ection NONE ld 0.8	0.00000 0.79660 0.83160 0.83210 0.84340 0.84270 0.84300 	0.00000 0.52938 0.59636 0.62396 0.62750 0.62822 0.62831 	
	Tre	Variance RFEC	FW Line ature Sel FWE f_c  Thresho CV Random Line FW Line ature Sel  Thresho CV Random	E chi2 earSVC ection lassif NONE ld 0.8 Forest SVC L1 E chi2 earSVC ection NONE ld 0.8 Forest	0.00000 0.79660 0.83160 0.83210 0.84340 0.84270 0.84300 	0.00000 0.52938 0.59636 0.62396 0.62750 0.62822 0.62831 	
	Tre	Variance RFEC	FW Line ature Sel FWE f_c  Thresho CV Random Line FW Line ature Sel  Thresho CV Random	E chi2 earSVC ection lassif NONE ld 0.8 Forest SVC L1 E chi2 earSVC ection NONE ld 0.8 Forest	0.00000 0.79660 0.83160 0.83210 0.84340 0.84270 0.84300 	0.00000 0.52938 0.59636 0.62396 0.62750 0.62822 0.62831 	
	Tre	Variance RFEC	FW Line ature Sel FWE f_c  Thresho CV Random Line FW Line ature Sel  Thresho CV Random	E chi2 earSVC ection lassif NONE ld 0.8 Forest SVC L1 E chi2 earSVC ection NONE ld 0.8 Forest	0.00000 0.79660 0.83160 0.83210 0.84340 0.84270 0.84300 	0.00000 0.52938 0.59636 0.62396 0.62750 0.62822 0.62831 	
MODEL: Rar	Tre	Variance RFEC	FW Line ature Sel FWE f_c  Thresho CV Random FW Line ature Sel  Thresho CV Random FWE f_c	E chi2 earSVC ection lassif NONE ld 0.8 Forest SVC L1 E chi2 earSVC ection NONE ld 0.8 Forest	0.00000 0.79660 0.83160 0.83210 0.84340 0.84270 0.84300 	0.00000 0.52938 0.59636 0.62396 0.62750 0.62822 0.62831 	
MODEL: Rar	Tre	Variance RFECTOR  Orest  Variance RFECTOR  ARSVC Model:	FW Lin ature Sel FWE f_c e Thresho CV Random Linear FW Lin ature Sel e Thresho CV Random FWE f_c	E chi2 earSVC ection lassif NONE ld 0.8 Forest SVC L1 E chi2 earSVC ection NONE ld 0.8 Forest ld 0.8	0.00000 0.79660 0.83160 0.83210 0.84340 0.84270 0.84300 	0.00000 0.52938 0.59636 0.62396 0.62750 0.62822 0.62831 	
MODEL: Rar	Tre	Variance RFEC	FW Lin ature Sel FWE f_c e Thresho CV Random Linear FW Lin ature Sel e Thresho CV Random FWE f_c	E chi2 earSVC ection lassif NONE ld 0.8 Forest SVC L1 E chi2 earSVC ection NONE ld 0.8 Forest	0.00000 0.79660 0.83160 0.83210 0.84340 0.84270 0.84300 	0.00000 0.52938 0.59636 0.62396 0.62750 0.62822 0.62831 	
MODEL: Rar	Tre	Variance RFECTOR  Orest  Variance RFECTOR  ARSVC Model:	FW Lin ature Sel FWE f_c e Thresho CV Random Linear FW Lin ature Sel e Thresho CV Random FWE f_c	E chi2 earSVC ection lassif NONE ld 0.8 Forest SVC L1 E chi2 earSVC ection NONE ld 0.8 Forest ld 0.8	0.00000 0.79660 0.83160 0.83210 0.84340 0.84270 0.84300 	0.00000 0.52938 0.59636 0.62396 0.62750 0.62822 0.62831 	

-			0.77	10000
macro avg	0.67	0.72	0.69	10000
weighted avg	0.81	0.77	0.78	10000
====>>>> Fail				
Selector: Var	precision	recall	f1-score	cuppont
	precision	recarr	11-2001-6	support
0	0.92	0.82	0.87	7963
1	0.50	0.72	0.59	2037
accuracy			0.80	10000
macro avg	0.71	0.77	0.73	10000
weighted avg	0.84	0.80	0.81	10000
Selector: RFE	CV PandomFon	ost Model	. 51/6	
Selector: KFE	precision		f1-score	support
	precision	IECAII	11-30016	suppor c
0	0.92	0.82	0.87	7963
1	0.50	0.72	0.59	2037
		• • • •		
accuracy			0.80	10000
macro avg	0.71	0.77	0.73	10000
weighted avg	0.83	0.80	0.81	10000
====>>>> Fail				
Selector: FWE	_			
	precision	recall	f1-score	support
0	0.92	0.00	0.87	7063
U				/463
		0.82 0.73		7963 2037
1	0.51	0.73	0.60	7963 2037
1			0.60	2037
1 accuracy	<ul><li>0.51</li><li>0.71</li></ul>	0.73	0.60 0.80	2037 10000
accuracy macro avg weighted avg	0.51 0.71 0.84	0.73 0.77 0.80	0.60 0.80 0.73 0.81	2037 10000 10000 10000
accuracy macro avg	0.51 0.71 0.84 e Based Feat	0.73 0.77 0.80 ure Selec	0.60 0.80 0.73 0.81 tion Model:	2037 10000 10000 10000 SVC
accuracy macro avg weighted avg	0.51 0.71 0.84	0.73 0.77 0.80 ure Selec	0.60 0.80 0.73 0.81	2037 10000 10000 10000
accuracy macro avg weighted avg Selector: Tre	0.51 0.71 0.84 e Based Feat precision	0.73 0.77 0.80 ure Selec recall	0.60 0.80 0.73 0.81 tion Model: f1-score	2037 10000 10000 10000 SVC support
accuracy macro avg weighted avg Selector: Tre	0.51 0.71 0.84 e Based Feat precision 0.91	0.73 0.77 0.80 ure Selec recall 0.80	0.60 0.80 0.73 0.81 tion Model: f1-score 0.85	2037 10000 10000 10000 SVC support 7963
accuracy macro avg weighted avg Selector: Tre	0.51 0.71 0.84 e Based Feat precision	0.73 0.77 0.80 ure Selec recall	0.60 0.80 0.73 0.81 tion Model: f1-score	2037 10000 10000 10000 SVC support
accuracy macro avg weighted avg Selector: Tre	0.51 0.71 0.84 e Based Feat precision 0.91	0.73 0.77 0.80 ure Selec recall 0.80	0.60 0.80 0.73 0.81 tion Model: f1-score 0.85 0.57	2037 10000 10000 10000 SVC support 7963 2037
accuracy macro avg weighted avg  Selector: Tre  0 1 accuracy	0.51 0.71 0.84 e Based Feat precision 0.91 0.47	0.73 0.77 0.80 Ture Selectine recall 0.80 0.71	0.60 0.80 0.73 0.81 tion Model: f1-score 0.85 0.57 0.78	2037 10000 10000 10000 SVC support 7963 2037 10000
accuracy macro avg weighted avg Selector: Tre	0.51 0.71 0.84 e Based Feat precision 0.91	0.73 0.77 0.80 ure Selec recall 0.80	0.60 0.80 0.73 0.81 tion Model: f1-score 0.85 0.57	2037 10000 10000 10000 SVC support 7963 2037
accuracy macro avg weighted avg Selector: Tre  0 1 accuracy macro avg	0.51 0.71 0.84 e Based Feat precision 0.91 0.47	0.73 0.77 0.80 Ture Selection recall 0.80 0.71	0.60 0.80 0.73 0.81 tion Model: f1-score 0.85 0.57 0.78 0.71	2037 10000 10000 10000 SVC support 7963 2037 10000 10000
accuracy macro avg weighted avg Selector: Tre  0 1 accuracy macro avg	0.51 0.71 0.84 e Based Feat precision 0.91 0.47 0.69 0.82	0.73 0.77 0.80 ure Selectorecall 0.80 0.71 0.75 0.78	0.60 0.80 0.73 0.81 tion Model: f1-score 0.85 0.57 0.78 0.71 0.79	2037 10000 10000 10000 SVC support 7963 2037 10000 10000
accuracy macro avg weighted avg  Selector: Tre  0 1 accuracy macro avg weighted avg	0.51 0.71 0.84 e Based Feat precision 0.91 0.47 0.69 0.82	0.73 0.77 0.80 ure Selectorecall 0.80 0.71 0.75 0.78	0.60 0.80 0.73 0.81 tion Model: f1-score 0.85 0.57 0.78 0.71	2037 10000 10000 SVC support 7963 2037 10000 10000
accuracy macro avg weighted avg  Selector: Tre  0 1 accuracy macro avg weighted avg  Selector: NON	0.51  0.71 0.84  e Based Feat precision  0.91 0.47  0.69 0.82  E Model: SVC precision	0.73 0.77 0.80 ure Selectorecall 0.80 0.71 0.75 0.78	0.60 0.80 0.73 0.81 tion Model: f1-score 0.85 0.57 0.78 0.71 0.79	2037 10000 10000 SVC support 7963 2037 10000 10000 support
accuracy macro avg weighted avg  Selector: Tre  0 1 accuracy macro avg weighted avg  Selector: NON	0.51 0.71 0.84 e Based Feat precision 0.91 0.47 0.69 0.82 E Model: SVC precision 0.92	0.73 0.77 0.80 Ture Selectorecall 0.80 0.71 0.75 0.78 recall 0.82	0.60 0.80 0.73 0.81 tion Model: f1-score 0.85 0.57 0.78 0.71 0.79 f1-score 0.87	2037 10000 10000 SVC support 7963 2037 10000 10000 10000 support
accuracy macro avg weighted avg  Selector: Tre  0 1 accuracy macro avg weighted avg  Selector: NON	0.51  0.71 0.84  e Based Feat precision  0.91 0.47  0.69 0.82  E Model: SVC precision	0.73 0.77 0.80 ure Selectorecall 0.80 0.71 0.75 0.78	0.60 0.80 0.73 0.81 tion Model: f1-score 0.85 0.57 0.78 0.71 0.79	2037 10000 10000 SVC support 7963 2037 10000 10000 support
accuracy macro avg weighted avg  Selector: Tre  0 1 accuracy macro avg weighted avg  Selector: NON  0 1	0.51 0.71 0.84 e Based Feat precision 0.91 0.47 0.69 0.82 E Model: SVC precision 0.92	0.73 0.77 0.80 Ture Selectorecall 0.80 0.71 0.75 0.78 recall 0.82	0.60 0.80 0.73 0.81 tion Model: f1-score 0.85 0.57 0.78 0.71 0.79 f1-score 0.87 0.59	2037 10000 10000 SVC support 7963 2037 10000 10000 support 7963 2037
accuracy macro avg weighted avg  Selector: Tre  0 1 accuracy macro avg weighted avg  Selector: NON  0 1 accuracy	0.51  0.71 0.84  e Based Feat precision  0.91 0.47  0.69 0.82  E Model: SVC precision  0.92 0.50	0.73 0.77 0.80 ure Selectorecall 0.80 0.71 0.75 0.78 recall 0.82 0.72	0.60 0.80 0.73 0.81 tion Model: f1-score 0.85 0.57 0.78 0.71 0.79 f1-score 0.87 0.59 0.80	2037 10000 10000 SVC support 7963 2037 10000 support 7963 2037 10000
accuracy macro avg weighted avg  Selector: Tre  0 1 accuracy macro avg weighted avg  Selector: NON  0 1	0.51 0.71 0.84 e Based Feat precision 0.91 0.47 0.69 0.82 E Model: SVC precision 0.92	0.73 0.77 0.80 Ture Selectorecall 0.80 0.71 0.75 0.78 recall 0.82	0.60 0.80 0.73 0.81 tion Model: f1-score 0.85 0.57 0.78 0.71 0.79 f1-score 0.87 0.59	2037 10000 10000 SVC support 7963 2037 10000 10000 support 7963 2037

#### MODEL: GradientBoosting LinearSVC L1 0.00000 0.00000 FWE chi2 0.00000 0.00000 LinearSVC 0.79660 0.52938 Tree Based Feature Selection 0.83160 0.59636 FWE f\_classif 0.83210 0.62396 NONE 0.84340 0.62750 Variance Threshold 0.8 0.84270 0.62822 RFECV RandomForest 0.84300 0.62831 MODEL: RandomForest LinearSVC L1 0.00000 0.00000 FWE chi2 0.00000 0.00000 LinearSVC 0.80910 0.51511 Tree Based Feature Selection 0.82420 0.59230 NONE 0.83520 0.61834 Variance Threshold 0.8 0.83520 0.61852 RFECV RandomForest 0.83620 0.62013 FWE f\_classif 0.83510 0.62684 ------MODEL: SVC LinearSVC L1 0.00000 0.00000 FWE chi2 0.00000 0.00000 LinearSVC 0.76740 0.52781 Tree Based Feature Selection 0.77990 0.56631 RFECV RandomForest 0.79840 0.59190 NONE 0.79760 0.59226 Variance Threshold 0.8 0.79800 0.59356 FWE f\_classif 0.80030 0.59827

- -

```
In [161]: def plot_graph_1(all_models):
              Plot the accuracy and F1 scores from the data output by the previous step
              am = ['GradientBoosting', 'RandomForest', 'SVC']
              i = 0
              accuracy_scores = None
              f1 scores = None
              selector_descs = None
              fig, ax = plt.subplots()
              for key, val in all_models.items():
                  model = am[i]
                  i+=1
                  accuracy_scores = val['accuracy']
                  f1_scores = val['f1']
                  selector_descs = val['selectors']
                  new_f1 = []
                  new acc = []
                  new_descs = []
                  for acc, f1, d in zip(accuracy_scores, f1_scores, selector_descs):
                       if 0 != f1:
                           new_f1.append(f1)
                           new_acc.append(acc)
                           new descs.append(d)
                   accuracy scores = new acc
                  f1_scores = new_f1
                  selector descs = new descs
                   ax.plot(accuracy_scores, label="%s-accuracy" % model)
                   ax.plot(f1_scores, label="%s-f1_score" % model)
              plt.xticks(np.arange(0, len(f1_scores), 1.0))
              ax.set_xticklabels(selector_descs, rotation=270)
              fig.legend()
          plot graph 1(all models)
```



```
In [126]:
         print("Best models with Feature Selection")
         print(desc string)
         Best models with Feature Selection
         MODEL: GradientBoosting
                                  LinearSVC L1
                                                0.00000
                                                          0.00000
                                      FWE chi2
                                                0.00000
                                                          0.00000
                                     LinearSVC
                                                0.79660
                                                          0.52938
                    Tree Based Feature Selection
                                                0.83160 0.59636
                                 FWE f classif
                                                0.83210 0.62396
                                         NONE
                                                0.84340 0.62750
                         Variance Threshold 0.8
                                                0.84270
                                                          0.62822
                             RFECV RandomForest
                                                0.84300
                                                          0.62831
         MODEL: RandomForest
         ______
                                  LinearSVC L1 0.00000
                                                          0.00000
                                     FWE chi2 0.00000 0.00000
                                     LinearSVC
                                                0.80910 0.51511
                    Tree Based Feature Selection
                                                0.82420 0.59230
                                         NONE
                                                0.83520
                                                          0.61834
                         Variance Threshold 0.8 0.83520 0.61852
                             RFECV RandomForest 0.83620 0.62013
                                 FWE f_classif
                                                0.83510
                                                          0.62684
         MODEL: SVC
                                  LinearSVC L1
                                                0.00000
                                                          0.00000
                                      FWE chi2
                                                0.00000
                                                          0.00000
                                     LinearSVC
                                                0.76740 0.52781
                    Tree Based Feature Selection 0.77990
                                                        0.56631
                             RFECV RandomForest
                                                0.79840
                                                        0.59190
                                         NONE
                                                0.79760
                                                          0.59226
                         Variance Threshold 0.8
                                                0.79800
                                                          0.59356
                                 FWE f_classif
                                                0.80030
                                                        0.59827
```

The best feature selectors are:

- 1. FWE f classif
- 2. NONE
- 3. RFECV Random Forest
- 4. Variance Threshold 0.8

However the difference is not much

Check if MinMaxScaling makes a difference

See if MinMax Scaling makes a difference. Standardization has already been performed earlier with unit SD

[('ms', MinMaxScaler()), ('sm', SMOTE()), ('sel', VarianceThreshold(threshold =0.159999999999999)), ('model', GradientBoostingClassifier(min\_samples\_spli t=42, random state=0))] Selector: Variance Threshold 0.8 Model: GradientBoosting precision recall f1-score support 0 0.86 0.69 0.77 7963 1 0.32 0.57 0.41 2037 0.66 10000 accuracy macro avg 0.59 0.63 0.59 10000 weighted avg 0.75 0.66 10000 0.69 [('ms', MinMaxScaler()), ('sm', SMOTE()), ('sel', RFECV(estimator=RandomFores tClassifier())), ('model', GradientBoostingClassifier(min\_samples\_split=42, r andom state=0))] Selector: RFECV RandomForest Model: GradientBoosting precision recall f1-score support 0 0.91 0.88 0.89 7963 1 0.58 0.67 0.62 2037 0.83 10000 accuracy 0.75 0.77 0.76 10000 macro avg weighted avg 0.84 10000 0.84 0.83 [('ms', MinMaxScaler()), ('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('m odel', GradientBoostingClassifier(min\_samples\_split=42, random\_state=0))] Selector: FWE f classif Model: GradientBoosting precision recall f1-score support 0 0.92 0.86 0.89 7963 1 0.56 0.70 2037 0.62 0.83 10000 accuracy macro avg 0.74 0.78 0.76 10000 weighted avg 0.85 0.83 0.83 10000 [('ms', MinMaxScaler()), ('sm', SMOTE()), ('model', GradientBoostingClassifie r(min samples split=42, random state=0))] Selector: NONE Model: GradientBoosting

	precision	recall	f1-score	support
0	0.91	0.88	0.89	7963
1	0.58	0.67	0.62	2037
accuracy			0.84	10000
macro avg	0.75	0.77	0.76	10000
weighted avg	0.84	0.84	0.84	10000

MODEL: GradientBoosting

-----

Variance Threshold 0.8	0.66370	0.40782
NONE	0.83510	0.62170
RFECV RandomForest	0.83440	0.62278
FWE f_classif	0.82730	0.62383

- -

[('ms', MinMaxScaler()), ('sm', SMOTE()), ('sel', VarianceThreshold(threshold
=0.1599999999999999)), ('model', RandomForestClassifier(class\_weight='balanc
ed\_subsample', criterion='entropy',

min\_samples\_split=100, n\_estimators=1000,
random state=0))]

Selector: Variance Threshold 0.8 Model: RandomForest

	precision	recall	f1-score	support
0	0.86	0.68	0.76	7963
1	0.32	0.58	0.41	2037
accuracy			0.66	10000
macro avg	0.59	0.63	0.58	10000
weighted avg	0.75	0.66	0.69	10000

[('ms', MinMaxScaler()), ('sm', SMOTE()), ('sel', RFECV(estimator=RandomFores
tClassifier())), ('model', RandomForestClassifier(class\_weight='balanced\_subs
ample', criterion='entropy',

min\_samples\_split=100, n\_estimators=1000,
random state=0))]

Selector: RFECV RandomForest Model: RandomForest

	precision	recall	f1-score	support
0	0.92	0.85	0.88	7963
1	0.55	0.69	0.61	2037
accuracy			0.82	10000
macro avg	0.73	0.77	0.75	10000
weighted avg	0.84	0.82	0.83	10000

[('ms', MinMaxScaler()), ('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('m
odel', RandomForestClassifier(class\_weight='balanced\_subsample', criterion='e
ntropy',

min\_samples\_split=100, n\_estimators=1000,
random state=0))]

Selector: FWE f\_classif Model: RandomForest

	idollii oi cac	HOUCE. Rui	1_6103311	SCICCOI. IWL
support	f1-score	recall	precision	
7963	0.88	0.85	0.92	0
2037	0.61	0.71	0.54	1
10000	0.82			accuracy
10000	0.75	0.78	0.73	macro avg
10000	0.83	0.82	0.84	weighted avg

[('ms', MinMaxScaler()), ('sm', SMOTE()), ('model', RandomForestClassifier(cl ass\_weight='balanced\_subsample', criterion='entropy',

min\_samples\_split=100, n\_estimators=1000,
nandom\_state\_0))]

random\_state=0))]

Selector: NONE Model: RandomForest

	precision	recall	f1-score	support
0	0.92	0.85	0.88	7963
1	0.55	0.69	0.61	2037

9		0.77 0.82		10000					
MODEL: GradientBoosting									
			NE	0.83510					
				0.83440 0.82730					
MODEL: RandomFore	st 								
		WE f_class	if	0.65870 0.81750 0.82070	0.61162				
	RFECV			0.82180					
=0.15999999999999999999999999999999999999	998)), ('mo e Thresholo	odel', SVC(	gamma: : SVC	='auto', ra	anceThreshold(threshold ndom_state=0))]				
0 1	0.87 0.31	0.65 0.61	0.74 0.41						
•	0.59 0.75	0.63 0.64		10000					
<pre>[('ms', MinMaxSca tClassifier())), Selector: RFECV R</pre>	('model', 9	SVC(gamma='	auto'		V(estimator=RandomFores ate=0))]				
		recall f1-		support					
0 1	0.92 0.44	0.76 0.73	0.83 0.54						
accuracy macro avg weighted avg	0.68 0.82	0.74 0.75	0.75 0.69 0.77	10000					
<pre>[('ms', MinMaxScaler()), ('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('m odel', SVC(gamma='auto', random_state=0))] Selector: FWE f_classif Model: SVC</pre>									
pre	cision r	recall f1-	score	support					
0 1	0.92 0.45	0.77 0.73	0.84 0.56						
accuracy			0.77	10000					

macro avg weighted avg	0.69 0.82	0.75 0.77						
<pre>[('ms', MinMaxScaler()), ('sm', SMOTE()), ('model', SVC(gamma='auto', random_ state=0))] Selector: NONE Model: SVC</pre>								
Selector: None	precision	recall	f1-score	suppor	t			
0	0.92	0.76	0.83	796	3			
1	0.43	0.73	0.54	203	7			
accuracy			0.75		0			
macro avg	0.67							
weighted avg	0.82	0.75	0.77	1000	Ю			
MODEL: Gradier	ntBoosting							
	Variance	e Thresho	ld 0.8	0.66370	0.40782			
				0.83510				
	RFE		Forest					
		FWE +_c	lassif 	0.82730	0.62383			
MODEL: RandomF								
		- Thresho	1d 0 8	0.65870	0.40798			
	rai zaire		lassif					
		_		0.82070				
	RFE	CV Random	Forest		0.61278			
MODEL: SVC								
	Variance	e Thresho	ld 0.8					
				0.75030				
	RFE		Forest					
		FWE +_c	lassif	0.76500	0.55976			

--

In [162]: plot\_graph\_1(all\_models\_mms)
 print(desc\_string\_mms)

# MODEL: GradientBoosting Variance Threshold 0.8 0.84200 0.62147 FWE f\_classif 0.83520 0.62323 RFECV RandomForest 0.84230 0.62372 NONE 0.84360 0.62726 MODEL: RandomForest \_\_\_\_\_\_ Variance Threshold 0.8 0.83390 0.61363 NONE 0.83510 0.61624 RFECV RandomForest 0.83610 0.61893 FWE f\_classif 0.83390 0.62514 MODEL: SVC 0.79730 Variance Threshold 0.8 0.59158 RFECV RandomForest 0.79850 0.59432 NONE 0.79980 0.59490 FWE f\_classif 0.79810 0.59837 GradientBoosting-accuracy GradientBoosting-f1\_score RandomForest-accuracy RandomForest-f1\_score SVC-accuracy 0.85 SVC-f1\_score 0.80 0.75 0.70 0.65 0.60 Variance Threshold 0.8 RFECV RandomForest FWE f\_classif

## Add PCA to the mix

\_\_\_\_\_\_\_

\_\_\_\_\_

#### PCA 5

\_\_\_\_\_

---

[('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.1599999999999999)),
('pca', PCA(n\_components=10)), ('model', GradientBoostingClassifier(min\_sampl
es\_split=42, random\_state=0))]

Selector: Variance Threshold 0.8 Model: GradientBoosting

	precision	recall	f1-score	support
0	0.92	0.81	0.86	7963
1	0.49	0.72	0.59	2037
accuracy			0.79	10000
macro avg weighted avg	0.71 0.83	0.77 0.79	0.72 0.81	10000 10000

[('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca',
PCA(n\_components=10)), ('model', GradientBoostingClassifier(min\_samples\_split
=42, random\_state=0))]

Selector: RFECV RandomForest Model: GradientBoosting

	precision	recall	f1-score	support
0	0.92	0.83	0.87	7963
1	0.51	0.70	0.59	2037
accuracy			0.80	10000
macro avg	0.71	0.77	0.73	10000
weighted avg	0.83	0.80	0.81	10000

[('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n\_components=1
0)), ('model', GradientBoostingClassifier(min\_samples\_split=42, random\_state=
0))]

====>>>> Failure in GradientBoosting, FWE f classif

[('sm', SMOTE()), ('pca', PCA(n\_components=10)), ('model', GradientBoostingCl
assifier(min\_samples\_split=42, random\_state=0))]

Selector: NONE Model: GradientBoosting

	precision	recall	†1-score	support
0	0.92	0.81	0.86	7963
1	0.49	0.73	0.59	2037
accuracy			0.79	10000
macro avg weighted avg	0.71 0.83	0.77 0.79	0.72 0.80	10000 10000
- 0 0				

#### MODEL: GradientBoosting

-----

FWE f_classif	0.00000	0.00000
NONE	0.78990	0.58552
Variance Threshold 0.8	0.79250	0.58690
RFECV RandomForest	0.80290	0.59285

```
[('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.159999999999999)),
('pca', PCA(n_components=10)), ('model', RandomForestClassifier(class_weight
='balanced_subsample', criterion='entropy',
                       min samples split=100, n estimators=1000,
                       random state=0))]
Selector: Variance Threshold 0.8 Model: RandomForest
              precision
                          recall f1-score
                                              support
           0
                   0.91
                             0.84
                                       0.88
                                                 7963
           1
                   0.52
                             0.69
                                       0.60
                                                 2037
                                       0.81
                                                10000
    accuracy
                   0.72
                             0.76
                                       0.74
                                                10000
   macro avg
                             0.81
                                                10000
weighted avg
                   0.83
                                       0.82
[('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca',
PCA(n components=10)), ('model', RandomForestClassifier(class weight='balance
d_subsample', criterion='entropy',
                       min samples split=100, n estimators=1000,
                       random state=0))]
Selector: RFECV RandomForest Model: RandomForest
              precision
                          recall f1-score
                                              support
                                       0.88
           0
                   0.91
                             0.85
                                                 7963
           1
                   0.53
                             0.69
                                       0.60
                                                 2037
    accuracy
                                       0.81
                                                10000
                   0.72
                             0.77
                                       0.74
                                                10000
   macro avg
weighted avg
                   0.84
                             0.81
                                       0.82
                                                10000
[('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n_components=1
0)), ('model', RandomForestClassifier(class_weight='balanced_subsample', crit
erion='entropy',
                       min samples split=100, n estimators=1000,
                       random state=0))]
====>>>> Failure in RandomForest, FWE f classif
[('sm', SMOTE()), ('pca', PCA(n_components=10)), ('model', RandomForestClassi
fier(class_weight='balanced_subsample', criterion='entropy',
                       min samples split=100, n estimators=1000,
                       random state=0))]
Selector: NONE Model: RandomForest
              precision
                          recall f1-score
                                              support
           0
                   0.92
                             0.84
                                       0.88
                                                 7963
                             0.70
           1
                   0.53
                                       0.61
                                                 2037
                                       0.81
                                                10000
    accuracy
                   0.72
                             0.77
                                       0.74
                                                10000
   macro avg
weighted avg
                   0.84
                             0.81
                                       0.82
                                                10000
MODEL: GradientBoosting
                           FWE f classif
                                            0.00000
                                                       0.00000
                                    NONE
                                            0.78990
                                                       0.58552
```

FWE f\_classif 0.00000 0.00000 NONE 0.78990 0.58552 Variance Threshold 0.8 0.79250 0.58690 RFECV RandomForest 0.80290 0.59285 MODEL: RandomForest \_\_\_\_\_\_ FWE f\_classif 0.00000 0.00000 Variance Threshold 0.8 0.80930 0.59537 RFECV RandomForest 0.81420 0.60214 NONE 0.81330 0.60604 [('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.159999999999999)), ('pca', PCA(n\_components=10)), ('model', SVC(gamma='auto', random\_state=0))] Selector: Variance Threshold 0.8 Model: SVC precision recall f1-score support 0 0.92 0.81 0.86 7963 1 0.50 0.73 0.59 2037 accuracy 0.80 10000 macro avg 0.71 0.77 0.73 10000 0.83 0.80 0.81 10000 weighted avg [('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca', PCA(n\_components=10)), ('model', SVC(gamma='auto', random\_state=0))] Selector: RFECV RandomForest Model: SVC precision recall f1-score support 0 0.92 0.81 0.86 7963 1 0.50 0.72 0.59 2037 accuracy 0.80 10000 macro avg 0.71 0.77 0.73 10000 weighted avg 0.83 0.80 0.81 10000 [('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n\_components=1 0)), ('model', SVC(gamma='auto', random\_state=0))] ====>>>> Failure in SVC, FWE f classif [('sm', SMOTE()), ('pca', PCA(n\_components=10)), ('model', SVC(gamma='auto', random state=0))] Selector: NONE Model: SVC precision recall f1-score support 0 0.92 0.82 0.87 7963 1 0.50 0.73 0.59 2037 0.80 10000 accuracy 0.71 0.73 macro avg 0.77 10000

MODEL: GradientBoosting

0.84

0.80

-----

weighted avg

	FWE f_classif	0.00000	0.00000
	NONE	0.78990	0.58552
Variance	Threshold 0.8	0.79250	0.58690

0.81

10000

	RFECV RandomF	orest	0.80290	0.59285	
MODEL: RandomFo	prest				
	<del>-</del>		0.00000	0.00000	
	Variance Threshol RFECV RandomF				
	KI LEV Kandomi		0.81330		
MODEL: SVC					
		: C	0.00000	0 00000	
	RFECV RandomF		0.00000 a 7955a	0.00000 0.58960	
	Variance Threshol				
			0.79820		
==========	.=========		:======	=======	========
PCA 6					
<pre>[('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.15999999999999999)), ('pca', PCA(n_components=10)), ('model', GradientBoostingClassifier(min_sampl es_split=42, random_state=0))] Selector: Variance Threshold 0.8 Model: GradientBoosting</pre>					
1	precision recall	f1-score	support		
0	0.92 0.81	0.86	7963		
1	0.49 0.72	0.59			
accuracy	0.71 0.77	0.79	10000		
macro avg weighted avg	0.71 0.77 0.83 0.79	0.72 0.81	10000 10000		
8	0.00	0.02			
<pre>[('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca', PCA(n_components=10)), ('model', GradientBoostingClassifier(min_samples_split =42, random_state=0))]</pre>					
Selector: RFECV RandomForest Model: GradientBoosting					
1	precision recall	f1-score	support		
0	0.92 0.81	0.86	7963		
1	0.49 0.72	0.58	2037		
		<b>2</b> = 5	4000-		
accuracy	0.70 0.70	0.79	10000		
macro avg weighted avg	0.70 0.76 0.83 0.79	0.72 0.80	10000 10000		
weighten avg	0.05	0.00	10000		
<pre>[('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n_components=1 0)), ('model', GradientBoostingClassifier(min_samples_split=42, random_state= 0))]</pre>					

====>>>> Failure in GradientBoosting, FWE f\_classif

```
[('sm', SMOTE()), ('pca', PCA(n_components=10)), ('model', GradientBoostingCl
assifier(min_samples_split=42, random_state=0))]
Selector: NONE Model: GradientBoosting
              precision
                           recall f1-score
                                              support
           0
                   0.92
                             0.81
                                       0.86
                                                 7963
           1
                   0.49
                             0.72
                                       0.58
                                                 2037
                                       0.79
    accuracy
                                                10000
                                       0.72
                                                10000
   macro avg
                   0.71
                             0.76
                             0.79
                                       0.80
weighted avg
                   0.83
                                                10000
MODEL: GradientBoosting
                           FWE f classif
                                            0.00000
                                                       0.00000
                      RFECV RandomForest
                                            0.79120
                                                       0.58390
                                    NONE
                                            0.79150
                                                       0.58458
                  Variance Threshold 0.8
                                            0.79280
                                                       0.58626
[('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.159999999999999)),
('pca', PCA(n_components=10)), ('model', RandomForestClassifier(class_weight
='balanced_subsample', criterion='entropy',
                       min_samples_split=100, n_estimators=1000,
                       random state=0))]
Selector: Variance Threshold 0.8 Model: RandomForest
              precision recall f1-score
                                              support
           0
                   0.91
                             0.84
                                       0.88
                                                 7963
           1
                   0.53
                             0.69
                                       0.60
                                                 2037
                                       0.81
                                                10000
    accuracy
   macro avg
                   0.72
                             0.77
                                       0.74
                                                10000
                   0.83
                             0.81
                                       0.82
                                                10000
weighted avg
[('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca',
PCA(n components=10)), ('model', RandomForestClassifier(class weight='balance
d_subsample', criterion='entropy',
                       min samples split=100, n estimators=1000,
                       random state=0))]
Selector: RFECV RandomForest Model: RandomForest
              precision
                        recall f1-score
                                              support
                   0.92
                             0.85
           0
                                       0.88
                                                 7963
           1
                   0.54
                             0.70
                                       0.61
                                                 2037
    accuracy
                                       0.82
                                                10000
   macro avg
                   0.73
                             0.77
                                       0.74
                                                10000
weighted avg
                   0.84
                             0.82
                                       0.82
                                                10000
[('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n_components=1
0)), ('model', RandomForestClassifier(class_weight='balanced_subsample', crit
```

min\_samples\_split=100, n\_estimators=1000,
random state=0))]

erion='entropy',

```
====>>>> Failure in RandomForest, FWE f classif
[('sm', SMOTE()), ('pca', PCA(n_components=10)), ('model', RandomForestClassi
fier(class weight='balanced subsample', criterion='entropy',
                      min samples split=100, n estimators=1000,
                      random state=0))]
Selector: NONE Model: RandomForest
             precision
                         recall f1-score
                                            support
          0
                  0.91
                           0.84
                                     0.88
                                               7963
          1
                  0.53
                           0.69
                                               2037
                                     0.60
                                     0.81
   accuracy
                                              10000
                  0.72
                           0.77
                                     0.74
                                              10000
   macro avg
                           0.81
weighted avg
                  0.84
                                     0.82
                                              10000
MODEL: GradientBoosting
______
                          FWE f classif
                                          0.00000
                                                    0.00000
                     RFECV RandomForest
                                          0.79120
                                                    0.58390
                                          0.79150
                                  NONE
                                                    0.58458
                 Variance Threshold 0.8
                                          0.79280
                                                     0.58626
MODEL: RandomForest
                                                     0.00000
                          FWE f_classif
                                          0.00000
                 Variance Threshold 0.8
                                          0.80980
                                                    0.59669
                                  NONE
                                                     0.59944
                                          0.81250
                     RFECV RandomForest
                                          0.81510
                                                     0.60601
[('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.159999999999999)),
('pca', PCA(n_components=10)), ('model', SVC(gamma='auto', random_state=0))]
Selector: Variance Threshold 0.8 Model: SVC
             precision recall f1-score
                                            support
          0
                  0.92
                           0.81
                                     0.86
                                               7963
          1
                  0.50
                           0.72
                                     0.59
                                               2037
   accuracy
                                     0.79
                                              10000
   macro avg
                  0.71
                           0.77
                                     0.73
                                              10000
weighted avg
                           0.79
                                              10000
                  0.83
                                     0.81
[('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca',
PCA(n_components=10)), ('model', SVC(gamma='auto', random_state=0))]
Selector: RFECV RandomForest Model: SVC
             precision
                       recall f1-score
                                            support
          0
                  0.92
                           0.82
                                     0.86
                                               7963
                  0.50
          1
                           0.72
                                     0.59
                                               2037
                                     0.80
   accuracy
                                              10000
                  0.71
                           0.77
                                     0.73
                                              10000
   macro avg
```

weighted avg

0.83

0.80

0.81

10000

```
[('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n_components=1
0)), ('model', SVC(gamma='auto', random_state=0))]
Selector: FWE f_classif Model: SVC
           precision
                    recall f1-score
                                      support
               0.92
         0
                        0.82
                                0.87
                                        7963
         1
               0.51
                        0.74
                                0.60
                                        2037
                                0.80
                                        10000
   accuracy
               0.71
0.84
                                0.73
  macro avg
                        0.78
                                        10000
weighted avg
                        0.80
                                        10000
                                0.81
[('sm', SMOTE()), ('pca', PCA(n_components=10)), ('model', SVC(gamma='auto',
random_state=0))]
Selector: NONE Model: SVC
           precision recall f1-score
                                      support
         0
               0.92
                        0.81
                                0.86
                                        7963
         1
               0.50
                        0.72
                                0.59
                                        2037
   accuracy
                                0.80
                                        10000
               0.71
0.83
                                0.73
                        0.77
                                        10000
  macro avg
                        0.80
weighted avg
                                0.81
                                        10000
MODEL: GradientBoosting
                      FWE f classif
                                    0.00000 0.00000
                  RFECV RandomForest
                                    0.79120 0.58390
                              NONE
                                    0.79150
                                             0.58458
               Variance Threshold 0.8
                                    0.79280
                                             0.58626
MODEL: RandomForest
                      FWE f_classif 0.00000 0.00000
               Variance Threshold 0.8
                                    0.80980 0.59669
                              NONE 0.81250 0.59944
                  RFECV RandomForest
                                    0.81510
                                             0.60601
MODEL: SVC
______
               Variance Threshold 0.8 0.79450 0.58957
                              NONE
                                    0.79550
                                             0.58960
                  RFECV RandomForest
                                    0.79670
                                             0.59037
                      FWE f_classif
                                    0.79990
                                             0.60004
______
______
PCA 7
```

[('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.159999999999999)), ('pca', PCA(n\_components=10)), ('model', GradientBoostingClassifier(min\_sampl es split=42, random state=0))] Selector: Variance Threshold 0.8 Model: GradientBoosting precision recall f1-score support 0.81 0.92 0.86 0 7963 1 0.49 0.73 0.59 2037 0.79 accuracy 10000 macro avg 0.71 0.77 0.72 10000 weighted avg 0.83 0.79 0.81 10000 [('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca', PCA(n\_components=10)), ('model', GradientBoostingClassifier(min\_samples\_split =42, random state=0))] Selector: RFECV RandomForest Model: GradientBoosting precision recall f1-score support 0 0.91 0.83 0.87 7963 1 0.51 0.70 0.59 2037 0.80 10000 accuracy macro avg 0.71 0.76 0.73 10000 weighted avg 0.83 0.80 0.81 10000 [('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n\_components=1 0)), ('model', GradientBoostingClassifier(min samples split=42, random state= 0))] Selector: FWE f\_classif Model: GradientBoosting precision recall f1-score support 0 0.92 0.82 0.87 7963 1 0.51 0.73 0.60 2037 0.80 accuracy 10000 macro avg 0.72 0.78 0.74 10000 weighted avg 0.84 0.80 0.81 10000 [('sm', SMOTE()), ('pca', PCA(n components=10)), ('model', GradientBoostingCl Selector: NONE Model: GradientBoosting

assifier(min samples split=42, random state=0))]

	precision	recall	f1-score	support
0	0.92	0.81	0.86	7963
1	0.49	0.72	0.58	2037
accuracy			0.79	10000
macro avg	0.70	0.76	0.72	10000
weighted avg	0.83	0.79	0.80	10000

MODEL: GradientBoosting

NONE 0.78970 0.58299 Variance Threshold 0.8 0.79190 0.58784 FWE f\_classif 0.80290 0.60142

```
[('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.159999999999999)),
('pca', PCA(n_components=10)), ('model', RandomForestClassifier(class_weight
='balanced_subsample', criterion='entropy',
```

min\_samples\_split=100, n\_estimators=1000, random state=0))]

Selector: Variance Threshold 0.8 Model: RandomForest

	precision	recall	f1-score	support
0	0.92	0.84	0.88	7963
1	0.53	0.70	0.60	2037
accuracy			0.81	10000
macro avg	0.72	0.77	0.74	10000
weighted avg	0.84	0.81	0.82	10000

[('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca', PCA(n\_components=10)), ('model', RandomForestClassifier(class\_weight='balance d\_subsample', criterion='entropy',

> min\_samples\_split=100, n\_estimators=1000, random\_state=0))]

Selector: RFECV RandomForest Model: RandomForest

	precision	recall	f1-score	support
0	0.92	0.84	0.88	7963
1	0.53	0.70	0.61	2037
accuracy			0.81	10000
macro avg	0.73	0.77	0.74	10000
weighted avg	0.84	0.81	0.82	10000

[('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n\_components=1 0)), ('model', RandomForestClassifier(class\_weight='balanced\_subsample', crit erion='entropy',

> min samples split=100, n estimators=1000, random state=0))]

====>>>> Failure in RandomForest, FWE f classif

[('sm', SMOTE()), ('pca', PCA(n\_components=10)), ('model', RandomForestClassi fier(class\_weight='balanced\_subsample', criterion='entropy',

> min samples split=100, n estimators=1000, random state=0))]

Selector: NONE Model: RandomForest

	precision	recall	f1-score	support
0	0.92	0.84	0.88	7963
1	0.53	0.70	0.60	2037
accuracy			0.81	10000
macro avg weighted avg	0.72 0.84	0.77 0.81	0.74 0.82	10000 10000

MODEL: GradientBoosting

-----NONE 0.78970 0.58299 Variance Threshold 0.8 0.79190 0.58784 RFECV RandomForest 0.80210 0.58984 FWE f\_classif 0.80290 0.60142 MODEL: RandomForest 0.00000 FWE f classif 0.00000 Variance Threshold 0.8 0.81050 0.59979 NONE 0.81160 0.60387 RFECV RandomForest 0.81470 0.60600 [('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.159999999999999)), ('pca', PCA(n\_components=10)), ('model', SVC(gamma='auto', random\_state=0))] Selector: Variance Threshold 0.8 Model: SVC precision recall f1-score support 0 0.92 0.82 0.87 7963 1 0.50 0.73 0.60 2037 0.80 10000 accuracy macro avg 0.71 0.77 0.73 10000 weighted avg 0.84 0.80 0.81 10000 [('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca', PCA(n\_components=10)), ('model', SVC(gamma='auto', random\_state=0))] Selector: RFECV RandomForest Model: SVC recall f1-score precision support 0 0.92 0.81 0.86 7963 1 0.50 0.73 0.59 2037 0.80 10000 accuracy macro avg 0.71 0.77 0.73 10000 weighted avg 0.84 0.80 0.81 10000 [('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n\_components=1 0)), ('model', SVC(gamma='auto', random\_state=0))] ====>>>> Failure in SVC, FWE f classif [('sm', SMOTE()), ('pca', PCA(n\_components=10)), ('model', SVC(gamma='auto', random state=0))] Selector: NONE Model: SVC 3

	precision	recall	f1-score	support
0 1	0.92 0.50	0.82 0.72	0.87 0.59	7963 2037
accuracy macro avg weighted avg	0.71 0.84	0.77 0.80	0.80 0.73 0.81	10000 10000 10000

```
MODEL: GradientBoosting
                                 NONE
                                        0.78970
                                                  0.58299
                Variance Threshold 0.8
                                        0.79190
                                                  0.58784
                    RFECV RandomForest
                                        0.80210 0.58984
                        FWE f_classif 0.80290
                                                  0.60142
MODEL: RandomForest
______
                         FWE f_classif 0.00000 0.00000
                Variance Threshold 0.8
                                        0.81050 0.59979
                                 NONE
                                        0.81160
                                                  0.60387
                    RFECV RandomForest
                                        0.81470 0.60600
MODEL: SVC
                         FWE f_classif 0.00000 0.00000
                                 NONE
                                        0.79810 0.59384
                    RFECV RandomForest
                                        0.79690
                                                  0.59404
                                        0.79770
                Variance Threshold 0.8
                                                  0.59516
PCA 8
[('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.1599999999999999)),
('pca', PCA(n_components=10)), ('model', GradientBoostingClassifier(min_sampl
es split=42, random state=0))]
Selector: Variance Threshold 0.8 Model: GradientBoosting
            precision recall f1-score
                                         support
                 0.92
                                   0.86
          0
                          0.82
                                             7963
                 0.50
          1
                          0.73
                                   0.59
                                             2037
   accuracy
                                   0.80
                                            10000
                 0.71
0.84
  macro avg
                          0.77
                                   0.73
                                            10000
                          0.80
weighted avg
                                   0.81
                                            10000
[('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca',
PCA(n_components=10)), ('model', GradientBoostingClassifier(min_samples_split
=42, random state=0))]
Selector: RFECV RandomForest Model: GradientBoosting
             precision recall f1-score
                                         support
                 0.92
          0
                          0.82
                                   0.87
                                             7963
                 0.50
          1
                          0.70
                                   0.59
                                             2037
                                   0.80
   accuracy
                                            10000
                 0.71
                          0.76
                                   0.73
                                            10000
  macro avg
```

weighted avg

0.83

0.80

0.81

```
[('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n_components=1
0)), ('model', GradientBoostingClassifier(min samples split=42, random state=
0))]
====>>>> Failure in GradientBoosting, FWE f classif
[('sm', SMOTE()), ('pca', PCA(n_components=10)), ('model', GradientBoostingCl
assifier(min samples split=42, random state=0))]
Selector: NONE Model: GradientBoosting
             precision
                        recall f1-score
                                            support
                  0.92
0.49
          0
                            0.81
                                     0.86
                                               7963
                                     0.59
          1
                            0.72
                                               2037
                                     0.79
                                              10000
    accuracy
                  0.710.83
                           0.77
0.79
   macro avg
                                     0.72
                                              10000
weighted avg
                                     0.80
                                              10000
MODEL: GradientBoosting
                          FWE f_classif 0.00000 0.00000
                                  NONE
                                          0.79170 0.58564
                     RFECV RandomForest
                                          0.79900
                                                     0.58778
                 Variance Threshold 0.8
                                          0.79700
                                                     0.59302
[('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.159999999999999)),
('pca', PCA(n components=10)), ('model', RandomForestClassifier(class weight
='balanced_subsample', criterion='entropy',
                      min_samples_split=100, n_estimators=1000,
                      random state=0))]
Selector: Variance Threshold 0.8 Model: RandomForest
             precision recall f1-score support
                            0.84
                                     0.88
          0
                  0.92
                                               7963
          1
                  0.53
                            0.70
                                     0.61
                                               2037
   accuracy
                                     0.81
                                              10000
                  0.72
0.84
   macro avg
                            0.77
                                     0.74
                                              10000
weighted avg
                  0.84
                            0.81
                                     0.82
                                              10000
[('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca',
PCA(n components=10)), ('model', RandomForestClassifier(class weight='balance
d_subsample', criterion='entropy',
                      min samples split=100, n estimators=1000,
                      random state=0))]
Selector: RFECV RandomForest Model: RandomForest
             precision recall f1-score
                                           support
          0
                  0.92
                            0.84
                                     0.88
                                               7963
          1
                  0.53
                            0.70
                                     0.60
                                               2037
                                     0.81
                                              10000
    accuracy
                  0.72
                                     0.74
   macro avg
                            0.77
                                              10000
```

weighted avg

0.84

0.81

0.82

```
[('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n components=1
('model', RandomForestClassifier(class_weight='balanced_subsample', crit
erion='entropy',
                      min samples split=100, n estimators=1000,
                      random state=0))]
====>>>> Failure in RandomForest, FWE f_classif
[('sm', SMOTE()), ('pca', PCA(n_components=10)), ('model', RandomForestClassi
fier(class_weight='balanced_subsample', criterion='entropy',
                      min_samples_split=100, n_estimators=1000,
                      random state=0))]
Selector: NONE Model: RandomForest
             precision
                          recall f1-score
                                             support
          0
                  0.92
                            0.84
                                      0.88
                                                7963
          1
                  0.53
                            0.70
                                      0.60
                                                2037
                                      0.81
                                               10000
   accuracy
                  0.72
                            0.77
                                      0.74
                                               10000
   macro avg
weighted avg
                  0.84
                            0.81
                                      0.82
                                               10000
MODEL: GradientBoosting
______
                          FWE f_classif
                                           0.00000
                                                      0.00000
                                   NONE
                                           0.79170
                                                      0.58564
                     RFECV RandomForest
                                           0.79900
                                                      0.58778
                 Variance Threshold 0.8
                                           0.79700
                                                      0.59302
MODEL: RandomForest
                          FWE f_classif
                                           0.00000
                                                      0.00000
                     RFECV RandomForest
                                           0.81230
                                                      0.60140
                                   NONE
                                           0.81250
                                                      0.60435
                 Variance Threshold 0.8
                                           0.81320
                                                      0.60574
[('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.159999999999999)),
('pca', PCA(n components=10)), ('model', SVC(gamma='auto', random state=0))]
Selector: Variance Threshold 0.8 Model: SVC
             precision recall f1-score
                                             support
                  0.92
          0
                            0.82
                                      0.87
                                                7963
          1
                  0.50
                                      0.59
                                                2037
                            0.73
                                      0.80
                                               10000
   accuracy
   macro avg
                  0.71
                            0.77
                                      0.73
                                               10000
weighted avg
                  0.84
                            0.80
                                      0.81
                                               10000
[('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca',
PCA(n_components=10)), ('model', SVC(gamma='auto', random_state=0))]
Selector: RFECV RandomForest Model: SVC
             precision
                          recall f1-score
                                             support
          0
                  0.92
                            0.81
                                      0.86
                                                7963
```

```
1
                0.50
                         0.73
                                  0.59
                                           2037
                                  0.79
                                          10000
   accuracy
                0.71
                         0.77
                                  0.73
                                          10000
  macro avg
weighted avg
                0.83
                         0.79
                                  0.81
                                          10000
[('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n_components=1
0)), ('model', SVC(gamma='auto', random_state=0))]
====>>>> Failure in SVC, FWE f_classif
[('sm', SMOTE()), ('pca', PCA(n components=10)), ('model', SVC(gamma='auto',
random state=0))]
Selector: NONE Model: SVC
            precision
                      recall f1-score
                                        support
                0.92
                         0.82
                                  0.87
                                           7963
         1
                0.50
                         0.73
                                  0.59
                                           2037
                                  0.80
                                          10000
   accuracy
                0.71
                         0.77
  macro avg
                                  0.73
                                          10000
weighted avg
                0.84
                         0.80
                                  0.81
                                          10000
MODEL: GradientBoosting
______
                       FWE f_classif
                                      0.00000
                                                0.00000
                               NONE
                                      0.79170
                                                0.58564
                   RFECV RandomForest
                                      0.79900
                                                0.58778
               Variance Threshold 0.8
                                      0.79700
                                                0.59302
MODEL: RandomForest
                       FWE f classif
                                      0.00000
                                                0.00000
                   RFECV RandomForest
                                      0.81230 0.60140
                               NONE
                                      0.81250
                                                0.60435
               Variance Threshold 0.8
                                      0.81320
                                                0.60574
MODEL: SVC
-----
                       FWE f_classif
                                      0.00000
                                                0.00000
                   RFECV RandomForest
                                      0.79430
                                                0.59081
                               NONE
                                      0.79770
                                                0.59369
               Variance Threshold 0.8
                                      0.79810
                                                0.59482
[('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.159999999999999)),
('pca', PCA(n_components=10)), ('model', GradientBoostingClassifier(min_sampl
es split=42, random state=0))]
```

```
Selector: Variance Threshold 0.8 Model: GradientBoosting
              precision
                            recall f1-score
                                               support
           0
                   0.92
                              0.81
                                        0.86
                                                  7963
           1
                   0.50
                              0.73
                                        0.59
                                                  2037
    accuracy
                                        0.80
                                                 10000
                   0.71
                                        0.73
                                                 10000
   macro avg
                              0.77
weighted avg
                   0.84
                              0.80
                                        0.81
                                                 10000
[('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca',
PCA(n_components=10)), ('model', GradientBoostingClassifier(min_samples_split
=42, random state=0))]
Selector: RFECV RandomForest Model: GradientBoosting
              precision
                            recall f1-score
                                               support
           0
                   0.92
                              0.81
                                        0.86
                                                  7963
           1
                   0.49
                              0.71
                                        0.58
                                                  2037
                                        0.79
                                                 10000
    accuracy
   macro avg
                   0.70
                              0.76
                                        0.72
                                                 10000
weighted avg
                   0.83
                              0.79
                                        0.80
                                                 10000
[('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n_components=1
0)), ('model', GradientBoostingClassifier(min_samples_split=42, random_state=
0))]
Selector: FWE f classif Model: GradientBoosting
                            recall f1-score
              precision
                                               support
                   0.92
                              0.82
                                        0.87
                                                  7963
           0
           1
                   0.51
                              0.73
                                        0.60
                                                  2037
    accuracy
                                        0.80
                                                 10000
   macro avg
                   0.71
                                        0.73
                                                 10000
                              0.78
weighted avg
                   0.84
                              0.80
                                        0.81
                                                 10000
[('sm', SMOTE()), ('pca', PCA(n_components=10)), ('model', GradientBoostingCl
assifier(min samples split=42, random state=0))]
Selector: NONE Model: GradientBoosting
              precision
                            recall f1-score
                                               support
           0
                   0.92
                              0.80
                                        0.86
                                                  7963
           1
                   0.48
                              0.72
                                        0.58
                                                  2037
                                        0.79
    accuracy
                                                 10000
                   0.70
                              0.76
                                        0.72
                                                 10000
   macro avg
                              0.79
weighted avg
                   0.83
                                        0.80
                                                 10000
MODEL: GradientBoosting
                                     NONE
                                             0.78660
                                                         0.57809
                       RFECV RandomForest
                                             0.79170
                                                         0.58181
                  Variance Threshold 0.8
                                             0.79690
                                                         0.59388
                            FWE f_classif
                                             0.80050
                                                         0.59932
```

- -

```
[('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.159999999999999)),
('pca', PCA(n_components=10)), ('model', RandomForestClassifier(class_weight
='balanced_subsample', criterion='entropy',
                       min samples split=100, n estimators=1000,
                       random state=0))]
Selector: Variance Threshold 0.8 Model: RandomForest
              precision
                           recall f1-score
                                              support
           0
                   0.92
                             0.84
                                       0.88
                                                  7963
           1
                   0.53
                             0.70
                                       0.60
                                                  2037
                                       0.81
                                                 10000
    accuracy
                   0.72
                             0.77
                                       0.74
                                                 10000
   macro avg
weighted avg
                   0.84
                             0.81
                                       0.82
                                                 10000
[('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca',
PCA(n_components=10)), ('model', RandomForestClassifier(class_weight='balance
d subsample', criterion='entropy',
                       min samples split=100, n estimators=1000,
                       random state=0))]
Selector: RFECV RandomForest Model: RandomForest
                           recall f1-score
              precision
                                              support
                                       0.88
           0
                   0.91
                             0.85
                                                  7963
           1
                   0.53
                             0.69
                                       0.60
                                                  2037
                                       0.81
                                                 10000
    accuracy
                   0.72
                                       0.74
                                                 10000
   macro avg
                             0.77
weighted avg
                   0.84
                             0.81
                                       0.82
                                                 10000
[('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n_components=1
0)), ('model', RandomForestClassifier(class_weight='balanced_subsample', crit
erion='entropy',
                       min samples split=100, n estimators=1000,
                       random state=0))]
Selector: FWE f_classif Model: RandomForest
              precision
                           recall f1-score
                                              support
           0
                   0.92
                             0.84
                                       0.88
                                                  7963
                   0.54
           1
                             0.72
                                                  2037
                                       0.62
    accuracy
                                       0.82
                                                 10000
   macro avg
                   0.73
                             0.78
                                       0.75
                                                 10000
weighted avg
                                       0.83
                   0.84
                             0.82
                                                 10000
[('sm', SMOTE()), ('pca', PCA(n_components=10)), ('model', RandomForestClassi
fier(class_weight='balanced_subsample', criterion='entropy',
                       min_samples_split=100, n_estimators=1000,
                       random state=0))]
Selector: NONE Model: RandomForest
              precision
                           recall f1-score
                                              support
           0
                   0.92
                             0.84
                                       0.88
                                                  7963
                   0.53
                             0.70
           1
                                       0.60
                                                  2037
```

0.81

10000

accuracy

```
0.72
                           0.77
                                    0.74
                                            10000
  macro avg
weighted avg
                 0.84
                           0.81
                                    0.82
                                            10000
MODEL: GradientBoosting
______
                                 NONE
                                         0.78660 0.57809
                    RFECV RandomForest
                                         0.79170
                                                   0.58181
                Variance Threshold 0.8
                                         0.79690
                                                   0.59388
                         FWE f classif
                                         0.80050
                                                   0.59932
------
MODEL: RandomForest
                                 NONE
                                        0.81140
                                                   0.60127
                    RFECV RandomForest
                                         0.81440
                                                   0.60137
                Variance Threshold 0.8
                                         0.81280
                                                   0.60289
                         FWE f_classif
                                         0.81890
                                                   0.61769
[('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.159999999999999)),
('pca', PCA(n_components=10)), ('model', SVC(gamma='auto', random_state=0))]
Selector: Variance Threshold 0.8 Model: SVC
             precision
                       recall f1-score
                                          support
          0
                 0.92
                           0.81
                                    0.86
                                             7963
                 0.50
                           0.72
                                    0.59
                                             2037
          1
                                    0.79
                                            10000
   accuracy
   macro avg
                 0.71
                          0.77
                                    0.73
                                            10000
weighted avg
                 0.83
                           0.79
                                    0.81
                                            10000
[('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca',
PCA(n_components=10)), ('model', SVC(gamma='auto', random_state=0))]
Selector: RFECV RandomForest Model: SVC
             precision
                        recall f1-score
                                          support
          0
                 0.92
                           0.81
                                    0.86
                                             7963
          1
                 0.50
                           0.72
                                    0.59
                                             2037
   accuracy
                                    0.79
                                            10000
  macro avg
                 0.71
                          0.77
                                    0.73
                                            10000
                           0.79
                                    0.81
weighted avg
                 0.83
                                            10000
[('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n_components=1
0)), ('model', SVC(gamma='auto', random_state=0))]
====>>>> Failure in SVC, FWE f_classif
[('sm', SMOTE()), ('pca', PCA(n_components=10)), ('model', SVC(gamma='auto',
random state=0))]
Selector: NONE Model: SVC
             precision
                         recall f1-score
                                          support
                                             7963
          0
                 0.92
                          0.82
                                    0.87
          1
                 0.50
                           0.72
                                    0.59
                                             2037
```

accuracy macro avg	0.71	0.77	0.80 0.73	10000 10000				
weighted avg	0.84	0.80	0.81	10000				
MODEL: GradientBoosting								
			NONE	0.78660	0.57809			
		V RandomFor Threshold		0.79170	0.58181 0.59388			
	vai Talice			0.80050				
MODEL: RandomFor	rest							
		ı	NONE	0.81140	0.60127			
	RFEC	V RandomFo			0.60137			
	Variance	Threshold		0.81280	0.60289			
		FWE +_clas	ss1† 	0.81890	0.61769			
MODEL: SVC								
		FWE f_clas	ssif	0.00000	0.00000			
		Threshold		0.79380	0.58842			
	RFEC	V RandomFo						
			NONE	0.79860 	0.59346			
					=======================================			
PCA 10		=======	=====					
					0.15999999999999999)),			
			el', Gra	adientBoost	<pre>ingClassifier(min_sampl</pre>			
es_split=42, rar Selector: Variar			ol. Gnad	dion+Roosti	nα			
		recall fi			iig			
r								
0	0.92	0.82	0.86					
1	0.50	0.72	0.59	2037				
accuracy			0.80	10000				
macro avg	0.71	0.77	0.73					
weighted avg	0.83	0.80	0.81	10000				
PCA(n_components	<pre>[('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca', PCA(n_components=10)), ('model', GradientBoostingClassifier(min_samples_split</pre>							
=42, random_stat Selector: RFECV	RandomFore	st Model: ( recall fi		_				
рі	CCTOTOII	rccail I.	- 3COLE	σαμμοί τ				
0	0.92	0.83	0.87	7963				

```
0.51
                             0.70
                                       0.59
                                                 2037
           1
                                       0.80
                                                10000
    accuracy
                   0.71
                                       0.73
   macro avg
                             0.76
                                                10000
weighted avg
                   0.83
                             0.80
                                       0.81
                                                10000
[('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n_components=1
0)), ('model', GradientBoostingClassifier(min_samples_split=42, random_state=
0))]
====>>>> Failure in GradientBoosting, FWE f classif
[('sm', SMOTE()), ('pca', PCA(n_components=10)), ('model', GradientBoostingCl
assifier(min_samples_split=42, random_state=0))]
Selector: NONE Model: GradientBoosting
              precision
                           recall f1-score
                                              support
           0
                   0.92
                             0.81
                                       0.86
                                                 7963
           1
                   0.50
                             0.72
                                       0.59
                                                 2037
                                       0.80
                                                10000
    accuracy
                   0.71
                             0.77
                                       0.73
                                                10000
   macro avg
weighted avg
                   0.83
                             0.80
                                       0.81
                                                10000
MODEL: GradientBoosting
                           FWE f classif
                                                       0.00000
                                            0.00000
                      RFECV RandomForest
                                            0.80030
                                                       0.58867
                                    NONE
                                            0.79500
                                                       0.58918
                  Variance Threshold 0.8
                                            0.79650
                                                       0.58947
[('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.159999999999999)),
('pca', PCA(n_components=10)), ('model', RandomForestClassifier(class_weight
='balanced_subsample', criterion='entropy',
                       min_samples_split=100, n_estimators=1000,
                       random state=0))]
Selector: Variance Threshold 0.8 Model: RandomForest
                         recall f1-score
              precision
                                              support
                   0.92
                             0.84
                                       0.88
                                                 7963
           0
           1
                   0.53
                             0.70
                                       0.60
                                                 2037
                                       0.81
                                                10000
    accuracy
                   0.72
   macro avg
                             0.77
                                       0.74
                                                10000
weighted avg
                   0.84
                             0.81
                                       0.82
                                                10000
[('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca',
PCA(n_components=10)), ('model', RandomForestClassifier(class_weight='balance
d_subsample', criterion='entropy',
                       min_samples_split=100, n_estimators=1000,
                       random state=0))]
Selector: RFECV RandomForest Model: RandomForest
              precision
                           recall f1-score
                                              support
           0
                   0.91
                             0.84
                                       0.88
                                                 7963
           1
                   0.53
                             0.69
                                       0.60
                                                 2037
```

```
accuracy
                                     0.81
                                              10000
                  0.72
                            0.77
                                     0.74
                                              10000
   macro avg
weighted avg
                  0.84
                            0.81
                                     0.82
                                              10000
[('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n_components=1
0)), ('model', RandomForestClassifier(class weight='balanced subsample', crit
erion='entropy',
                      min_samples_split=100, n_estimators=1000,
                      random state=0))]
Selector: FWE f classif Model: RandomForest
                         recall f1-score
             precision
                                            support
                  0.92
          0
                           0.85
                                     0.88
                                               7963
          1
                  0.54
                            0.72
                                     0.62
                                               2037
                                     0.82
                                              10000
   accuracy
   macro avg
                  0.73
                            0.78
                                     0.75
                                              10000
weighted avg
                  0.84
                            0.82
                                     0.83
                                              10000
[('sm', SMOTE()), ('pca', PCA(n_components=10)), ('model', RandomForestClassi
fier(class weight='balanced subsample', criterion='entropy',
                      min samples split=100, n estimators=1000,
                      random state=0))]
Selector: NONE Model: RandomForest
             precision
                        recall f1-score
                                            support
          0
                  0.92
                           0.84
                                     0.88
                                               7963
          1
                  0.53
                            0.70
                                               2037
                                     0.61
                                     0.81
                                              10000
   accuracy
                  0.73 0.77
  macro avg
                                     0.74
                                              10000
weighted avg
                  0.84
                            0.81
                                     0.82
                                              10000
MODEL: GradientBoosting
                          FWE f classif
                                          0.00000
                                                     0.00000
                     RFECV RandomForest
                                          0.80030
                                                     0.58867
                                  NONE
                                          0.79500
                                                     0.58918
                 Variance Threshold 0.8
                                          0.79650
                                                     0.58947
MODEL: RandomForest
                     RFECV RandomForest
                                          0.81080
                                                     0.59881
                 Variance Threshold 0.8
                                          0.81170
                                                     0.60165
                                  NONE
                                          0.81460
                                                     0.60737
                          FWE f_classif
                                          0.81890
                                                     0.61704
[('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.159999999999999)),
('pca', PCA(n_components=10)), ('model', SVC(gamma='auto', random_state=0))]
Selector: Variance Threshold 0.8 Model: SVC
```

precision

recall f1-score

support

```
0
                   0.92
                             0.82
                                        0.86
                                                  7963
           1
                   0.50
                             0.72
                                        0.59
                                                  2037
    accuracy
                                        0.80
                                                 10000
                   0.71
                             0.77
                                        0.73
   macro avg
                                                 10000
weighted avg
                   0.83
                             0.80
                                        0.81
                                                 10000
[('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca',
PCA(n components=10)), ('model', SVC(gamma='auto', random state=0))]
Selector: RFECV RandomForest Model: SVC
              precision
                           recall f1-score
                                               support
           0
                   0.92
                             0.81
                                        0.86
                                                  7963
           1
                   0.49
                             0.73
                                        0.59
                                                  2037
                                        0.79
                                                 10000
    accuracy
   macro avg
                   0.71
                             0.77
                                        0.72
                                                 10000
weighted avg
                   0.83
                             0.79
                                        0.80
                                                 10000
[('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n_components=1
0)), ('model', SVC(gamma='auto', random state=0))]
Selector: FWE f_classif Model: SVC
                           recall f1-score
              precision
                                               support
           0
                   0.92
                             0.82
                                        0.87
                                                  7963
           1
                   0.51
                             0.74
                                        0.60
                                                  2037
                                        0.80
                                                 10000
    accuracy
                   0.72
                             0.78
                                        0.74
   macro avg
                                                 10000
                                        0.81
                                                 10000
weighted avg
                   0.84
                             0.80
[('sm', SMOTE()), ('pca', PCA(n_components=10)), ('model', SVC(gamma='auto',
random state=0))]
Selector: NONE Model: SVC
              precision
                         recall f1-score
                                               support
           0
                   0.92
                             0.82
                                        0.87
                                                  7963
                   0.50
                             0.73
           1
                                        0.60
                                                  2037
                                        0.80
                                                 10000
    accuracy
   macro avg
                   0.71
                             0.77
                                        0.73
                                                 10000
weighted avg
                   0.84
                             0.80
                                        0.81
                                                 10000
MODEL: GradientBoosting
                           FWE f classif
                                             0.00000
                                                        0.00000
                      RFECV RandomForest
                                             0.80030
                                                        0.58867
                                     NONE
                                             0.79500
                                                        0.58918
                  Variance Threshold 0.8
                                             0.79650
                                                        0.58947
MODEL: RandomForest
```

```
Variance Threshold 0.8
                                          0.81170
                                                    0.60165
                                  NONE
                                          0.81460
                                                    0.60737
                          FWE f_classif
                                          0.81890
                                                     0.61704
MODEL: SVC
                     RFECV RandomForest
                                          0.79120
                                                    0.58653
                 Variance Threshold 0.8
                                          0.79710
                                                    0.59232
                                  NONE
                                          0.79820
                                                    0.59510
                          FWE f_classif
                                          0.80200
                                                     0.60337
_____
PCA 11
[('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.1599999999999999)),
('pca', PCA(n components=10)), ('model', GradientBoostingClassifier(min sampl
es split=42, random state=0))]
Selector: Variance Threshold 0.8 Model: GradientBoosting
             precision
                       recall f1-score
                                            support
          0
                  0.92
                           0.81
                                     0.86
                                               7963
          1
                  0.50
                           0.72
                                     0.59
                                               2037
                                     0.80
    accuracy
                                              10000
   macro avg
                  0.71
                           0.77
                                     0.73
                                              10000
weighted avg
                  0.83
                           0.80
                                     0.81
                                              10000
[('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca',
PCA(n_components=10)), ('model', GradientBoostingClassifier(min_samples_split
=42, random state=0))]
Selector: RFECV RandomForest Model: GradientBoosting
             precision recall f1-score
                                            support
          0
                  0.92
                           0.82
                                     0.87
                                               7963
          1
                  0.51
                           0.71
                                     0.59
                                               2037
    accuracy
                                     0.80
                                              10000
   macro avg
                  0.71
                           0.77
                                     0.73
                                              10000
weighted avg
                  0.83
                           0.80
                                     0.81
                                              10000
[('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n_components=1
0)), ('model', GradientBoostingClassifier(min_samples_split=42, random_state=
0))]
Selector: FWE f_classif Model: GradientBoosting
             precision recall f1-score
                                            support
                  0.92
                           0.82
                                     0.87
          0
                                               7963
          1
                  0.51
                           0.73
                                     0.60
                                               2037
                                     0.80
                                              10000
    accuracy
   macro avg
                  0.72
                           0.77
                                     0.74
                                              10000
```

weighted avg 0.84 0.80 0.82 10000

[('sm', SMOTE()), ('pca', PCA(n\_components=10)), ('model', GradientBoostingCl
assifier(min\_samples\_split=42, random\_state=0))]

Selector: NONE Model: GradientBoosting

	precision	recall	f1-score	support
0	0.92	0.81	0.86	7963
1	0.49	0.72	0.59	2037
accuracy			0.79	10000
macro avg	0.71	0.77	0.72	10000
weighted avg	0.83	0.79	0.81	10000

## MODEL: GradientBoosting

-----

NONE	0.79300	0.58781
Variance Threshold 0.8	0.79520	0.58859
RFECV RandomForest	0.80020	0.59141
FWE f classif	0.80360	0.60114

-----

- -

[('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.15999999999999999)),
('pca', PCA(n\_components=10)), ('model', RandomForestClassifier(class\_weight
='balanced\_subsample', criterion='entropy',

min\_samples\_split=100, n\_estimators=1000,
random\_state=0))]

Selector: Variance Threshold 0.8 Model: RandomForest

		precision	recall	f1-score	support
	0	0.92	0.84	0.88	7963
	1	0.53	0.70	0.60	2037
accura	су			0.81	10000
macro a	vg	0.72	0.77	0.74	10000
weighted a	vg	0.84	0.81	0.82	10000

[('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca',
PCA(n\_components=10)), ('model', RandomForestClassifier(class\_weight='balance
d\_subsample', criterion='entropy',

min\_samples\_split=100, n\_estimators=1000,
random\_state=0))]

Selector: RFECV RandomForest Model: RandomForest

	precision	recall	f1-score	support
0	0.91	0.85	0.88	7963
1	0.53	0.67	0.59	2037
accuracy			0.81	10000
macro avg	0.72	0.76	0.74	10000
veighted avg	0.83	0.81	0.82	10000

[('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n\_components=1
0)), ('model', RandomForestClassifier(class\_weight='balanced\_subsample', crit
erion='entropy',

## min\_samples\_split=100, n\_estimators=1000, random\_state=0))]

Selector: FWE			≘=0))]		
	f_classif N	Model: Rand	domForest		
	precision	recall	f1-score	support	
0	0.92	0.84	0.88	7963	
1		0.71	0.61	2037	
_	0.54	0.71	0.01	2037	
			0 00	40000	
accuracy			0.82		
macro avg	0.73				
weighted avg	0.84	0.82	0.83	10000	
[('sm', SMOTE	()), ('pca',	PCA(n con	mponents=1	L0)), ('mod	del', RandomForestClass
fier(class_we					
110. (01035_N	-				itors=1000,
				, II_ESCIIII	1001 3-1000,
6 1 1 NON		andom_state	==0))]		
Selector: NON			•		
	precision	recall	f1-score	support	
0	0.92	0.84	0.88	7963	
1	0.53	0.70	0.60	2037	
accuracy			0.81	10000	
<del>-</del>	0.72	0 77			
macro avg	0.72				
weighted avg	0.84	0.81	0.82	10000	
MODEL: Gradie	ntBoosting				
			NONE	0.79300	0.58781
	Variano	e Thresho			0.58859
			ld 0.8	0.79520	
		CV Random	ld 0.8 Forest	0.79520 0.80020	0.59141
		CV Random	ld 0.8 Forest	0.79520	0.59141
		CV Random	ld 0.8 Forest	0.79520 0.80020	0.59141
		CV Random	ld 0.8 Forest	0.79520 0.80020	0.59141
		CV Random	ld 0.8 Forest	0.79520 0.80020	0.59141
	RFE	CV Random	ld 0.8 Forest	0.79520 0.80020	0.59141
MODEL: Random	RFE	CV Random	ld 0.8 Forest	0.79520 0.80020	0.59141
MODEL: Random	RFE	ECV RandomF FWE f_c	ld 0.8 Forest lassif	0.79520 0.80020 0.80360	0.59141 0.60114
MODEL: Random	RFE  Forest  RFE	ECV Random FWE f_c 	ld 0.8 Forest lassif	0.79520 0.80020 0.80360 	0.59141 0.60114 0.59288
MODEL: Random	RFE  Forest  RFE	ECV Random FWE f_c 	ld 0.8 Forest lassif  Forest ld 0.8	0.79520 0.80020 0.80360  0.81240 0.81210	0.59141 0.60114 0.59288 0.60400
MODEL: Random	RFE  Forest  RFE	ECV Random FWE f_c  ECV Random Ce Thresho	ld 0.8 Forest lassif  Forest ld 0.8 NONE	0.79520 0.80020 0.80360  0.81240 0.81210 0.81270	0.59141 0.60114 0.59288 0.60400 0.60460
MODEL: Random	RFE  Forest  RFE	ECV Random FWE f_c  ECV Random Ce Thresho	ld 0.8 Forest lassif  Forest ld 0.8 NONE	0.79520 0.80020 0.80360  0.81240 0.81210	0.59141 0.60114 0.59288 0.60400 0.60460
MODEL: Random	RFE  Forest  RFE	ECV Random FWE f_c  ECV Random Ce Thresho	ld 0.8 Forest lassif  Forest ld 0.8 NONE	0.79520 0.80020 0.80360  0.81240 0.81210 0.81270	0.59141 0.60114 0.59288 0.60400 0.60460
MODEL: Random	RFE  Forest  RFE	ECV Random FWE f_c  ECV Random Ce Thresho	ld 0.8 Forest lassif  Forest ld 0.8 NONE	0.79520 0.80020 0.80360  0.81240 0.81210 0.81270	0.59141 0.60114 0.59288 0.60400 0.60460
MODEL: Random	RFE  Forest  RFE	ECV Random FWE f_c  ECV Random Ce Thresho	ld 0.8 Forest lassif  Forest ld 0.8 NONE	0.79520 0.80020 0.80360  0.81240 0.81210 0.81270	0.59141 0.60114 0.59288 0.60400 0.60460
	RFE  Forest  RFE Variand	ECV Random FWE f_c ECV Random Ee Thresho FWE f_c	ld 0.8 Forest lassif ONE lassif	0.79520 0.80020 0.80360  0.81240 0.81210 0.81270 0.81740	0.59141 0.60114 0.59288 0.60400 0.60460 0.61379
	RFE	ECV Random FWE f_cl ECV Random Threshol FWE f_cl Variance	ld 0.8 Forest lassif Orest ld 0.8 NONE lassif	0.79520 0.80020 0.80360  0.81240 0.81210 0.81270 0.81740 	0.59141 0.60114 0.59288 0.60400 0.60460 0.61379
[('sm', SMOTE	RFE	ECV Random FWE f_cl ECV Random EE Threshol FWE f_cl Variance	Id 0.8 Forest Lassif Orest Id 0.8 NONE Lassif Oreshold(	0.79520 0.80020 0.80360  0.81240 0.81210 0.81270 0.81740 	0.59141 0.60114 0.59288 0.60400 0.60460 0.61379
	RFE  Forest  RFE  Variance  ()), ('sel',  _components= iance Thresh	ECV Random FWE f_cl ECV Random Threshol FWE f_cl Variance ( Variance)	Id 0.8 Forest lassif Orest Id 0.8 NONE lassif Odel', SVO	0.79520 0.80020 0.80360 0.81240 0.81210 0.81270 0.81740 0.81740 (threshold=	0.59141 0.60114 0.59288 0.60400 0.60460 0.61379
[('sm', SMOTE	RFE  Forest  RFE  Variance  ()), ('sel',  _components= iance Thresh	ECV Random FWE f_cl ECV Random EE Threshol FWE f_cl Variance	Id 0.8 Forest lassif Orest Id 0.8 NONE lassif Odel', SVO	0.79520 0.80020 0.80360  0.81240 0.81210 0.81270 0.81740 	0.59141 0.60114 0.59288 0.60400 0.60460 0.61379
[('sm', SMOTE ('pca', PCA(n Selector: Var	RFE  Forest  RFE  Variance  ()), ('sel',  _components= iance Thresh precision	ECV Random FWE f_cl ECV Random Threshol FWE f_cl Variance ( Variance ( Variance) ( Variance) ( Variance)	Id 0.8 Forest Lassif Orest Id 0.8 NONE Lassif Odel', SVO Ddel: SVC	0.79520 0.80020 0.80360 0.81240 0.81210 0.81270 0.81740 (threshold= C(gamma='au	0.59141 0.60114 0.59288 0.60400 0.60460 0.61379
[('sm', SMOTE	RFE  Forest  RFE  Variance  ()), ('sel',  _components= iance Thresh precision  0.92	ECV RandomF FWE f_cl ECV RandomF Threshol FWE f_cl Variance (10)), ('monold 0.8 Monold 0.8 Monorecall 0.82	Id 0.8 Forest lassif Forest Id 0.8 NONE lassif Foreshold( odel', SVC odel: SVC f1-score	0.79520 0.80020 0.80360 0.81240 0.81210 0.81270 0.81740 0.81740 (threshold=	0.59141 0.60114 0.59288 0.60400 0.60460 0.61379
[('sm', SMOTE ('pca', PCA(n Selector: Var	RFE  Forest  RFE  Variance  ()), ('sel',  _components= iance Thresh precision	ECV Random FWE f_cl ECV Random Threshol FWE f_cl Variance ( Variance ( Variance) ( Variance) ( Variance)	Id 0.8 Forest Lassif Orest Id 0.8 NONE Lassif Odel', SVO Ddel: SVC	0.79520 0.80020 0.80360 0.81240 0.81210 0.81270 0.81740 (threshold= C(gamma='au	0.59141 0.60114 0.59288 0.60400 0.60460 0.61379
[('sm', SMOTE ('pca', PCA(n Selector: Var	RFE  Forest  RFE  Variance  ()), ('sel',  _components= iance Thresh precision  0.92	ECV RandomF FWE f_cl ECV RandomF Threshol FWE f_cl Variance (10)), ('monold 0.8 Monold 0.8 Monorecall 0.82	Id 0.8 Forest lassif Forest Id 0.8 NONE lassif Foreshold( odel', SVC odel: SVC f1-score	0.79520 0.80020 0.80360 0.81240 0.81210 0.81270 0.81740 0.81740 0.81740 0.81740 0.81740 0.81740 0.81740 0.81740	0.59141 0.60114 0.59288 0.60400 0.60460 0.61379
[('sm', SMOTE ('pca', PCA(n Selector: Var	RFE  Forest  RFE  Variance  ()), ('sel',  _components= iance Thresh precision  0.92	ECV RandomF FWE f_cl ECV RandomF Threshol FWE f_cl Variance (10)), ('monold 0.8 Monold 0.8 Monorecall 0.82	Id 0.8 Forest lassif Forest Id 0.8 NONE lassif Odel', SVO odel: SVC f1-score 0.87 0.59	0.79520 0.80020 0.80360 0.81240 0.81210 0.81270 0.81740 (threshold= C(gamma='au support 7963 2037	0.59141 0.60114 0.59288 0.60400 0.60460 0.61379
[('sm', SMOTE ('pca', PCA(n Selector: Var	RFE   Forest   Variance  ()), ('sel',  _components= iance Thresh precision  0.92 0.50	ECV Random FWE f_cl ECV Random Threshol FWE f_cl Variance 10)), ('monold 0.8 Monold 0.8 Monold 0.8 Monold 0.8 Monold 0.82 0.72	Id 0.8 Forest lassif Forest Id 0.8 NONE lassif Odel', SVO pdel: SVC f1-score 0.87 0.59 0.80	0.79520 0.80020 0.80360 0.81240 0.81210 0.81270 0.81740 0.81740 (threshold= (gamma='au support 7963 2037 10000	0.59141 0.60114 0.59288 0.60400 0.60460 0.61379
[('sm', SMOTE ('pca', PCA(n Selector: Var	RFE  Forest  RFE  Variance  ()), ('sel',  _components= iance Thresh precision  0.92	ECV RandomF FWE f_cl ECV RandomF Threshol FWE f_cl Variance (10)), ('monold 0.8 Monold 0.8 Monorecall 0.82	Id 0.8 Forest lassif Forest Id 0.8 NONE lassif Odel', SVO odel: SVC f1-score 0.87 0.59	0.79520 0.80020 0.80360 0.81240 0.81210 0.81270 0.81740 (threshold= C(gamma='au support 7963 2037	0.59141 0.60114 0.59288 0.60400 0.60460 0.61379

```
[('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca',
PCA(n_components=10)), ('model', SVC(gamma='auto', random_state=0))]
Selector: RFECV RandomForest Model: SVC
             precision recall f1-score
                                            support
                                     0.86
                                               7963
                  0.92
          0
                           0.81
          1
                  0.50
                            0.73
                                     0.59
                                               2037
    accuracy
                                     0.80
                                              10000
   macro avg
                  0.71
                            0.77
                                     0.73
                                              10000
weighted avg
                  0.83
                                     0.81
                                              10000
                            0.80
[('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n_components=1
0)), ('model', SVC(gamma='auto', random_state=0))]
====>>>> Failure in SVC, FWE f_classif
[('sm', SMOTE()), ('pca', PCA(n_components=10)), ('model', SVC(gamma='auto',
random state=0))]
Selector: NONE Model: SVC
             precision
                         recall f1-score
                                            support
                  0.92
                            0.82
                                     0.87
                                               7963
          1
                  0.50
                            0.73
                                     0.59
                                               2037
    accuracy
                                     0.80
                                              10000
   macro avg
                  0.71
                            0.77
                                     0.73
                                              10000
                                              10000
weighted avg
                  0.84
                            0.80
                                     0.81
MODEL: GradientBoosting
 NONE
                                          0.79300
                                                     0.58781
                 Variance Threshold 0.8
                                          0.79520
                                                     0.58859
                     RFECV RandomForest
                                          0.80020
                                                     0.59141
                         FWE f_classif
                                          0.80360
                                                     0.60114
MODEL: RandomForest
                     RFECV RandomForest
                                          0.81240
                                                     0.59288
                 Variance Threshold 0.8
                                          0.81210
                                                     0.60400
                                  NONE
                                          0.81270
                                                     0.60460
                         FWE f_classif
                                          0.81740
                                                     0.61379
MODEL: SVC
------
                          FWE f_classif
                                          0.00000
                                                     0.00000
                     RFECV RandomForest
                                          0.79630
                                                     0.59187
                 Variance Threshold 0.8
                                          0.79810
                                                     0.59335
                                  NONE
                                          0.79780
                                                     0.59365
```

PCA 12 [('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.1599999999999999)), ('pca', PCA(n\_components=10)), ('model', GradientBoostingClassifier(min\_sampl es\_split=42, random\_state=0))] Selector: Variance Threshold 0.8 Model: GradientBoosting precision recall f1-score support 0.92 0.86 0 0.81 7963 0.50 1 0.73 0.59 2037 0.79 10000 accuracy 0.71 10000 macro avg 0.77 0.73 weighted avg 0.83 0.79 0.81 10000 [('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca', PCA(n\_components=10)), ('model', GradientBoostingClassifier(min\_samples\_split =42, random state=0))] Selector: RFECV RandomForest Model: GradientBoosting precision recall f1-score support 0 0.92 0.83 0.87 7963 1 0.51 0.71 0.59 2037 0.80 10000 accuracy macro avg 0.71 0.77 0.73 10000 0.81 10000 weighted avg 0.83 0.80 [('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n\_components=1 0)), ('model', GradientBoostingClassifier(min\_samples\_split=42, random\_state= 0))] ====>>>> Failure in GradientBoosting, FWE f\_classif [('sm', SMOTE()), ('pca', PCA(n\_components=10)), ('model', GradientBoostingCl assifier(min samples split=42, random state=0))] Selector: NONE Model: GradientBoosting precision recall f1-score support 0 0.92 0.81 0.86 7963 1 0.50 0.73 0.59 2037 0.79 10000 accuracy macro avg 0.71 0.77 0.73 10000 weighted avg 0.83 0.79 0.81 10000 MODEL: GradientBoosting -----FWE f\_classif 0.00000 0.00000

-----

- -

[('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.1599999999999999)),
('pca', PCA(n\_components=10)), ('model', RandomForestClassifier(class\_weight

```
='balanced_subsample', criterion='entropy',
                       min_samples_split=100, n_estimators=1000,
                       random state=0))]
Selector: Variance Threshold 0.8 Model: RandomForest
                           recall f1-score
              precision
                                               support
           0
                   0.92
                             0.84
                                        0.88
                                                  7963
           1
                   0.53
                             0.70
                                        0.60
                                                  2037
                                        0.81
    accuracy
                                                 10000
                                        0.74
   macro avg
                   0.72
                             0.77
                                                 10000
weighted avg
                   0.84
                             0.81
                                        0.82
                                                 10000
[('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca',
PCA(n_components=10)), ('model', RandomForestClassifier(class_weight='balance
d subsample', criterion='entropy',
                       min samples split=100, n estimators=1000,
                       random state=0))]
Selector: RFECV RandomForest Model: RandomForest
                           recall f1-score
              precision
                                               support
                                                  7963
           0
                   0.92
                             0.85
                                        0.88
           1
                   0.54
                             0.70
                                                  2037
                                        0.61
                                       0.82
                                                 10000
    accuracy
                   0.73
                             0.77
                                        0.74
                                                 10000
   macro avg
weighted avg
                   0.84
                             0.82
                                        0.82
                                                 10000
[('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n components=1
0)), ('model', RandomForestClassifier(class_weight='balanced_subsample', crit
erion='entropy',
                       min samples split=100, n estimators=1000,
                       random state=0))]
Selector: FWE f_classif Model: RandomForest
                           recall f1-score
              precision
                                               support
                   0.92
                             0.84
           0
                                        0.88
                                                  7963
           1
                   0.53
                             0.71
                                        0.61
                                                  2037
                                        0.81
                                                 10000
    accuracy
                   0.73
                             0.77
                                        0.74
                                                 10000
   macro avg
weighted avg
                   0.84
                             0.81
                                        0.82
                                                 10000
[('sm', SMOTE()), ('pca', PCA(n_components=10)), ('model', RandomForestClassi
fier(class_weight='balanced_subsample', criterion='entropy',
                       min samples split=100, n estimators=1000,
                       random_state=0))]
Selector: NONE Model: RandomForest
              precision
                           recall f1-score
                                               support
                             0.84
           0
                   0.92
                                        0.87
                                                  7963
           1
                   0.52
                             0.70
                                        0.60
                                                  2037
                                        0.81
                                                 10000
    accuracy
```

0.74

0.82

10000

10000

0.72

0.84

0.77

0.81

macro avg

weighted avg

weighted avg

0.84

0.80

0.81

```
MODEL: GradientBoosting
. - - - - - - - - - - - - - - - - - -
                       FWE f classif
                                     0.00000
                                              0.00000
               Variance Threshold 0.8
                                     0.79380 0.58941
                              NONE
                                     0.79400
                                              0.58997
                  RFECV RandomForest 0.80150 0.59182
MODEL: RandomForest
------
                              NONE
                                     0.80930 0.59929
               Variance Threshold 0.8 0.81100 0.60059
                  RFECV RandomForest
                                     0.81520 0.60597
                       FWE f_classif
                                     0.81350
                                              0.60778
MODEL: SVC
------
                       FWE f classif
                                     0.00000 0.00000
                  RFECV RandomForest
                                     0.79530
                                              0.59068
               Variance Threshold 0.8
                                     0.79730
                                              0.59256
                              NONE
                                     0.79650
                                              0.59259
______
______
[('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.1599999999999999)),
('pca', PCA(n_components=10)), ('model', GradientBoostingClassifier(min_samp
les_split=42, random_state=0))]
Selector: Variance Threshold 0.8 Model: GradientBoosting
            precision recall f1-score
                                      support
                0.92
                        0.81
                                 0.86
                                         7963
         0
         1
                0.49
                        0.72
                                 0.59
                                         2037
                                 0.79
                                         10000
   accuracy
                0.71 0.77
                                 0.72
  macro avg
                                         10000
weighted avg
                0.83
                        0.79
                                 0.81
                                         10000
[('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca',
PCA(n_components=10)), ('model', GradientBoostingClassifier(min_samples_split
=42, random state=0))]
Selector: RFECV RandomForest Model: GradientBoosting
            precision recall f1-score
                                       support
         0
                0.92
                        0.81
                                 0.86
                                         7963
                0.49
                        0.72
                                 0.59
                                         2037
         1
```

0.79

10000

accuracy

```
0.71
                             0.77
                                        0.72
   macro avg
                                                 10000
                             0.79
                                        0.81
                                                 10000
weighted avg
                   0.83
[('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n_components=1
0)), ('model', GradientBoostingClassifier(min_samples_split=42, random_state=
0))]
====>>>> Failure in GradientBoosting, FWE f classif
[('sm', SMOTE()), ('pca', PCA(n_components=10)), ('model', GradientBoostingCl
assifier(min_samples_split=42, random_state=0))]
Selector: NONE Model: GradientBoosting
              precision
                           recall f1-score
                                               support
           0
                   0.92
                             0.81
                                        0.86
                                                  7963
           1
                   0.50
                             0.73
                                        0.59
                                                  2037
                                        0.79
    accuracy
                                                 10000
                   0.71
                             0.77
                                        0.73
                                                 10000
   macro avg
weighted avg
                   0.83
                             0.79
                                        0.81
                                                 10000
MODEL: GradientBoosting
                           FWE f classif
                                             0.00000
                                                        0.00000
                      RFECV RandomForest
                                             0.79200
                                                        0.58615
                  Variance Threshold 0.8
                                             0.79260
                                                        0.58735
                                    NONE
                                             0.79380
                                                        0.58908
[('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.1599999999999999)),
 ('pca', PCA(n_components=10)), ('model', RandomForestClassifier(class_weight
='balanced_subsample', criterion='entropy',
                       min samples split=100, n estimators=1000,
                       random state=0))]
Selector: Variance Threshold 0.8 Model: RandomForest
              precision
                           recall f1-score
                                               support
           0
                   0.92
                             0.84
                                        0.88
                                                  7963
           1
                   0.53
                             0.70
                                        0.61
                                                  2037
                                                 10000
                                        0.81
    accuracy
   macro avg
                   0.73
                             0.77
                                        0.74
                                                 10000
weighted avg
                   0.84
                             0.81
                                        0.82
                                                 10000
[('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca',
PCA(n_components=10)), ('model', RandomForestClassifier(class_weight='balance
d_subsample', criterion='entropy',
                       min samples split=100, n estimators=1000,
                       random state=0))]
Selector: RFECV RandomForest Model: RandomForest
              precision
                           recall f1-score
                                               support
                             0.84
                                        0.88
           0
                   0.92
                                                  7963
           1
                   0.53
                             0.69
                                        0.60
                                                  2037
                                        0.81
                                                 10000
    accuracy
                                        0.74
```

macro avg

0.72

0.77

weighted avg 0.84 0.81 0.82 10000

```
[('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n_components=1
0)), ('model', RandomForestClassifier(class_weight='balanced_subsample', crit
erion='entropy',
                      min_samples_split=100, n_estimators=1000,
                      random state=0))]
Selector: FWE f_classif Model: RandomForest
             precision recall f1-score
                                            support
                  0.92
          0
                            0.84
                                      0.88
                                               7963
                  0.54
          1
                            0.71
                                      0.61
                                               2037
                                     0.82
                                              10000
   accuracy
   macro avg
                  0.73
                            0.78
                                      0.75
                                              10000
weighted avg
                  0.84
                            0.82
                                      0.83
                                              10000
[('sm', SMOTE()), ('pca', PCA(n_components=10)), ('model', RandomForestClassi
fier(class_weight='balanced_subsample', criterion='entropy',
                      min_samples_split=100, n_estimators=1000,
                      random state=0))]
Selector: NONE Model: RandomForest
             precision
                         recall f1-score
                                            support
          0
                  0.91
                            0.84
                                      0.87
                                               7963
          1
                  0.52
                            0.69
                                               2037
                                     0.60
                                     0.81
                                              10000
   accuracy
                  0.72
                            0.77
                                     0.74
                                              10000
   macro avg
weighted avg
                  0.84
                            0.81
                                     0.82
                                              10000
MODEL: GradientBoosting
                          FWE f classif
                                                     0.00000
                                          0.00000
                     RFECV RandomForest
                                          0.79200
                                                     0.58615
                 Variance Threshold 0.8
                                          0.79260
                                                     0.58735
                                  NONE
                                          0.79380
                                                     0.58908
MODEL: RandomForest
                                  NONE
                                          0.80920
                                                     0.59730
                     RFECV RandomForest 0.81220
                                                     0.60093
                 Variance Threshold 0.8
                                          0.81490
                                                     0.60726
                          FWE f_classif
                                          0.81710
                                                     0.61225
[('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.1599999999999999)),
 ('pca', PCA(n_components=10)), ('model', SVC(gamma='auto', random_state=0))]
Selector: Variance Threshold 0.8 Model: SVC
             precision recall f1-score
                                           support
          0
                  0.92
                            0.82
                                      0.87
                                               7963
```

1

0.50

0.73

0.59

```
accuracy
                                    0.80
                                             10000
                 0.71
                           0.77
                                    0.73
                                             10000
  macro avg
weighted avg
                 0.84
                           0.80
                                    0.81
                                             10000
[('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca',
PCA(n_components=10)), ('model', SVC(gamma='auto', random_state=0))]
Selector: RFECV RandomForest Model: SVC
             precision recall f1-score
                                           support
                 0.920.820.500.73
          0
                                    0.87
                                              7963
          1
                                    0.60
                                              2037
                                    0.80
                                             10000
   accuracy
  macro avg
                 0.71
                           0.77
                                    0.73
                                             10000
weighted avg
                 0.84
                           0.80
                                    0.81
                                             10000
[('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n_components=1
0)), ('model', SVC(gamma='auto', random_state=0))]
====>>>> Failure in SVC, FWE f_classif
[('sm', SMOTE()), ('pca', PCA(n_components=10)), ('model', SVC(gamma='auto',
random state=0))]
Selector: NONE Model: SVC
             precision recall f1-score
                                           support
                 0.92
          0
                           0.82
                                    0.87
                                              7963
          1
                 0.50
                           0.72
                                    0.59
                                              2037
                                    0.80
                                             10000
   accuracy
                 0.71 0.77
0.84 0.80
                                    0.73
0.81
  macro avg
                                             10000
weighted avg
                                             10000
MODEL: GradientBoosting
                         FWE f_classif 0.00000 0.00000
                    RFECV RandomForest 0.79200 0.58615
                Variance Threshold 0.8
                                         0.79260
                                                   0.58735
                                 NONE
                                         0.79380 0.58908
MODEL: RandomForest
                                  NONE
                                         0.80920 0.59730
                    RFECV RandomForest
                                         0.81220
                                                   0.60093
                Variance Threshold 0.8 0.81490
                                                   0.60726
                         FWE f_classif
                                         0.81710
                                                   0.61225
MODEL: SVC
______
                                         0.00000
                                                   0.00000
                         FWE f classif
                                  NONE
                                         0.79810
                                                   0.59368
                Variance Threshold 0.8
                                         0.79810
                                                   0.59450
                    RFECV RandomForest
                                         0.79870
                                                   0.59586
```

-----

- -

-----

**PCA 14** 

-----

---

[('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.1599999999999999)),
 ('pca', PCA(n\_components=10)), ('model', GradientBoostingClassifier(min\_samp
les split=42, random state=0))]

Selector: Variance Threshold 0.8 Model: GradientBoosting

	precision	recall	f1-score	support
0 1	0.92 0.50	0.81 0.73	0.86 0.59	7963 2037
accuracy macro avg weighted avg	0.71 0.83	0.77 0.79	0.79 0.73 0.81	10000 10000 10000

[('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca',
PCA(n\_components=10)), ('model', GradientBoostingClassifier(min\_samples\_split
=42, random\_state=0))]

Selector: RFECV RandomForest Model: GradientBoosting

	precision	recall	f1-score	support
0	0.92	0.82	0.86	7963
1	0.50	0.71	0.59	2037
accuracy			0.80	10000
macro avg	0.71	0.76	0.73	10000
weighted avg	0.83	0.80	0.81	10000

[('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n\_components=1
0)), ('model', GradientBoostingClassifier(min\_samples\_split=42, random\_state=
0))]

Selector: FWE f\_classif Model: GradientBoosting

support	f1-score	recall	precision	
7963	0.87	0.82	0.92	0
2037	0.60	0.73	0.51	1
10000	0.80			accuracy
10000 10000	0.73 0.81	0.77 0.80	0.71 0.84	macro avg weighted avg

[('sm', SMOTE()), ('pca', PCA(n\_components=10)), ('model', GradientBoostingCl
assifier(min\_samples\_split=42, random\_state=0))]

Selector: NONE Model: GradientBoosting

	precision	recall	f1-score	support
0	0.92	0.81	0.86	7963
1	0.49	0.72	0.58	2037
				40000
accuracy			0.79	10000
macro avg	0.70	0.76	0.72	10000

weighted avg 0.83 0.79 0.80 10000

MODEL: GradientBoosting \_\_\_\_\_\_ NONE 0.79010 0.58328 RFECV RandomForest 0.79540 0.58650 Variance Threshold 0.8 0.79430 0.58967 FWE f\_classif 0.79940 0.59686 [('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.159999999999999)), ('pca', PCA(n\_components=10)), ('model', RandomForestClassifier(class\_weight ='balanced\_subsample', criterion='entropy', min samples split=100, n estimators=1000, random state=0))] Selector: Variance Threshold 0.8 Model: RandomForest precision recall f1-score support 0 0.92 0.84 0.88 7963 1 0.53 0.70 0.60 2037 0.81 10000 accuracy 0.72 0.77 0.74 10000 macro avg 0.82 10000 weighted avg 0.84 0.81 [('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca', PCA(n components=10)), ('model', RandomForestClassifier(class weight='balance d\_subsample', criterion='entropy', min\_samples\_split=100, n\_estimators=1000, random state=0))] Selector: RFECV RandomForest Model: RandomForest precision recall f1-score support 0 0.91 0.84 0.88 7963 1 0.53 0.69 0.60 2037 accuracy 0.81 10000 0.72 0.74 macro avg 0.77 10000 weighted avg 0.84 0.81 0.82 10000 [('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n\_components=1 0)), ('model', RandomForestClassifier(class\_weight='balanced\_subsample', crit erion='entropy', min\_samples\_split=100, n\_estimators=1000, random\_state=0))] ====>>>> Failure in RandomForest, FWE f\_classif [('sm', SMOTE()), ('pca', PCA(n\_components=10)), ('model', RandomForestClassi fier(class\_weight='balanced\_subsample', criterion='entropy', min\_samples\_split=100, n\_estimators=1000, random state=0))] Selector: NONE Model: RandomForest precision recall f1-score support 0 0.92 0.84 0.87 7963

1

0.52

0.70

0.60

```
accuracy
                                       0.81
                                               10000
                   0.72
                            0.77
                                       0.74
                                               10000
   macro avg
weighted avg
                   0.84
                            0.81
                                       0.82
                                               10000
MODEL: GradientBoosting
                                    NONE
                                           0.79010
                                                      0.58328
                      RFECV RandomForest
                                           0.79540
                                                      0.58650
                  Variance Threshold 0.8
                                           0.79430
                                                      0.58967
                           FWE f_classif
                                           0.79940
                                                      0.59686
MODEL: RandomForest
------
                           FWE f_classif
                                           0.00000
                                                      0.00000
                      RFECV RandomForest
                                           0.81150
                                                      0.59902
                                   NONE
                                           0.80950
                                                      0.60038
                  Variance Threshold 0.8
                                           0.81070
                                                      0.60156
[('sm', SMOTE()), ('sel', VarianceThreshold(threshold=0.159999999999999)),
 ('pca', PCA(n_components=10)), ('model', SVC(gamma='auto', random_state=0))]
Selector: Variance Threshold 0.8 Model: SVC
              precision recall f1-score
                                             support
                   0.92
                            0.82
                                       0.87
                                                7963
           0
           1
                   0.50
                            0.73
                                      0.59
                                                 2037
    accuracy
                                      0.80
                                               10000
   macro avg
                   0.71
                            0.77
                                       0.73
                                               10000
weighted avg
                  0.84
                            0.80
                                       0.81
                                               10000
[('sm', SMOTE()), ('sel', RFECV(estimator=RandomForestClassifier())), ('pca',
PCA(n_components=10)), ('model', SVC(gamma='auto', random_state=0))]
Selector: RFECV RandomForest Model: SVC
              precision
                          recall f1-score
                                             support
           0
                   0.92
                            0.81
                                      0.86
                                                 7963
           1
                   0.49
                            0.72
                                       0.59
                                                 2037
                                       0.79
    accuracy
                                               10000
                   0.71
                                       0.72
                                               10000
   macro avg
                            0.77
                            0.79
                                       0.80
weighted avg
                   0.83
                                               10000
[('sm', SMOTE()), ('sel', SelectFwe(alpha=0.01)), ('pca', PCA(n_components=1
0)), ('model', SVC(gamma='auto', random_state=0))]
====>>>> Failure in SVC, FWE f classif
[('sm', SMOTE()), ('pca', PCA(n_components=10)), ('model', SVC(gamma='auto',
 random_state=0))]
Selector: NONE Model: SVC
              precision
                          recall f1-score
                                             support
           0
                   0.92
                            0.81
                                       0.86
                                                 7963
```

1	0.50	0.73	0.59	2037		
accuracy			0.80	10000		
macro avg	0.71	0.77	0.73			
weighted avg	0.83		0.81			
8						
MODEL . Co. di	LD					
MODEL: Gradien	tBoosting					
			NONE	0.79010	0.58328	
	RFECV	/ RandomFo	rest	0.79540	0.58650	
	Variance	Threshold	0.8	0.79430	0.58967	
		FWE f_cla	ssif	0.79940	0.59686	
MODEL: RandomF	orest					
		FWE f cla	ssif	0.00000	0.00000	
				0.81150		
				0.80950		
	Variance			0.81070		
MODEL: SVC						
		C 1				
				0.00000		
	RFECV			0.79180		
				0.79650	0.59210	
	Variance	Threshold	0.8	0.79800	0.59405	

In [168]: | npca\_a = ncap\_a

```
In [169]: for i, d, m in zip(ncap_a, desc_strings_mmspca_a, all_models_mmspca_a):
    print(f'PCA {i}')
    print('-' * 120)
    print(d)
    plot_graph_1(m)
```

PCA 5				
MODEL: GradientBoo	osting			
	 FWE f_classif	0.00000	0.00000	
	NONE	0.78990	0.58552	
	Variance Threshold 0.8			
		0.80290		
MODEL: RandomFores	s†			
	FWE f_classif	0.00000	0.00000	
	Variance Threshold 0.8	0.80930	0.59537	
	RFECV RandomForest	0.81420	0.60214	
	NONE	0.81330	0.60604	
MODEL: SVC				
	FWE f_classif	0.00000	0.00000	
		0.79550		
	Variance Threshold 0.8			
	NONE	0.79820	0.59478	
PCA 6				
MODEL: GradientBoo	<del>-</del>			
	 FWE f classif	0 00000	0 00000	
	<del>-</del>			
	RFECV RandomForest	0.79150	0.58458	
	Variance Threshold 0.8			
MODEL: RandomFores	st			
	 FME f classif	0 00000	0 00000	
	FWE f_classif Variance Threshold 0.8	0.00000 0.00000	0.00000 0.59669	
		0.81250		
	RFECV RandomForest			
MODEL: SVC				
		0.70450	0 50057	
	Variance Threshold 0.8	u./9450	0.5895/	

NONE

0.79550

0.58960

RFECV RandomForest FWE f_classif			
PCA 7			
MODEL: GradientBoosting			
NONE		0.58299	
Variance Threshold 0.8 RFECV RandomForest			
FWE f_classif			
· wc · _ctassi.			
MODEL: RandomForest			
FWE f_classif	0.00000	0.00000	
Variance Threshold 0.8			
	0.81160		
RFECV RandomForest	0.81470	0.60600	
MODEL: SVC			
FWE f_classif			
	0.79810		
RFECV RandomForest Variance Threshold 0.8			
variance inresnoid 0.8	0./9//0	0.59516	
PCA 8			
MODEL: GradientBoosting			
FWE f_classif			
	0.79170		
RFECV RandomForest Variance Threshold 0.8			
valiance in eshou 0.8			
MODEL: RandomForest			
FWE f_classif	0.00000	0.00000	
RFECV RandomForest			
	0.81250		
Variance Threshold 0.8	0.81320	0.60574	

MODEL: SVC	
FWE f class	if 0.00000 0.00000
<del>-</del>	st 0.79430 0.59081
	NE 0.79770 0.59369
Variance Threshold 0	.8 0.79810 0.59482
PCA 9	
MODEL: GradientBoosting	
NO	NE 0.78660 0.57809
	st 0.79170 0.58181
Variance Threshold 0	.8 0.79690 0.59388
	if 0.80050 0.59932
MODEL . Dandon Farrach	
MODEL: RandomForest	
	NE 0.81140 0.60127
RFECV RandomFore	st 0.81440 0.60137
Variance Threshold 0	.8 0.81280 0.60289
FWE f_class	if 0.81890 0.61769
MODEL: SVC	
	if 0.00000 0.00000
Variance Threshold 0	
RFECV RandomFore	
	NE 0.79860 0.59346
PCA 10	
MODEL: GradientBoosting	
<del>-</del>	if 0.00000 0.00000
RFECV RandomFore	
	NE 0.79500 0.58918
Variance Threshold 0	.8 0.79650 0.58947
MODEL: RandomForest	

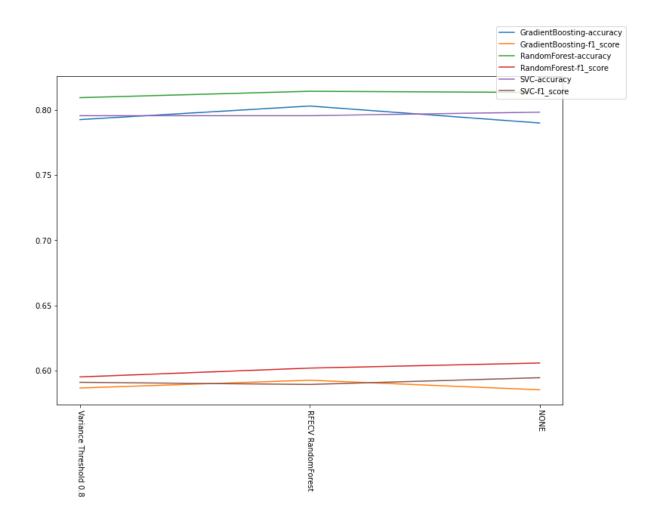
RFECV RandomForest 0.81080 0.59881

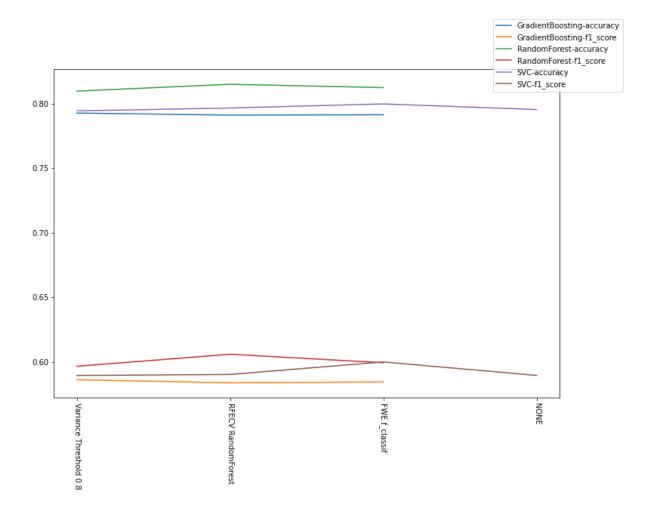
		0.81460	0.60737	
	FWE f_classif	0.81890	0.61/04 	
MODEL: SVC				
	RFECV RandomForest	0 79120	0 58653	
	Variance Threshold 0.8			
		0.79820		
	FWE f_classif	0.80200	0.60337	
PCA 11				
MODEL: GradientBo	posting			
	NONF	0.79300	0.58781	
	Variance Threshold 0.8			
	RFECV RandomForest			
	FWE f_classif	0.80360	0.60114	
MODEL: RandomFore	est 			
	RFECV RandomForest	0.81240	0.59288	
	Variance Threshold 0.8	0.81210	0.60400	
		0.81270		
	FWE f_classif	0.81740	0.61379	
MODEL: SVC				
	FWE f_classif	0.00000	0.00000	
	RFECV RandomForest	0.79630	0.59187	
	Variance Threshold 0.8			
	NONE	0.79780	0.59365	
PCA 12				
MODEL: GradientBo	oosting			
	 FWE f classif	0.00000	0.00000	
	Variance Threshold 0.8			
		0.79400		
	RFECV RandomForest	0.80150	0.59182	

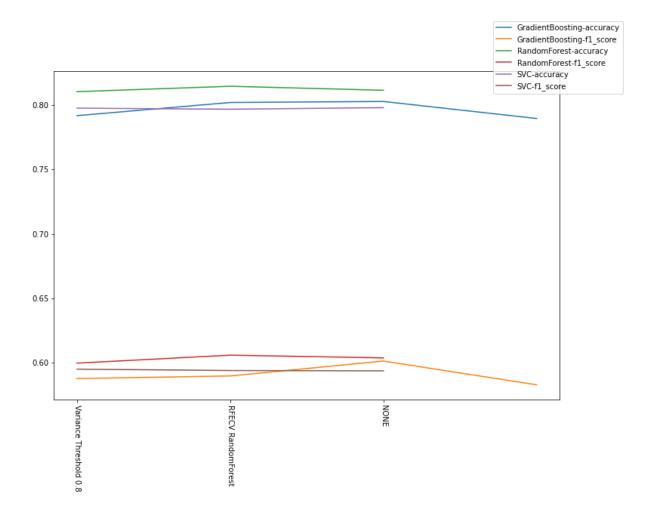
-----

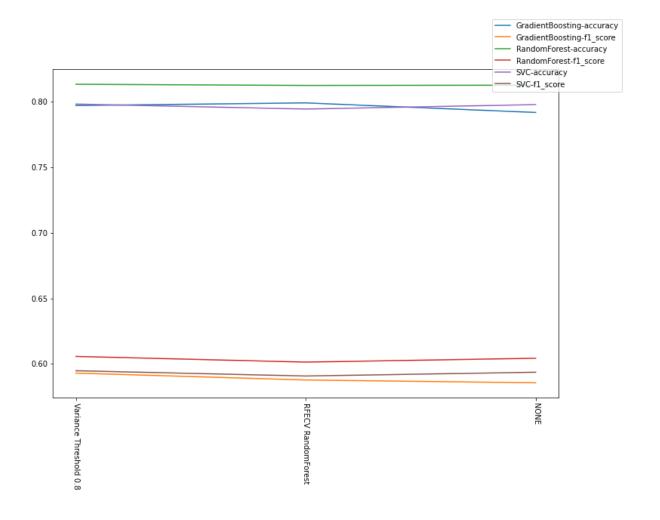
MODEL: RandomForest			
NONE Variance Threshold 0.8 RFECV RandomForest FWE f_classif	0.81100 0.81520	0.60059 0.60597	
MODEL: SVC			
FWE f_classif RFECV RandomForest Variance Threshold 0.8 NONE	0.79530	0.59068 0.59256	
 PCA 13			
MODEL: GradientBoosting			
FWE f_classif RFECV RandomForest	0.79200	0.58615	
Variance Threshold 0.8 NONE	0.79260 0.79380		
MODEL: RandomForest			
NONE		0.59730	
RFECV RandomForest Variance Threshold 0.8		0.60093	
FWE f_classif			
MODEL: SVC			
FWE f_classif		0.00000	
NONE	0.79810	0.59368	
Variance Threshold 0.8 RFECV RandomForest			
PCA 14			
	-		
MODEL: GradientBoosting			

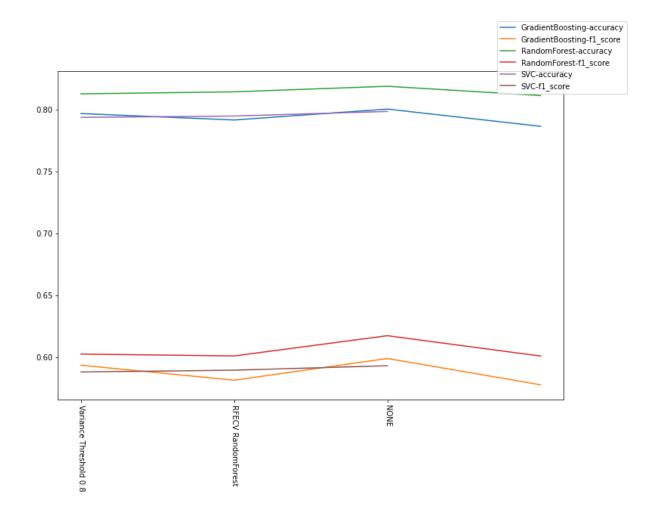
	NONE RFECV RandomForest Variance Threshold 0.8 FWE f_classif	0.79430	0.58967
MODEL: RandomFore	st		
	FWE f_classif	0.00000	0.00000
	RFECV RandomForest		
		0.80950	
	Variance Threshold 0.8	0.81070	0.60156
MODEL: SVC			
	FWE f_classif	0.00000	0.00000
	<del></del>	0.79180	0.58559
		0.79650	
	Variance Threshold 0.8		0.59405

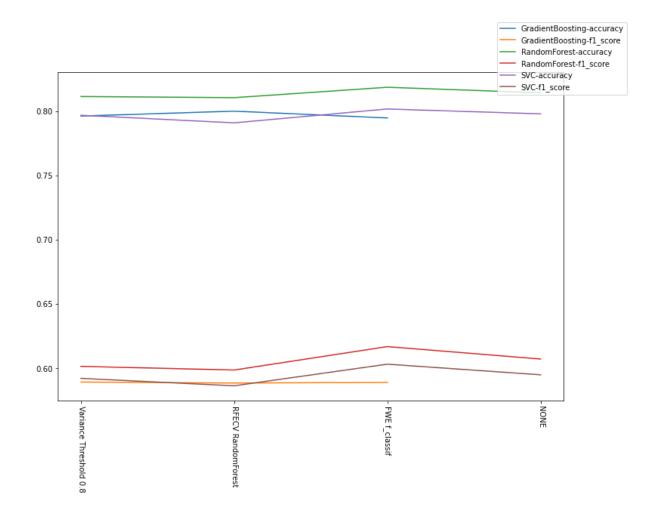


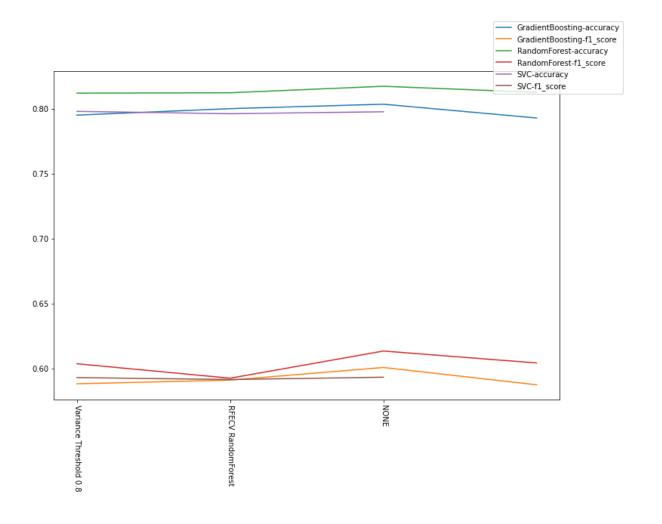


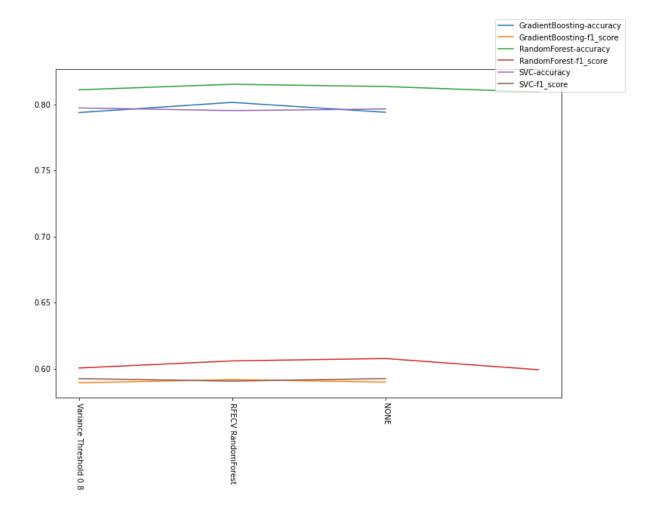


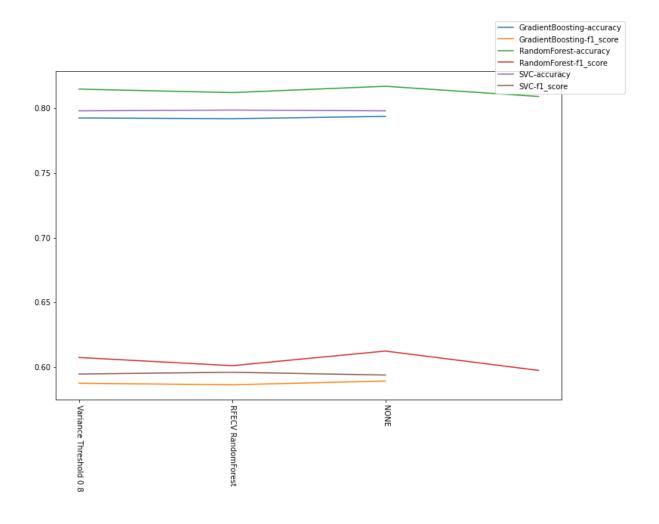


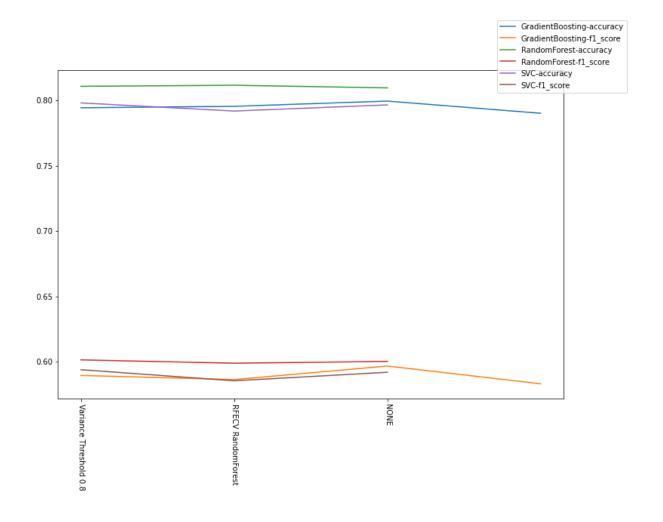












## Research

**Experiment with weighted features in Random Forest** 

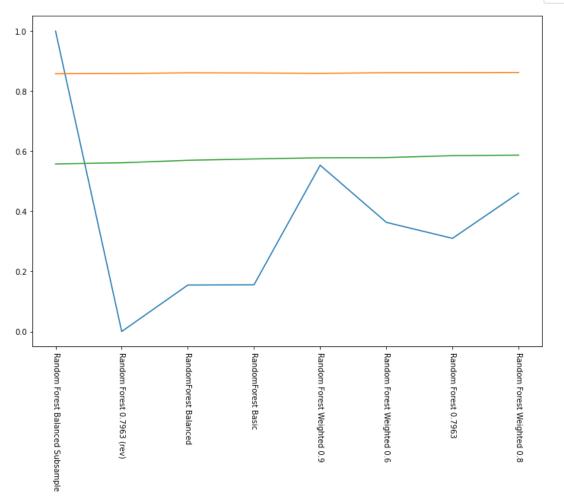
```
In [182]:
          # First check with default parameters, with just this parameter changed
          basic models = [
              ('RandomForest Basic', lambda: RandomForestClassifier(random_state=1)),
              ('RandomForest Balanced', lambda: RandomForestClassifier(random state=1, c
          lass weight='balanced')),
              ('Random Forest Balanced Subsample', lambda: RandomForestClassifier(random
          _state=1, class_weight='balanced_subsample')),
              ('Random Forest Weighted 0.8', lambda: RandomForestClassifier(random state
          =1, class_weight={1: 0.2, 0: 0.8})),
              ('Random Forest Weighted 0.6', lambda: RandomForestClassifier(random_state
          =1, class weight={1: 0.4, 0: 0.6})),
              ('Random Forest Weighted 0.9', lambda: RandomForestClassifier(random_state
          =1, class_weight={1: 0.1, 0: 0.9})),
              ('Random Forest 0.7963', lambda: RandomForestClassifier(random state=1, cl
          ass weight={1: 2037, 0: 7963})),
              ('Random Forest 0.7963 (rev)', lambda: RandomForestClassifier(random_state
          =1, class_weight={0: 2037, 1: 7963})),
          compare_basic_models(run_basic_model, basic_models)
```

RandomForest Basic <function <lambda> at 0x000001EC43C1E5E0>
RandomForest Balanced <function <lambda> at 0x000001EC43C1E550>
Random Forest Balanced Subsample <function <lambda> at 0x0000001EC43FA2D30>
Random Forest Weighted 0.8 <function <lambda> at 0x000001EC43FA23A0>
Random Forest Weighted 0.6 <function <lambda> at 0x0000001EC43FA2550>
Random Forest Weighted 0.9 <function <lambda> at 0x0000001EC43FA2700>
Random Forest 0.7963 <function <lambda> at 0x0000001EC43FA28B0>
Random Forest 0.7963 (rev) <function <lambda> at 0x0000001EC43FA2A60>

[0.8603999488217603, 0.8609005089738355, 0.8579003286977658, 0.86200028901782 67, 0.8613003689778251, 0.8592000087377483, 0.8617004690298435, 0.85870076884 98315]

[0.5742653346426003, 0.5697548162058486, 0.5577101012466912, 0.58704804321528 09, 0.578779895715622, 0.5782437561433658, 0.5852288885179198, 0.561476369894 2369]

Time		Accuracy		
1.000	Random Forest Balanced Subsample	 0.858	0.558	
1.000	Random Forest 0.7963 (rev)	 0.859	0.561	
0.000	, ,			
0.155	RandomForest Balanced	 0.861	0.570	
0.133	RandomForest Basic	 0.860	0.574	
0.155				
0.553	Random Forest Weighted 0.9	 0.859	0.578	
0.333	Random Forest Weighted 0.6	 0.861	0.579	
0.363	D		0 505	
0.310	Random Forest 0.7963	 0.862	0.585	
0.310	Random Forest Weighted 0.8	 0.862	0.587	
0.461				



```
In [245]:
          # Now check with the parameters obtained in the baseline
          basic models = [
               ('RandomForest Basic', lambda: RandomForestClassifier(random state=1, crit
          erion='entropy', min samples split=100, n estimators=1000)),
               ('RandomForest Balanced', lambda: RandomForestClassifier(random_state=1, c
          lass_weight='balanced', criterion='entropy', min_samples_split=100, n_estimato
          rs=1000)),
               ('Random Forest Balanced Subsample', lambda: RandomForestClassifier(random
           _state=1, class_weight='b<mark>alanced_subsample</mark>', criterion='<mark>entropy</mark>', min_samples_
          split=100, n_estimators=1000)),
               ('Random Forest Weighted 1:4', lambda: RandomForestClassifier(random_state
          =1, class_weight={1: 0.2, 0: 0.8}, criterion='entropy', min_samples_split=100,
          n estimators=1000)),
               ('Random Forest Weighted 2:3', lambda: RandomForestClassifier(random state
           =1, class weight={1: 0.4, 0: 0.6}, criterion='entropy', min samples split=100,
          n_estimators=1000)),
               ('Random Forest Weighted 4:1', lambda: RandomForestClassifier(random state
          =1, class_weight={1: 4, 0: 1}, criterion='entropy', min_samples_split=100, n_e
          stimators=1000)),
               ('Random Forest Weighted 4:1.33', lambda: RandomForestClassifier(random st
          ate=1, class_weight={1: 4, 0: 1.33}, criterion='entropy', min_samples_split=10
          0, n estimators=1000)),
               ('Random Forest Weighted 6:1', lambda: RandomForestClassifier(random state
          =1, class_weight={1: 6, 0: 1}, criterion='entropy', min_samples_split=100, n_e
           stimators=1000)),
               ('Random Forest Weighted 1:9', lambda: RandomForestClassifier(random state
          =1, class_weight={1: 0.1, 0: 0.9}, criterion='entropy', min_samples_split=100,
          n estimators=1000)),
               ('Random Forest Weighted 9:1', lambda: RandomForestClassifier(random state
          =1, class weight={1: 0.9, 0: 0.1}, criterion='entropy', min samples split=100,
          n_estimators=1000)),
          compare_basic_models(run_basic_model, basic_models)
```

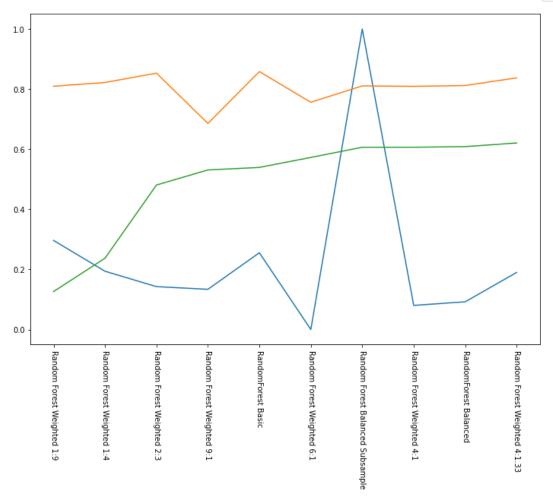
RandomForest Basic <function <lambda> at 0x000001D924FA4EE0>
RandomForest Balanced <function <lambda> at 0x000001D924FA4940>
Random Forest Balanced Subsample <function <lambda> at 0x000001D924FA4960>
Random Forest Weighted 1:4 <function <lambda> at 0x000001D924FA40D0>
Random Forest Weighted 2:3 <function <lambda> at 0x000001D924FA41F0>
Random Forest Weighted 4:1 <function <lambda> at 0x000001D924FA4D30>
Random Forest Weighted 4:1.33 <function <lambda> at 0x000001D924FA4DC0>
Random Forest Weighted 6:1 <function <lambda> at 0x000001D924FA4DC0>
Random Forest Weighted 1:9 <function <lambda> at 0x000001D924FA49D0>
Random Forest Weighted 9:1 <function <lambda> at 0x000001D924FA4700>
Random Forest Weighted 9:1 <function <lambda> at 0x000001D924FA4280>

[0.8584003487417762, 0.8120002050010166, 0.8108001448929902, 0.82210010578916 64, 0.8532003683296954, 0.8092998447129301, 0.8374993068613168, 0.75629928036 00144, 0.8096005648810214, 0.6858987146148201]

[0.5394276378866715, 0.6085874497161349, 0.6065484980293258, 0.23713837863635 384, 0.48082540319439254, 0.6066130501763358, 0.6208418323813478, 0.572675352 1214187, 0.12621721702162433, 0.5309915187663873]

Time	Model	 Accuracy	F1 Score	Run
0.200	Random Forest Weighted 1:9	 0.810	0.126	
0.296	Dandom Fanast Waighted 1.4	0 022	0 227	
0 104	Random Forest Weighted 1:4	 0.822	0.237	
0.194	Random Forest Weighted 2:3	 0.853	0.481	
0.143				
	Random Forest Weighted 9:1	 0.686	0.531	
0.134				
	RandomForest Basic	 0.858	0.539	
0.255				
	Random Forest Weighted 6:1	 0.756	0.573	
0.000				
	Random Forest Balanced Subsample	 0.811	0.607	
1.000				
	Random Forest Weighted 4:1	 0.809	0.607	
0.080				
	RandomForest Balanced	 0.812	0.609	
0.092				
0.400	Random Forest Weighted 4:1.33	 0.837	0.621	
0.190				

-----



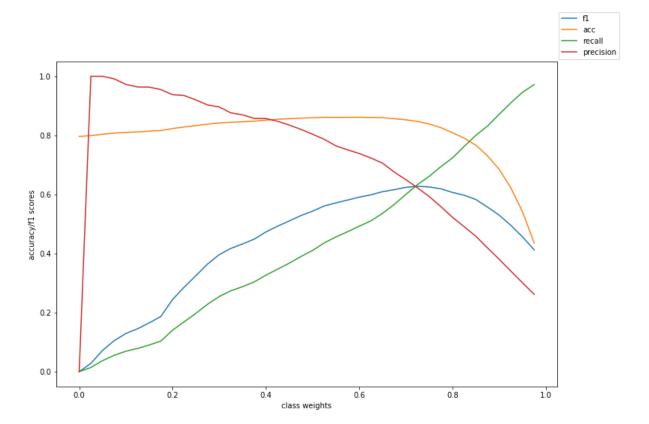
```
In [282]:
          # Now plot with by varying the params
          import copy
          def compare_weighted_rfcs(X, y):
              Compare the effect of alternate class weights on RFC
              weights = []
              f1 scores = []
              acc_scores = []
              recall_scores = []
              precision scores = []
              for i in np.arange(0, 1.1, 0.025):
                  desc = "Random Forest: Weight = %f" % i
                   if (1-i) > 0:
                       cw = \{1: i, 0: 1-i\}
                       rfc = RandomForestClassifier(random_state=1, class_weight=cw, crit
          erion='entropy',\
                                                             min samples split=100, n est
          imators=1000)
                       rfc.fit(X, y)
                      y_pred = cross_val_predict(rfc, X, y,\
                                                  cv=model_selection.StratifiedKFold(n_sp
          lits=6, shuffle=True, random state=2),\
                                                  n_jobs=7)
                       f1 = metrics.f1_score(y, y_pred)
                      weights.append(i)
                       f1 scores.append(f1)
                       acc_scores.append(metrics.accuracy_score(y, y_pred))
                       recall scores.append(metrics.recall score(y, y pred))
                       precision_scores.append(metrics.precision_score(y, y_pred))
                       print("%8.4f %8.4f" % (i, f1, ))
              return weights, f1_scores, acc_scores, recall_scores, precision_scores
          weights, f1_scores, acc_scores, rec_scores, prec_scores = compare_weighted_rfc
          s(scaled_X.to_numpy(), y.to_numpy())
```

0.0000	0.0000
0.0250	0.0281
0.0500	0.0719
0.0750	0.1051
0.1000	0.1292
0.1250	0.1453
0.1500	0.1652
0.1750	0.1869
0.2000	0.2442
0.2250	0.2861
0.2500	0.3250
0.2750	0.3644
0.3000	0.3962
0.3250	0.4174
0.3500	0.4323
0.3750	0.4487
0.4000	0.4728
0.4250	0.4925
0.4500	0.5101
0.4750	0.5285
0.5000	0.5434
0.5250	0.5610
0.5500	0.5713
0.5750	0.5810
0.6000	0.5909
0.6250	0.5986
0.6500	0.6094
0.6750	0.6161
0.7000	0.6239
0.7250	0.6277
0.7500	0.6252
0.7750	0.6191
0.8000	0.6067
0.8250	0.5973
0.8500	0.5828
0.8750	0.5571
0.9000	0.5296
0.9250	0.4953

0.95000.45630.97500.4123

```
In [284]: fig, ax = plt.subplots()
    ax.plot(weights, f1_scores, label='f1')
    ax.plot(weights, acc_scores, label='acc')
    ax.plot(weights, rec_scores, label="recall")
    ax.plot(weights, prec_scores, label="precision")
    ax.set_xlabel('class weights')
    ax.set_ylabel('accuracy/f1 scores')
    fig.legend()
```

Out[284]: <matplotlib.legend.Legend at 0x1d85a47c850>



```
weight = 0.00 | f1 = 0.0000 | accuracy = 0.7963 | recall = 0.0000 | precision
= 0.0000
weight = 0.03 | f1 = 0.0281 | accuracy = 0.7992 | recall = 0.0142 | precision
= 1.0000
weight = 0.05 | f1 = 0.0719 | accuracy = 0.8039 | recall = 0.0373 | precision
= 1.0000
weight = 0.08 | f1 = 0.1051 | accuracy = 0.8075 | recall = 0.0555 | precision
= 0.9912
weight = 0.10 | f1 = 0.1292 | accuracy = 0.8100 | recall = 0.0692 | precision
= 0.9724
weight = 0.12 | f1 = 0.1453 | accuracy = 0.8117 | recall = 0.0785 | precision
= 0.9639
weight = 0.15 | f1 = 0.1652 | accuracy = 0.8140 | recall = 0.0903 | precision
= 0.9634
weight = 0.18 | f1 = 0.1869 | accuracy = 0.8164 | recall = 0.1036 | precision
= 0.9548
weight = 0.20 | f1 = 0.2442 | accuracy = 0.8230 | recall = 0.1404 | precision
= 0.9377
weight = 0.23 | f1 = 0.2861 | accuracy = 0.8283 | recall = 0.1689 | precision
= 0.9348
weight = 0.25 | f1 = 0.3250 | accuracy = 0.8330 | recall = 0.1973 | precision
= 0.9199
weight = 0.28 | f1 = 0.3644 | accuracy = 0.8378 | recall = 0.2283 | precision
= 0.9029
weight = 0.30 | f1 = 0.3962 | accuracy = 0.8421 | recall = 0.2543 | precision
= 0.8962
weight = 0.33 | f1 = 0.4174 | accuracy = 0.8442 | recall = 0.2739 | precision
= 0.8760
weight = 0.35 | f1 = 0.4323 | accuracy = 0.8461 | recall = 0.2877 | precision
= 0.8694
weight = 0.38 | f1 = 0.4487 | accuracy = 0.8479 | recall = 0.3039 | precision
= 0.8573
weight = 0.40 | f1 = 0.4728 | accuracy = 0.8517 | recall = 0.3265 | precision
= 0.8570
weight = 0.43 | f1 = 0.4925 | accuracy = 0.8543 | recall = 0.3471 | precision
= 0.8477
weight = 0.45 | f1 = 0.5101 | accuracy = 0.8563 | recall = 0.3672 | precision
= 0.8348
weight = 0.48 | f1 = 0.5285 | accuracy = 0.8583 | recall = 0.3898 | precision
= 0.8202
weight = 0.50 | f1 = 0.5434 | accuracy = 0.8595 | recall = 0.4104 | precision
= 0.8038
weight = 0.53 | f1 = 0.5610 | accuracy = 0.8610 | recall = 0.4359 | precision
= 0.7865
weight = 0.55 \mid f1 = 0.5713 \mid accuracy = 0.8606 \mid recall = 0.4561 \mid precision
= 0.7646
weight = 0.58 | f1 = 0.5810 | accuracy = 0.8608 | recall = 0.4737 | precision
= 0.7510
weight = 0.60 | f1 = 0.5909 | accuracy = 0.8611 | recall = 0.4924 | precision
= 0.7386
weight = 0.62 | f1 = 0.5986 | accuracy = 0.8605 | recall = 0.5106 | precision
= 0.7232
weight = 0.65 | f1 = 0.6094 | accuracy = 0.8600 | recall = 0.5361 | precision
= 0.7059
weight = 0.68 | f1 = 0.6161 | accuracy = 0.8563 | recall = 0.5660 | precision
= 0.6758
weight = 0.70 | f1 = 0.6239 | accuracy = 0.8527 | recall = 0.5999 | precision
```

```
= 0.6500
weight = 0.73 | f1 = 0.6277 | accuracy = 0.8470 | recall = 0.6333 | precision
weight = 0.75 | f1 = 0.6252 | accuracy = 0.8385 | recall = 0.6613 | precision
= 0.5929
weight = 0.78 | f1 = 0.6191 | accuracy = 0.8260 | recall = 0.6942 | precision
= 0.5587
weight = 0.80 | f1 = 0.6067 | accuracy = 0.8089 | recall = 0.7236 | precision
= 0.5223
weight = 0.83 | f1 = 0.5973 | accuracy = 0.7905 | recall = 0.7629 | precision
= 0.4908
weight = 0.85 | f1 = 0.5828 | accuracy = 0.7666 | recall = 0.8002 | precision
= 0.4583
weight = 0.88 | f1 = 0.5571 | accuracy = 0.7306 | recall = 0.8316 | precision
= 0.4188
weight = 0.90 | f1 = 0.5296 | accuracy = 0.6847 | recall = 0.8714 | precision
= 0.3804
weight = 0.93 | f1 = 0.4953 | accuracy = 0.6222 | recall = 0.9102 | precision
= 0.3402
weight = 0.95 | f1 = 0.4563 | accuracy = 0.5407 | recall = 0.9460 | precision
= 0.3006
weight = 0.98 | f1 = 0.4123 | accuracy = 0.4355 | recall = 0.9720 | precision
= 0.2616
```

**Grid Search different weights (no SMOTE)** 

```
In [243]: # Try a grid wearch with different weights, and with some other parameters
          def grid search random forest weighted(X, y):
              rfc = RandomForestClassifier(random state=1)
              p = imblearn.pipeline.Pipeline([('rfc', rfc)])
              class_weights = [{0: i, 1: 1-i}] for i in np.arange(0, 1, 0.05)]
              class weights.append({1: 4, 0: 1.33})
              class weights.append({1: 4, 0: 1})
              class_weights.append({1: 5, 0: 1})
              param_grid = {
                  'rfc n estimators': [10, 100, 1000],
                   'rfc__criterion': ['gini', 'entropy'],
                  'rfc__min_samples_split': [i for i in range(20, 200, 40)],
                  'rfc__class_weight': class_weights
              }
              #scoring = {'F1': make_scorer(metrics.f1_score), 'ACC': make_scorer(metric
          s.accuracy score)}
              scorer = make_scorer(metrics.f1_score)
              cv1 = StratifiedKFold(n_splits=6, random_state=10, shuffle=True)
              search = GridSearchCV(p, param grid, n jobs=7, cv=cv1, scoring=scorer)
              search.fit(X, y)
              print("RANDOM FOREST")
              print(f"Best Params = {search.best params }")
              print(f"Best score = {search.best_score_}")
              print('-' * 120)
              return search.best estimator
          random.seed(22)
          rfcw = grid search random forest weighted(scaled X.to numpy(), y.to numpy())
          RANDOM FOREST
          Best Params = {'rfc__class_weight': {0: 0.2, 1: 0.8}, 'rfc__criterion': 'gin
          i', 'rfc__min_samples_split': 20, 'rfc__n_estimators': 100}
          Best score = 0.6323982569024459
```

GridSearch BalancedRandomForestClassifier from Imblearn

```
In [276]: def grid search balanced random forest weighted(X, y):
              rfc = BalancedRandomForestClassifier(random state=1)
              p = imblearn.pipeline.Pipeline([('rfc', rfc)])
              param_grid = {
                  'rfc__n_estimators': [10, 100, 1000],
                   'rfc__criterion': ['gini', 'entropy'],
                  'rfc min samples split': [i for i in range(20, 200, 40)],
                  'rfc__class_weight': [{0: i, 1: 1-i} for i in np.arange(0, 1, 0.1)],
                  'rfc__bootstrap': [True, False],
                   'rfc oob score': [True, False],
              }
              #scoring = {'F1': make_scorer(metrics.f1_score), 'ACC': make_scorer(metric
          s.accuracy score)}
              scorer = make scorer(metrics.f1 score)
              cv1 = StratifiedKFold(n_splits=6, random_state=10, shuffle=True)
              search = GridSearchCV(p, param_grid, n_jobs=7, cv=cv1, scoring=scorer)
              search.fit(X, y)
              print("RANDOM FOREST")
              print(f"Best Params = {search.best params }")
              print(f"Best score = {search.best score }")
              print('-' * 120)
              return search.best estimator
          random.seed(22)
          bwrfc = grid_search_balanced_random_forest_weighted(scaled_X.to_numpy(), y.to
          _numpy())
          RANDOM FOREST
          Best Params = {'rfc_bootstrap': True, 'rfc_class_weight': {0: 0.700000000000
          00001, 1: 0.299999999999999, 'rfc__criterion': 'entropy', 'rfc__min_sample
          s_split': 20, 'rfc__n_estimators': 100, 'rfc__oob_score': True}
          Best score = 0.6329106408569257
```

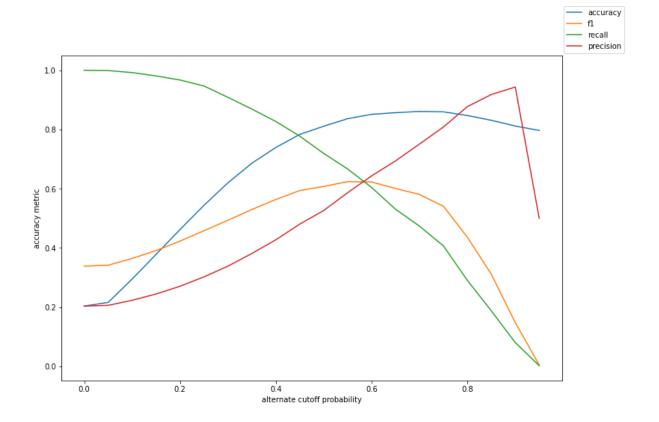
Random Forest with Adjusted Cutoff

```
In [290]:
          def random forest alternate cutoff(X, y, cutoff, use smote=False):
              Using Alternate cut-off is different from assigning class weights. Assigni
          ng class weights
              actually changes the mode, but this doens't change the model.
              The only thing that changes is that we predict a probability instead of a
           definite value,
              and then set a cut-off on that probability.
              f1_scores = []
              accuracy_scores = []
              recall_scores = []
              precision_scores = []
              kf = model selection.StratifiedKFold(n splits=6, shuffle=True, random stat
          e=2)
              for train_index, test_index in kf.split(X, y):
                   rfc = RandomForestClassifier(random_state=0, class_weight='balanced_su
          bsample', criterion='entropy', min_samples_split=100, n_estimators=1000, n_job
          s=7)
                  X train, y train = X[train index, :], y[train index]
                  X_test, y_test = X[test_index, :], y[test_index]
                  if use_smote:
                      sm = SMOTE()
                      X_train, y_train = sm.fit_resample(X_train, y_train)
                   rfc.fit(X_train, y_train)
                  y pred proba = rfc.predict proba(X test)
                  y_pred = (y_pred_proba[:, 1] >= cutoff).astype('int')
                  f1s = metrics.f1_score(y_test, y_pred)
                  accs = metrics.accuracy score(y test, y pred)
                  f1 scores.append(f1s)
                   recall_scores.append(metrics.recall_score(y_test, y_pred))
                   precision_scores.append(metrics.precision_score(y_test, y_pred))
                   accuracy scores.append(accs)
              return mean(accuracy_scores), mean(f1_scores), mean(recall_scores), mean(p
          recision_scores)
```

```
In [295]: | acc_scores = []
          f1_scores = []
          rec_scores = []
          prec scores = []
          cutoffs = []
          for i in np.arange(0, 1, 0.05):
              accs, f1s, rec, prec = random_forest_alternate_cutoff(scaled_X.to_numpy(),
          y.to_numpy(), i)
              acc_scores.append(accs)
              f1_scores.append(f1s)
              rec_scores.append(rec)
              prec_scores.append(prec)
              cutoffs.append(i)
          for i, j, k, l, m in zip(cutoffs, acc_scores, f1_scores, prec_scores, rec_scor
          es):
              print("%10.5f %20.10f %20.10f %20.10f %20.10f" % (i, j, k, l, m))
          fig, ax = plt.subplots()
          ax.plot(cutoffs, acc_scores, label="accuracy")
          ax.plot(cutoffs, f1 scores, label="f1")
          ax.plot(cutoffs, rec_scores, label="recall")
          ax.plot(cutoffs, prec_scores, label="precision")
          ax.set_xlabel('alternate cutoff probability')
          ax.set_ylabel('accuracy metric')
          fig.legend()
```

0.00000	0.2036999563	0.3384562867	0.2036999563
1.0000000000		0.0447600004	0 0064400054
0.05000	0.2160994572	0.3417630524	0.2061432254
0.9990196078	0. 2040004225	0.2644404072	0 2222227060
0.10000	0.2948994235	0.3644404072	0.2232327069
0.9921481867	0. 2701006104	0.2012567066	0 2442055444
0.15000	0.3781006104	0.3912567866	0.2443955141
0.9808534328	0.4621000172	0 4220522014	0 2706222405
0.20000 0.9671149286	0.4621009172	0.4228523914	0.2706232485
0.25000	0 5429002426	0 4591076530	0 2022200060
0.9469879692	0.5438002436	0.4581976529	0.3022398960
	0.6196992894	A 4022122262	A 220E166226
0.30000 0.9086977269	0.0190992894	0.4932123362	0.3385166326
0.35000	0.6860987346	0.5298961211	0.3812581525
0.8689325583	0.0800987340	0.3298901211	0.3612361323
0.40000	0.7391995791	0.5636000164	0.4274662791
0.8271863613	0.7331333731	0.3030000104	0.42/4002/31
0.45000	0.7834007028	0.5938122218	0.4804228219
0.7776100411	0.7031007020	0.3330122210	0.4004220213
0.50000	0.8109996848	0.6077237432	0.5263355725
0.7191957314			
0.55000	0.8364991467	0.6239149352	0.5866065161
0.6666724507			
0.60000	0.8511993880	0.6228026748	0.6433087395
0.6043250043			
0.65000	0.8567996485	0.6008714103	0.6939108514
0.5306900341			
0.70000	0.8608993087	0.5805817165	0.7508777502
0.4737463127			
0.75000	0.8595988485	0.5407757560	0.8083735609
0.4069625195			
0.80000	0.8471991076	0.4362901545	0.8772331919
0.2906327723			
0.85000	0.8310996864	0.3112439357	0.9182463509
0.1875368732			
0.90000	0.8116997249	0.1480510594	0.9432839380
0.0805092834			
0.95000	0.7967999437	0.0048790276	0.5000000000
0.0024524264			

Out[295]: <matplotlib.legend.Legend at 0x1d8c551d4c0>

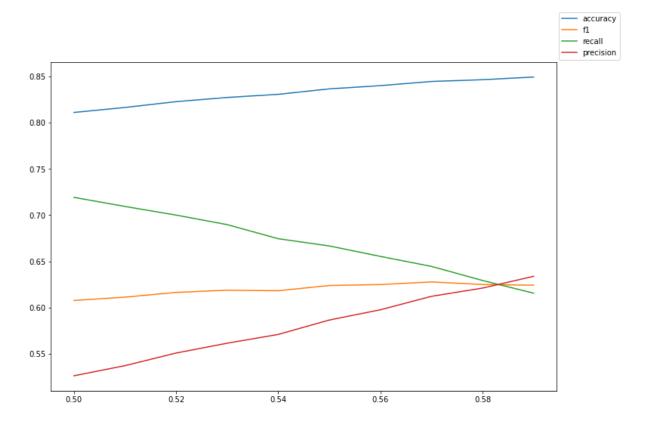


Do a more fine grained one with a smalle rinterval

```
In [297]: | acc_scores = []
          f1_scores = []
          recall_scores = []
          precision scores = []
          cutoffs = []
          for i in np.arange(0.50, 0.60, 0.01):
              accs, f1s, rec, prec = random_forest_alternate_cutoff(scaled_X.to_numpy(),
          y.to_numpy(), i)
              acc_scores.append(accs)
              f1_scores.append(f1s)
              recall_scores.append(rec)
              precision_scores.append(prec)
              cutoffs.append(i)
          for i, j, k, l, m in zip(cutoffs, acc_scores, f1_scores, prec_scores, rec_scor
          es):
              print("%10.5f %20.10f %20.10f %20.10f %20.10f" % (i, j, k, l, m))
          fig, ax = plt.subplots()
          ax.plot(cutoffs, acc_scores, label="accuracy")
          ax.plot(cutoffs, f1 scores, label="f1")
          ax.plot(cutoffs, recall_scores, label="recall")
          ax.plot(cutoffs, precision_scores, label="precision")
          fig.legend()
```

0.50000 1.0000000000	0.8109996848	0.6077237432	0.2036999563
0.51000 0.9990196078	0.8163995052	0.6113544138	0.2061432254
0.52000 0.52000 0.9921481867	0.8226993257	0.6164340182	0.2232327069
0.53000 0.9808534328	0.8271995661	0.6188813431	0.2443955141
0.54000	0.8305992463	0.6182805793	0.2706232485
0.9671149286 0.55000	0.8364991467	0.6239149352	0.3022398960
0.9469879692 0.56000	0.8399995871	0.6249520083	0.3385166326
0.9086977269 0.57000	0.8444997675	0.6277714194	0.3812581525
0.8689325583 0.58000	0.8463995676	0.6250138357	0.4274662791
0.8271863613 0.59000 0.7776100411	0.8492999479	0.6242401350	0.4804228219
0.7770100411			

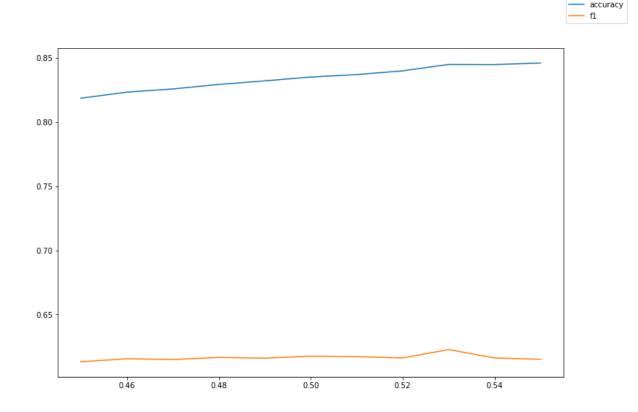
Out[297]: <matplotlib.legend.Legend at 0x1d92481f160>



Do a more fine-grained for a smaller interval, use SMOTE this time

```
In [343]:
          acc_scores = []
          f1_scores = []
           cutoffs = []
           for i in np.arange(0.45, 0.55, 0.01):
               accs, f1s = random_forest_alternate_cutoff(scaled_X.to_numpy(), y.to_numpy
           (), i, use_smote=True)
               acc scores.append(accs)
               f1 scores.append(f1s)
               cutoffs.append(i)
           for i, j, k in zip(cutoffs, acc_scores, f1_scores):
               print("%3.2f %10.5f %10.5f" % (i, j, k))
           fig, ax = plt.subplots()
           ax.plot(cutoffs, acc_scores, label="accuracy")
           ax.plot(cutoffs, f1_scores, label="f1")
           fig.legend()
          0.45
                  0.81860
                              0.61310
          0.46
                  0.82330
                              0.61544
          0.47
                  0.82580
                              0.61486
          0.48
                  0.82930
                              0.61650
          0.49
                  0.83210
                              0.61591
          0.50
                  0.83510
                              0.61746
          0.51
                  0.83700
                              0.61707
          0.52
                  0.83990
                              0.61611
          0.53
                  0.84490
                              0.62261
          0.54
                  0.84480
                              0.61607
          0.55
                  0.84600
                              0.61504
```

Out[343]: <matplotlib.legend.Legend at 0x1ec41eb7ee0>



## Classification with Other Samplers (over-samplers and combine-samplers)

```
In [117]: | samplers = [
               ('NONE', None),
               ('RandomOversampling', RandomOverSampler(sampling_strategy='minority')),
               ('RandomUnderSampling', RandomUnderSampler(sampling_strategy='majority')),
               ('smote', SMOTE(random_state=0)),
               ('adasyn', ADASYN(random_state=0)),
               ('smotenc', SMOTENC([False, False, True, True, True, True, False, True, Tr
          ue, True, True], random_state=0)),
               ('borderlinesmote', BorderlineSMOTE(random_state=0)),
               ('svmsmote', SVMSMOTE(random_state=0)),
               ('smoteenn', SMOTEENN(random_state=0)),
               ('smotetomek', SMOTETomek(random_state=0))
          1
          models = [
               ('<mark>GradientBoosting', lambda</mark>: GradientBoostingClassifier(random_state=0, ma
          x_depth=3, min_samples_split=42, n_estimators=100)),
               ('RandomForest', lambda: RandomForestClassifier(class_weight='balanced_sub
          sample', criterion='entropy', max_depth=None, min_samples_split=100, n_estimat
          ors=1000, random state=0)),
               ('SVC', lambda: SVC(class weight=None, gamma='auto', kernel='rbf', random
          state=0)),
          descriptions = []
          sorted_indices = None
          indices calculated = False
          comparison_string = ""
          plot data = []
          for a, b in models:
              fig1, ax1 = plt.subplots()
              accuracy_scores = []
              f1 scores = []
              descriptions = []
              print('-' * 80)
              print(f"ITERATING FOR {a.upper()}\n")
              for i, j in samplers:
                   accs, f1s = all_models_compare_samplers2(scaled_X.to_numpy(), y.to_num
          py(), j, i, b, a)
                  accuracy_scores.append(accs)
                  f1_scores.append(f1s)
                  descriptions.append(i)
              if False == indices_calculated:
                   sorted_indices = np.argsort(f1_scores)
                   indices calculated = True
              plot_data.append(([f1_scores[i] for i in sorted_indices], f"{a}-f1",) )
              plot_data.append(([accuracy_scores[i] for i in sorted_indices], f'{a}-accu
          racy',))
              ax1.plot([f1 scores[i] for i in np.argsort(f1 scores)], label=a)
              plt.xticks(np.arange(0, len(f1 scores), 1.0))
              ax1.set xticklabels([descriptions[i].upper() for i in np.argsort(f1 scores
          )], rotation=270)
              fig1.legend()
              fig1.suptitle(a)
              plt.show()
              sorted_order = np.argsort(f1_scores)
```

```
comparison_string += "EVALUATING MODEL %s\n" % a
   comparison_string += '----\n'
   for i in sorted_order:
       comparison_string += "%20s %10.5f %10.5f" % (descriptions[i], accuracy
_scores[i], f1_scores[i])
       comparison_string += "\n"
   comparison_string += "-----\n"
   print(comparison_string)
fig, ax = plt.subplots()
plt.xticks(np.arange(0, len(f1_scores), 1.0))
for array, lab in plot_data:
   ax.plot(array, label=lab)
ax.set_xticklabels([descriptions[i].upper() for i in sorted_indices], rotation
=270)
ax.legend()
fig.legend()
print(plot_data)
```

\_\_\_\_\_

-	

TTFRATTNG	FOR	GRADTENTROOSTING

ITERATING FOR	R GRADIENTBOC	STING			
Description:	NONE Model:	GradientR	oosting		
besch iption.	precision			support	
	precision	, ccarr	11 30010	заррог с	
0	0.88	0.97	0.92	7963	
1		0.47		2037	
accuracy			0.86	10000	
macro avg	0.83	0.72	0.75	10000	
weighted avg	0.86	0.86	0.85	10000	
Description:					5
	precision	recall	†1-score	support	
0	0.03	0.00	0.07	7063	
0 1	0.93 0.51		0.87		
1	0.51	0.74	0.60	2037	
accuracy			0.80	10000	
macro avg	a 72	0.78			
weighted avg				10000	
weighted avg	0.04	0.00	0.01	10000	
Description:	RandomUnderS	ampling M	odel: Grad:	ientBoostir	ng
·	precision				
0	0.93	0.80	0.86	7963	
1	0.49	0.76	0.60	2037	
accuracy			0.79		
macro avg		0.78			
weighted avg	0.84	0.79	0.81	10000	
Description:	smote Model:	Gradient	Boosting		
besci ipcion.			f1-score	support	
	precision	rccarr	11 30010	заррог с	
0	0.91	0.89	0.90	7963	
1	0.60	0.65	0.62	2037	
_					
accuracy			0.84	10000	
macro avg	0.76	0.77	0.76	10000	
weighted avg	0.85	0.84	0.84	10000	
Description:	-		_		
	precision	recall	f1-score	support	

0.91

0.60

0 1 0.89

0.65

0.90

0.63

7963

2037

accuracy				10000	
	0.76				
weighted avg	0.85	0.84	0.84	10000	
Description:	smotenc Mode	ol. Gradie	ntRoosting		
beset iperon.	precision		_	sunnort	
	precision	rccarr	11 30010	заррог с	
0	0.92	0.83	0.87	7963	
1			0.60		
_	0.32	0.72	0.00	2037	
accuracy			0.81	10000	
macro avg		0.77		10000	
weighted avg					
weighted avg	0.04	0.01	0.02	10000	
Description:	borderlinesm	ote Model	: Gradient	Boosting	
•	precision			_	
	•				
0	0.91	0.87	0.89	7963	
1			0.62		
accuracy			0.83	10000	
	0.74	0.77			
weighted avg					
0 0					
Description:			_		
	precision	recall	f1-score	support	
0		0.88		7963	
1	0.59	0.67	0.63	2037	
accuracy			0.84		
macro avg	0.75	0.78	0.76	10000	
weighted avg	0.85	0.84			
Description:	smoteenn Mod		_		
	precision	recall	f1-score	support	
0		0.79			
1	0.48	0.77	0.59	2037	
accuracy			0.78	10000	
macro avg	0.70	0.78	0.72	10000	
weighted avg	0.84	0.78	0.80	10000	

Description: smotetomek Model: GradientBoosting

precision recall f1-score support

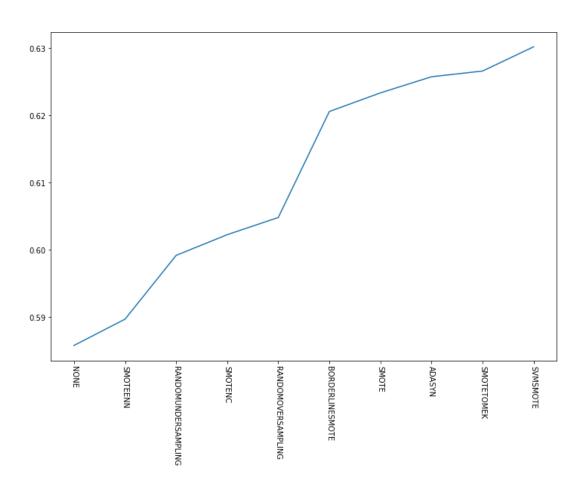
	0	0.91	0.89	0.90	7963
	1	0.60	0.65	0.63	2037
accurac	cv			0.84	10000
macro av	•	0.76	0.77		10000
weighted av	O	0.85	0.84		10000

-----

---



GradientBoosting



EVALUATING MO	DDEL Gradier	ntBoosting			
		0.86480	0.58578		
		0.78250	0.58970		
RandomUnderS		0.79290	0.59919		
<b>D</b> 1 0		0.80690			
RandomOvers		0.80190			
borderli		0.83090			
		0.84110			
	adasyn	0.84140	0.625//		
		0.84150	0.62662		
9	svmsmote	0.83900	0.63023		
ITERATING FOR	R RANDOMFORE	ST			
Description:	NONE Model:	RandomFor	est		
•			f1-score	support	
0	0.92	0.83	0.88	7963	
1	0.53	0.72	0.61	2037	
accuracy			0.0-	10000	
macro avg	0.72	0.78	0.74	10000	
weighted avg	0.84			10000	
	B 1 6	<b>.</b>		<u>.</u>	
Description:					
	precision	recall	f1-score	support	
0	0.00	0.04	A 00	7063	
0	0.92	0.84	0.88	7963	
1	0.54	0.71	0.61	2037	
0.0011100.001			0.00	10000	
accuracy	0.73	0.70	0.82	10000	
	0.73			10000	
weighted avg	0.84	0.82	0.82	10000	
Docenintian:	Dandami Inda	Sampline M	lodol. Pand	omEonos+	
Description:		. •			
	precision	recall	f1-score	support	
0	a 02	0.00	a 96	7062	
0	0.93 0.49			7963	
1	0.49	0.75	0.59	2037	
266112261			0.70	10000	
accuracy	0 71	0 77	0.79	10000	
macro avg				10000	
weighted avg	0.84	0.79	0.80	10000	

-----

Description: smote Model: RandomForest

precision recall f1-score support

0	0.91 0.59	0.88 0.65	0.90 0.62	7963 2037	
1	0.39	0.03	0.02	2037	
accuracy			0.84	10000	
macro avg	0.75	0.77	0.76	10000	
weighted avg	0.84	0.84	0.84	10000	
Description:	adasyn Model:				
	precision	recall	†1-score	support	
0	0.91	0.86	0.88	7963	
1		0.68			
accuracy			0.82		
macro avg					
weighted avg	0.84	0.82	0.83	10000	
 Decemberies	amakana Mada	1 . Danadam	Fanash		
pescription:	smotenc Model precision			sunnont	
	precision	recarr	11-30016	suppor c	
0	0.91	0.84	0.88	7963	
1	0.53	0.69	0.60	2037	
accuracy			0.81		
macro avg					
weighted avg	0.84	0.81	0.82	10000	
Description:	borderlinesmo	nte Model	· RandomFor	rest	
	precision				
0		0.85	0.88	7963	
1	0.54	0.70	0.61	2037	
accuracy			0.82	10000	
macro avg	0.73	a 77			
weighted avg	0.75				
weighted avg					
		0.82	0.83	10000	
Description:	0.84	0.82	0.83		
Description:	0.84 svmsmote Mode	0.82  el: Rando	0.83  mForest	10000	
Description:	0.84	0.82  el: Rando	0.83  mForest	10000	
Description:	0.84  svmsmote Mode	0.82  el: Rando	0.83  mForest	10000	
	0.84  svmsmote Mode precision  0.91	0.82  el: Rando recall	0.83  mForest f1-score	10000  support	
0	0.84  svmsmote Mode precision  0.91	0.82  el: Rando recall 0.87	0.83 mForest f1-score 0.89 0.62	10000 support 7963 2037	
0 1 accuracy	0.84  svmsmote Mode precision  0.91 0.58	0.82 el: Rando recall 0.87 0.68	0.83  mForest f1-score 0.89 0.62 0.83	10000 support 7963 2037 10000	
0 1 accuracy macro avg	0.84  svmsmote Mode precision  0.91 0.58	0.82 el: Rando recall 0.87 0.68	0.83  mForest f1-score 0.89 0.62 0.83 0.76	10000 support 7963 2037 10000 10000	
0 1 accuracy	0.84  svmsmote Mode precision  0.91 0.58	0.82 el: Rando recall 0.87 0.68	0.83  mForest f1-score 0.89 0.62 0.83	10000 support 7963 2037 10000	

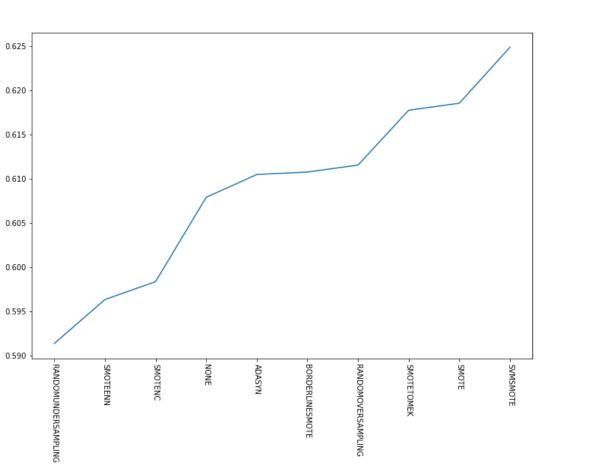
Description:	smoteenn Mod	lel: Rando	mForest		
	precision	recall	f1-score	support	
0	A 92	0.81	0.86	7963	
1	0.50	0.74	0.60	2037	
_	0.00	• • • • • • • • • • • • • • • • • • • •		_00,	
accuracy			0.80	10000	
macro avg	0.71	0.77	0.73	10000	
weighted avg	0.84	0.80	0.81	10000	
Description:	smotetomek M	lodel· Ran	domEorest		
beset ipeion.	precision			support	
	p. 002520		555. 5	опрро. с	
0	0.91	0.88	0.89	7963	
1	0.58	0.65	0.62	2037	
accuracy			0.83		
macro avg		0.77			
weighted avg	0.84	0.83	0.84	10000	

-----

---



RandomForest



	NONE	0.86480	0.58578	
CM	noteenn	0.78250	0.58970	
RandomUnderSa		0.78230	0.59919	
	smotenc	0.80690	0.60227	
RandomOversa		0.80190	0.60483	
borderlin		0.83090	0.62060	
55. 55. 22	smote	0.84110	0.62337	
		0.84140	0.62577	
	•	0.84150	0.62662	
sv	msmote		0.63023	
EVALUATING MOD		Forest		
RandomUnderSa	ampling	0.78840	0.59135	
sm	noteenn	0.79610	0.59632	
S	smotenc	0.81110	0.59834	
		0.81100	0.60788	
	•	0.82210	0.61047	
borderlin		0.81950		
RandomOversa		0.81540	0.0225	
smot			0.61774	
			0.61853	
SV	/msmote	0.83310	0.62486	
 ITERATING FOR	SVC			
 ITERATING FOR	SVC			
		: SVC		
Description: N	NONE Model	: SVC recall	f1-score	support
Description: N	NONE Model		f1-score	support
Description: N	NONE Model		f1-score 0.92	support 7963
Description: N	NONE Model precision	recall 0.98		
Description: N	NONE Model precision 0.86	recall 0.98	0.92 0.52	7963
Description: N	NONE Model precision 0.86	recall 0.98	0.92	7963
Description: N 0 1	NONE Model precision 0.86	recall 0.98 0.38	0.92 0.52	7963 2037
Description: N  0 1  accuracy macro avg	NONE Model precision 0.86 0.82	recall 0.98 0.38 0.68	0.92 0.52 0.86	7963 2037 10000
Description: N  0 1  accuracy macro avg	NONE Model precision 0.86 0.82	recall 0.98 0.38 0.68	0.92 0.52 0.86 0.72	7963 2037 10000 10000
Description: N  0 1  accuracy macro avg weighted avg	NONE Model precision 0.86 0.82 0.84 0.85	recall 0.98 0.38 0.68 0.86	0.92 0.52 0.86 0.72 0.84	7963 2037 10000 10000
Description: N  0 1  accuracy macro avg weighted avg	NONE Model precision 0.86 0.82 0.84 0.85	recall 0.98 0.38 0.68 0.86	0.92 0.52 0.86 0.72 0.84	7963 2037 10000 10000
accuracy macro avg weighted avg Description: R	NONE Model precision 0.86 0.82 0.84 0.85	recall 0.98 0.38 0.68 0.86	0.92 0.52 0.86 0.72 0.84	7963 2037 10000 10000 10000
Description: N  0 1  accuracy macro avg weighted avg  Description: R	NONE Model precision  0.86 0.82  0.84 0.85  RandomOvers	recall 0.98 0.38 0.68 0.86  sampling Mo	0.92 0.52 0.86 0.72 0.84 	7963 2037 10000 10000 10000 support
Description: N  0 1  accuracy macro avg weighted avg  Description: R	NONE Model precision  0.86 0.82  0.84 0.85  RandomOvers precision  0.92	recall 0.98 0.38 0.68 0.86sampling Morecall 0.81	0.92 0.52 0.86 0.72 0.84 	7963 2037 10000 10000 10000 support
Description: N  0 1  accuracy macro avg weighted avg  Description: R	NONE Model precision  0.86 0.82  0.84 0.85  RandomOvers	recall 0.98 0.38 0.68 0.86sampling Morecall 0.81	0.92 0.52 0.86 0.72 0.84 	7963 2037 10000 10000 10000 support
Description: N  0 1  accuracy macro avg weighted avg  Description: R  0 1	NONE Model precision  0.86 0.82  0.84 0.85  RandomOvers precision  0.92	recall 0.98 0.38 0.68 0.86sampling Morecall 0.81	0.92 0.52 0.86 0.72 0.84 	7963 2037 10000 10000 10000 support 7963 2037
Description: N  0 1  accuracy macro avg weighted avg  Description: R  0 1  accuracy	NONE Model precision  0.86 0.82  0.84 0.85  RandomOvers precision  0.92 0.50	recall  0.98  0.38  0.68  0.86	0.92 0.52 0.86 0.72 0.84 	7963 2037 10000 10000 10000 support 7963 2037
Description: N  0 1  accuracy macro avg weighted avg  Description: R  0 1	NONE Model precision  0.86 0.82  0.84 0.85  RandomOvers precision  0.92	recall  0.98  0.38  0.68  0.86  sampling Morecall  0.81  0.74	0.92 0.52 0.86 0.72 0.84 	7963 2037 10000 10000 10000 support 7963 2037

Description: RandomUnderSampling Model: SVC

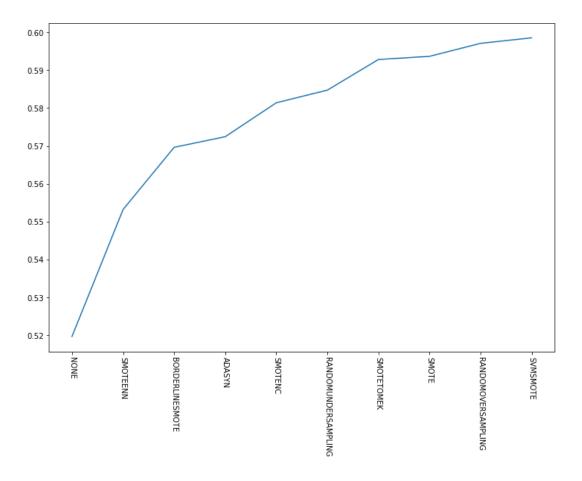
precision recall f1-score support

	0.00		0.05	70.60	
0 1	0.92 0.48	0.80 0.74	0.86 0.58	7963 2037	
1	0.40	0.74	0.38	2037	
accuracy			0.79	10000	
macro avg	0.70	0.77			
weighted avg					
 Desembles	Madal.	CVC			
Description:	<pre>smote Model:   precision</pre>		f1-scono	cunnont	
	precision	recarr	11-30016	Suppor c	
0	0.92	0.82	0.87	7963	
1		0.72			
accuracy			0.80		
macro avg					
weighted avg	0.84	0.80	0.81	10000	
Description:	adasyn Model:	SVC			
Desci Iperon.	precision		f1-score	support	
	p. 002520		555. 5	Juppo. C	
0	0.93	0.77	0.84	7963	
1	0.46	0.77	0.57	2037	
accuracy			0.77		
macro avg					
weighted avg	0.83	0.77	0.79	10000	
Description:	smotenc Model	: SVC			
F ·			f1-score	support	
	·				
0	0.91	0.83	0.87	7963	
1	0.50	0.69	0.58	2037	
accuracy	A =-	a = -	0.80		
macro avg					
weighted avg	0.83	0.80	0.81	10000	
	·	· – – <b>–</b>	<b>_</b>	<b>-</b>	
Description:	borderlinesmo	te Model	: SVC		
,	precision			support	
0			0.84		
1	0.45	0.77	0.57	2037	
			<b>A</b> = -		
accuracy	A	a ==	0.76		
macro avg					
weighted avg	0.83	0.76	0.78	10000	
		. <b></b> =			

_	-	_	

Description:	svmsmote Mo	odel: SVC			
	precision	recall	f1-score	support	
0	0.92	0.81	0.87	7963	
1	0.50	0.74	0.60	2037	
accuracy			0.80	10000	
macro avg		0.78	0.73	10000	
weighted avg	0.84	0.80	0.81	10000	
8					
Description:	smoteenn Mo	odel: SVC			
2000. 1pc1011.			f1-score	support	
	precision	rccair	11 30010	заррог с	
0	0.93	Q 72	0 81	7963	
1			0.55		
_	0.42	0.00	0.55	2037	
accuracy			0.74	10000	
macro avg		0.76			
weighted avg	0.83	0.74	0.76	10000	
December		Madal. CVC			
Description:					
	precision	recall	f1-score	support	
_	2 22	2 22	2 2=	=0.55	
0			0.87		
1	0.50	0.72	0.59	2037	
accuracy			0.80		
macro avg					
weighted avg	0.84	0.80	0.81	10000	

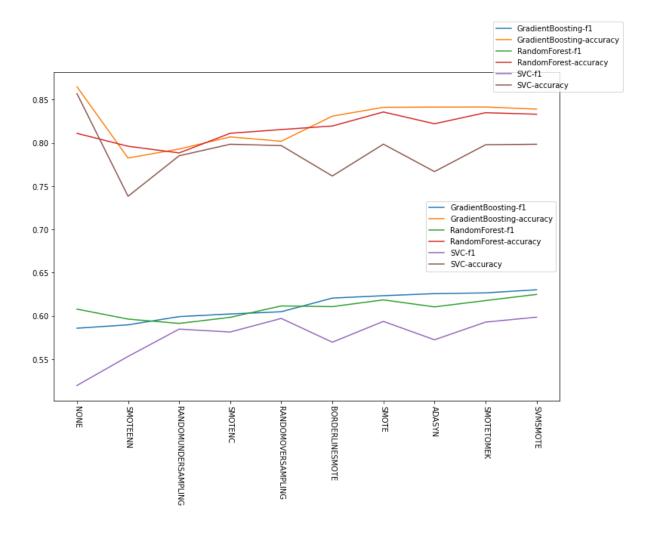
SVC — SVC



## **EVALUATING MODEL GradientBoosting**

EVALUATING MODEL Grad	iientBoosting	
NONE smoteenn RandomUnderSampling smotenc RandomOversampling borderlinesmote smote adasyn smotetomek svmsmote	0.86480 0.78250 0.79290 0.80690 0.80190 0.83090 0.84110 0.84140 0.84150 0.83900	0.58578 0.58970 0.59919 0.60227 0.60483 0.62060 0.62337 0.62577 0.62662 0.63023
EVALUATING MODEL Rand	lomForest	
RandomUnderSampling smoteenn smotenc NONE adasyn borderlinesmote RandomOversampling smotetomek smote svmsmote	0.78840 0.79610 0.81110 0.81100 0.82210 0.81950 0.81540 0.83490 0.83570 0.83310	0.59135 0.59632 0.59834 0.60788 0.61047 0.61074 0.61153 0.61774 0.61853 0.62486
EVALUATING MODEL SVC		
NONE smoteenn borderlinesmote adasyn smotenc RandomUnderSampling smotetomek smote RandomOversampling svmsmote	0.85690 0.73830 0.76170 0.76680 0.79840 0.78510 0.79780 0.79850 0.79690 0.79840	0.51964 0.55318 0.56962 0.57242 0.58140 0.58473 0.59283 0.59367 0.59710 0.59857

[([0.5857843137254902, 0.5897000565930957, 0.5991871492161795, 0.602265705458 2904, 0.6048274486335529, 0.6205968140004489, 0.6233704669352926, 0.625766871 1656442, 0.6266195524146054, 0.630225080385852], 'GradientBoosting-f1'), ([0.8648, 0.7825, 0.7929, 0.8069, 0.8019, 0.8309, 0.8411, 0.8414, 0.8415, 0.839], 'GradientBoosting-accuracy'), ([0.6078838174273858, 0.5963175608790339, 0.591 3480108149866, 0.5983414841590474, 0.6115319865319865, 0.6107397023937892, 0.6185279777107034, 0.6104663893146486, 0.6177355869414217, 0.624859518993032], 'RandomForest-f1'), ([0.811, 0.7961, 0.7884, 0.8111, 0.8154, 0.8195, 0.8357, 0.8221, 0.8349, 0.8331], 'RandomForest-accuracy'), ([0.5196374622356495, 0.55 31842240054635, 0.5847342995169083, 0.5813953488372092, 0.5971037492560999, 0.5696225392811992, 0.593668078241581, 0.5724239090575723, 0.592831252517116 4, 0.5985663082437276], 'SVC-f1'), ([0.8569, 0.7383, 0.7851, 0.7984, 0.7969, 0.7617, 0.7985, 0.7668, 0.7978, 0.7984], 'SVC-accuracy')]



## **Classifiers Samplers and Feature-Selectors compared**

```
In [100]: def model_compare_3(X, y,\
                               sampler, description,\
                               model_gen_fn, model_description,\
                               feature_selector_fn, feature_selector_description):
              p = []
              if None != sampler:
                   p.append(('sampler', sampler,))
              if None != feature_selector_fn:
                   p.append(('fs', feature_selector_fn(),))
              p.append(('model', model_gen_fn(),))
              p = imblearn.pipeline.Pipeline(p)
              y_pred = cross_val_predict(p, X, y, \
                                          cv=model_selection.StratifiedKFold(n_splits=6,
          shuffle=True, random_state=2),\
                                          n_jobs=7)
              print(f"Description: {description} Model: {model_description}\n{metrics.cl
          assification_report(y, y_pred)}")
              print('-' * 80)
              return metrics.accuracy_score(y, y_pred), metrics.f1_score(y, y_pred)
```

```
In [244]:
          selectors = [
               ('Variance Threshold 0.8', lambda: VarianceThreshold(threshold=(.8 * (1 -
           .8)))),
               ('RFECV RandomForest', lambda: RFECV(estimator=RandomForestClassifier())),
               ('FWE f_classif', lambda: SelectFwe(f_classif, alpha=0.01)),
               ('NONE', None),
          1
          samplers = [
               ('smote', SMOTE(random_state=0)),
               ('adasyn', ADASYN(random_state=0)),
              ('svmsmote', SVMSMOTE(random_state=0)),
              ('smotetomek', SMOTETomek(random state=0)),
               ('RandomOversampling', RandomOverSampler(sampling strategy='minority')),
               ('NONE', None)
          1
          models = [
               ('GradientBoosting', lambda: GradientBoostingClassifier(random state=0, ma
          x depth=3, min samples split=42, n estimators=100)),
               ('RandomForest', lambda: RandomForestClassifier(class_weight='balanced_sub
          sample', criterion='entropy', max depth=None, min samples split=100, n estimat
          ors=1000, random state=0)),
               ('SVC', lambda: SVC(class_weight=None, gamma='auto', kernel='rbf', random_
          state=0)),
          all descs = ""
          all results = []
          for model_desc, model_fn in models:
              for sampler_desc, sampler in samplers:
                   for selector desc, selector in selectors:
                       print('-'*80)
                       print("EVALUATING: ", model_desc, sampler_desc, selector_desc)
                       acc, f1 = model_compare_3(scaled_X.to_numpy(), y.to_numpy(),\
                                                sampler, sampler_desc,\
                                                model_fn, model_desc,\
                                                selector, selector desc)
                       print("%20s %30s %30s %20.10f %20.10f" % (model_desc, sampler_desc
           , selector_desc, acc, f1))
                       all descs += "%20s %30s %30s %20.10f %20.10f" % (model desc, sampl
          er desc, selector desc, acc, f1)
                       all descs += '\n'
                       all results.append((model desc, sampler desc, selector desc, acc,
          f1,))
          print(all_descs)
```

 FVALUATTNG:	GradientBoost	ing smot	e Variance	Threshold	0.8
	smote Model:	•		TIII CSIIOIG	•••
	precision	recall	f1-score	support	
0	0.91	0.89	0.90	7963	
1	0.60	0.65	0.62	2037	
accuracy			0.84	10000	
macro avg	0.76	0.77	0.76	10000	
weighted avg	0.85	0.84	0.84	10000	
Gradient					Variance Threshol
d 0.8	0.8411000000		0.6233704	669 	
	GradientBoost			ndomForest	
Description:	<pre>smote Model:   precision</pre>		_	sunnont	
	pi ecision	recarr	11-30016	suppor c	
0			0.90		
1	0.61	0.65	0.63	2037	
accuracy			0.84	10000	
-	0.76	0.77	0.76	10000	
	0.85				
Gradient	_		0 6260150	smote	RFECV RandomF
orest	0.8433000000 		0.6268159	085 	
	<pre>GradientBoost smote Model:</pre>	_	_	assif	
besci ipcion.	precision		_	support	
	•				
0			0.89		
1	0.58	0.67	0.62	2037	
accuracy			0.83	10000	
macro avg	0.75	0.77	0.76	10000	
weighted avg	0.84	0.83	0.84	10000	
Gradient			0 622777	smote	FWE f_cl
assif	0.8340000000 		0.6227272	/ <i>L</i> / 	
	<pre>GradientBoost smote Model:</pre>	•			

Description: smote Model: GradientBoosting precision recall f1-score support

0 1		0.89 0.65	0.90 0.62	7963 2037	
accuracy macro avg	0.76			10000	
weighted avg	0.85	0.84	0.84	10000	
Gradient				smote	
	0.8411000000 				
	GradientBoost adasyn Model:			• Threshold	0.8
	precision		_	support	
0 1			0.90 0.63		
accuracy			0.84	10000	
macro avg	0.76	0.77	0.76	10000	
weighted avg	0.85	0.84	0.84	10000	
Gradient d 0.8	Boosting 0.8414000000		0.62576687	-	Variance Threshol
	GradientBoost			andomForest	
	adasyn Model:	Gradien	tBoosting		
		Gradien	tBoosting		
Description:	<pre>adasyn Model:   precision   0.91</pre>	Gradien recall 0.89	tBoosting f1-score 0.90	support 7963	
Description:	adasyn Model: precision	Gradien recall	tBoosting f1-score	support	
Description: 0 1 accuracy	adasyn Model: precision 0.91 0.60	Gradien recall 0.89 0.65	tBoosting f1-score 0.90 0.63 0.84	support 7963 2037 10000	
Description:  0 1 accuracy macro avg	adasyn Model: precision 0.91 0.60	Gradien recall  0.89  0.65	tBoosting f1-score 0.90 0.63 0.84 0.76	support 7963 2037 10000 10000	
Description: 0 1 accuracy	adasyn Model: precision 0.91 0.60	Gradien recall 0.89 0.65	tBoosting f1-score 0.90 0.63 0.84 0.76	support 7963 2037 10000 10000	
Description:  0 1 accuracy macro avg	adasyn Model: precision 0.91 0.60	Gradien recall  0.89  0.65	tBoosting f1-score 0.90 0.63 0.84 0.76	support 7963 2037 10000 10000	
Description:  0 1 accuracy macro avg	adasyn Model:     precision	Gradien recall 0.89 0.65 0.77 0.84	tBoosting f1-score 0.90 0.63 0.84 0.76 0.84	support 7963 2037 10000 10000 10000	RFECV RandomF
Description:  0 1 accuracy macro avg weighted avg  Gradient orest	adasyn Model:     precision	Gradien recall  0.89 0.65  0.77 0.84	1Boosting f1-score 0.90 0.63 0.84 0.76 0.84	support  7963 2037  10000 10000 10000  adasyn 943	RFECV RandomF
Description:  0 1 accuracy macro avg weighted avg  Gradient orest EVALUATING:	adasyn Model:     precision	Gradien recall	tBoosting f1-score 0.90 0.63 0.84 0.76 0.84	support  7963 2037  10000 10000 10000  adasyn 943	RFECV RandomF
Description:  0 1 accuracy macro avg weighted avg  Gradient orest EVALUATING:	adasyn Model:     precision	Gradien recall  0.89 0.65  0.77 0.84  ing adas Gradien	### The strain of the strain o	support  7963 2037  10000 10000  adasyn 943	RFECV RandomF
Description:  0 1 accuracy macro avg weighted avg  Gradient orest EVALUATING:	adasyn Model: precision  0.91 0.60  0.76 0.85  Boosting 0.8417000000  GradientBoost adasyn Model: precision	Gradien recall  0.89 0.65  0.77 0.84 ing adas Gradien recall	### The strain of the strain o	support  7963 2037  10000 10000 10000  adasyn 943  Lassif  support	RFECV RandomF
Description:  0 1 accuracy macro avg weighted avg  Gradient orest EVALUATING: Description:	adasyn Model: precision  0.91 0.60  0.76 0.85   Boosting 0.8417000000 GradientBoost adasyn Model: precision  0.92	Gradien recall  0.89 0.65  0.77 0.84 ing adas Gradien recall	### The strain of the strain o	support  7963 2037  10000 10000 10000  adasyn 043 Lassif support 7963	RFECV RandomF
Description:  0 1 accuracy macro avg weighted avg  Gradient orest EVALUATING: Description:	adasyn Model: precision  0.91 0.60  0.76 0.85  Boosting 0.8417000000  GradientBoost adasyn Model: precision  0.92 0.55	Gradien recall  0.89 0.65  0.77 0.84  ing adas Gradien recall  0.86	### The strain of the strain o	support  7963 2037  10000 10000  adasyn 943  Lassif  support  7963 2037	RFECV RandomF
Description:  0 1 accuracy macro avg weighted avg  Gradient orest EVALUATING: Description:	adasyn Model: precision  0.91 0.60  0.76 0.85   Boosting 0.8417000000  GradientBoost adasyn Model: precision  0.92 0.55	Gradien recall  0.89 0.65  0.77 0.84  ing adas Gradien recall  0.86	### TBOOSTING ### 1-score  ### 0.90 ### 0.63  ### 0.76 ### 0.84  ### 0.76 ### 0.84  ### 0.62638659  ### 0.62638659  ### 1-score  ### 0.89 ### 0.62	support  7963 2037  10000 10000 10000  adasyn 943  Lassif  support  7963 2037 10000	RFECV RandomF

 Gradient	Boosting			adasyn	FWE f_cl
assif	0.8239000000		0.61692408	309	
 EVALUATING:	GradientBoost	ing adas	vn NONE		
	adasyn Model:				
- года: -р с-о	precision			support	
	•				
0			0.90	7963	
1	0.60	0.65	0.63	2037	
			0.04	10000	
accuracy	0.76	0.77	0.84 0.76	10000 10000	
macro avg weighted avg		0.77			
weighted avg	0.05	0.04	0.04	10000	
Gradient				adasyn	
NONE	0.8414000000		0.625766873	12	
EVALUATING:	GradientBoost	ing svms	smote Varia	nce Threshol	d 0.8
	svmsmote Mode	_			
·	precision				
0			0.90		
1	0.59	0.67	0.63	2037	
			0.04	10000	
accuracy	0.75	0.70	0.84 0.76		
macro avg weighted avg		0.78 0.84	0.76	10000 10000	
weighted avg	0.05	0.04	0.04	10000	
Gradient	_				Variance Threshol
	0.8390000000				
 EVALUATING:	GradientBoost	ing cymc	moto PEECV	PandomEonos:	+
	svmsmote Model	_			L
beschiption.	precision				
	precision	rccarr	11 30010	заррог с	
0	0.91	0.88	0.90	7963	
1			0.63		
accuracy				10000	
macro avg					
weighted avg	0.85	0.84	0.84	10000	
Gradient	Roosting		c v	/msmote	RFECV RandomF
orest	0.8383000000		_		AT LCV TATIOUTI

_	-	-	

Description:	GradientBoost svmsmote Mode precision	l: Gradi	entBoosting.		
	•				
0			0.89		
1	0.57	0.70	0.63	2037	
accuracy			0.83	10000	
macro avg	0.75	0.78		10000	
weighted avg					
 Gradient	Boosting		SV	msmote	FWE f_cl
	0.8324000000				· WE 1_C1
		_			
	GradientBoost				
Description:	svmsmote Mode precision		_		
	precision	rccarr	11 30010	заррог с	
0	0.91	0.88	0.90	7963	
1	0.59	0.67	0.63	2037	
			0.04	10000	
accuracy macro avg		0.79		10000 10000	
weighted avg					
mergineed and	0.03	0.01	0.0.	2000	
 Coadiant					
( <sub>1</sub> ranianti				4	
	Boosting a 8390000000			msmote 4	
	300sting 0.8390000000 		sv 0.630225080		
NONE	0.8390000000		0.630225080	4	
NONE EVALUATING:	0.8390000000  GradientBoost	ing smot	0.630225080  etomek Vari	4  ance Thresh	old 0.8
NONE EVALUATING:	0.8390000000 GradientBoost smotetomek Mo	ing smot	0.630225080 etomek Vari	4 ance Thresh	old 0.8
NONE EVALUATING:	0.8390000000  GradientBoost	ing smot	0.630225080 etomek Vari	4 ance Thresh	old 0.8
NONE EVALUATING:	0.8390000000 GradientBoost smotetomek Mo precision	ing smot del: Gra recall	0.630225080 etomek Vari	4 ance Thresh ng support	old 0.8
NONE EVALUATING: Description:	0.8390000000 GradientBoost smotetomek Mo precision 0.91	ing smot del: Gra recall 0.89	0.630225080 etomek Vari dientBoosti f1-score	4 ance Thresh ng support 7963	old 0.8
NONE EVALUATING: Description:	0.8390000000 GradientBoost smotetomek Mo precision 0.91 0.60	ing smot del: Gra recall 0.89	0.630225080 setomek Vari dientBoosti f1-score 0.90 0.63	4 ance Thresh ng support 7963 2037	old 0.8
NONE EVALUATING: Description:  0 1 accuracy	0.8390000000 GradientBoost smotetomek Mo precision 0.91 0.60	ing smot del: Gra recall 0.89 0.65	0.630225080 eetomek Vari dientBoosti f1-score 0.90 0.63	4 ance Thresh ng support 7963 2037	old 0.8
NONE EVALUATING: Description:  0 1 accuracy macro avg	0.8390000000 GradientBoost smotetomek Mo precision 0.91 0.60	ing smot del: Gra recall 0.89 0.65	0.630225080 eetomek Vari dientBoosti f1-score 0.90 0.63 0.84 0.76	4 ance Thresh ng support 7963 2037 10000 10000	old 0.8
NONE EVALUATING: Description:  0 1 accuracy macro avg	0.8390000000 GradientBoost smotetomek Mo precision 0.91 0.60	ing smot del: Gra recall 0.89 0.65	0.630225080 eetomek Vari dientBoosti f1-score 0.90 0.63 0.84 0.76	4 ance Thresh ng support 7963 2037 10000 10000	old 0.8
NONE EVALUATING: Description:  0 1 accuracy macro avg	0.8390000000 GradientBoost smotetomek Mo precision 0.91 0.60	ing smot del: Gra recall 0.89 0.65	0.630225080 eetomek Vari dientBoosti f1-score 0.90 0.63 0.84 0.76	4 ance Thresh ng support 7963 2037 10000 10000	old 0.8
NONE EVALUATING: Description:  0 1 accuracy macro avg weighted avg	0.8390000000 GradientBoost smotetomek Mo precision 0.91 0.60 0.76 0.85	ing smot del: Gra recall 0.89 0.65	0.630225080 eetomek Vari dientBoosti f1-score 0.90 0.63 0.84 0.76 0.84	4 ance Thresh ng support 7963 2037 10000 10000	
NONE EVALUATING: Description:  0 1 accuracy macro avg weighted avg GradientI	0.8390000000 GradientBoost smotetomek Mo precision 0.91 0.60 0.76 0.85	ing smot del: Gra recall 0.89 0.65 0.77 0.84	0.630225080 eetomek Vari dientBoosti f1-score  0.90 0.63  0.84 0.76 0.84	4 ance Thresh ng support 7963 2037 10000 10000 10000	old 0.8  Variance Threshol
NONE	0.8390000000 GradientBoost smotetomek Mo precision 0.91 0.60 0.76 0.85	ing smot del: Gra recall 0.89 0.65 0.77 0.84	0.630225080 etomek VaridientBoostif1-score 0.90 0.63 0.84 0.76 0.84	4 ance Thresh ng support 7963 2037 10000 10000 10000	
NONE	0.8390000000 GradientBoost smotetomek Mo precision 0.91 0.60 0.76 0.85	ing smot del: Gra recall 0.89 0.65 0.77 0.84	0.630225080 etomek VaridientBoostif1-score 0.90 0.63 0.84 0.76 0.84	4 ance Thresh ng support 7963 2037 10000 10000 10000	
NONE	0.8390000000 GradientBoost smotetomek Mo precision 0.91 0.60 0.76 0.85	ing smot del: Gra recall 0.89 0.65 0.77 0.84	0.630225080 eetomek Vari dientBoosti f1-score 0.90 0.63 0.84 0.76 0.84 smot 0.62661955	4 ance Thresh ng support 7963 2037 10000 10000 etomek 24	Variance Threshol
NONE	O.8390000000 GradientBoost smotetomek Mo precision O.91 O.60 O.76 O.85  Boosting O.8415000000 GradientBoost smotetomek Mo	ing smot del: Gra recall 0.89 0.65 0.77 0.84	etomek VaridientBoostif1-score  0.90 0.63  0.84 0.76 0.84  smot 0.62661955 eetomek RFEC	4 ance Thresh ng support 7963 2037 10000 10000 etomek 24 V RandomFor ng	Variance Threshol
NONE	0.8390000000 GradientBoost smotetomek Mo precision 0.91 0.60 0.76 0.85  Boosting 0.8415000000 GradientBoost	ing smot del: Gra recall 0.89 0.65 0.77 0.84	etomek VaridientBoostif1-score  0.90 0.63  0.84 0.76 0.84  smot 0.62661955 eetomek RFEC	4 ance Thresh ng support 7963 2037 10000 10000 etomek 24 V RandomFor ng	Variance Threshol

1	0.61	0.65	0.63	2037	
266110261			0.94	10000	
accuracy macro avg		0 77			
weighted avg		0.77		10000	
weighted avg	0.05	0.04	0.85	10000	
Gradient				tetomek	RFECV RandomF
	0.8432000000				
EVALUATING:	GradientBoosti	ing smot	etomek FWF	f classif	
	smotetomek Mod	_			
2000: =p 0=0v	precision			_	
	,				
0	0.91	0.88	0.89	7963	
1	0.58	0.67	0.63	2037	
accuracy			0.84	10000	
macro avg		0.77	0.76	10000	
weighted avg	0.85	0.84	0.84	10000	
Gradient	Boosting		smo <sup>-</sup>	tetomek	FWE f_cl
	0.8358000000				
	GradientBoosti				
Description:	smotetomek Mod				
	precision	recall	f1-score	support	
0	0.91	0.89	0.90	7963	
1			0.63		
_	0.00	0.05	0.05	2037	
accuracy			0.84	10000	
macro avg		0.77	0.76	10000	
weighted avg	0.85	0.84		10000	
 C	D+:				
Gradient	_			tetomek	
NONE	0.8415000000			2 <del>4</del> 	
EVALUATING:	GradientBoosti	ing Rand	lomOversamp	ling Varian	ce Threshold 0.8
	RandomOversamp	_		_	
·	precision	_		_	
0	0.93	0.82	0.87	7963	
1	0.51	0.75	0.61	2037	
accuracy		0.70	0.80	10000	
macro avg			0.74		
weighted avg	0.84	0.80	0.81	10000	

	Roosting	D ~	ndom0vonc-	molina	Vanianco Throchal
	GradientBoosting 0.8 0.8023000000		0.60656716	variance inresnoi	
 EVALUATING:	GradientBoost:	ing Rando	omOversamp]	ing RFECV Ra	ındomForest
	RandomOversam				
•	precision				
0			0.87		
1	0.51	0.74	0.61	2037	
266118261			0 80	10000	
accuracy macro avg		0 78			
weighted avg					
icignica avg	0.04	0.00	0.02	10000	
·					
	Boosting				RFECV RandomF
orest	0.8043000000		0.60773702	214	
VALUATING:	GradientBoost:	ing Rando	mOversamp]	ing FWF f cl	assif
	RandomOversam	•	•	_	
<b>F</b>	precision				
	,				
0	0.93	0.81	0.87	7963	
1	0.51	0.75	0.61	2037	
_			0.00	2037	
_					
accuracy			0.80	10000	
accuracy macro avg	0.72	0.78	0.80 0.74	10000 10000	
accuracy macro avg	0.72		0.80 0.74	10000	
accuracy macro avg	0.72	0.78	0.80 0.74	10000 10000	
accuracy macro avg weighted avg	0.72 0.84	0.78 0.80	0.80 0.74 0.81	10000 10000 10000	
accuracy macro avg veighted avg  Gradient	0.72 0.84 Boosting	0.78 0.80 	0.80 0.74 0.81	10000 10000 10000	FWE <b>f_c</b> ]
accuracy macro avg weighted avg  Gradient	0.72 0.84 Boosting 0.8011000000	0.78 0.80 	0.80 0.74 0.81  andomOversa 0.60637245	10000 10000 10000 mmpling	FWE f_c]
accuracy macro avg veighted avg  Gradient	0.72 0.84 Boosting	0.78 0.80 	0.80 0.74 0.81	10000 10000 10000 mmpling	FWE f_c
accuracy macro avg weighted avg Gradient assif	0.72 0.84 Boosting 0.8011000000	0.78 0.80  Ra	0.80 0.74 0.81  andomOversa 0.60637245	10000 10000 10000 	FWE f_c]
accuracy macro avg weighted avg  Gradient assif Gradient	0.72 0.84 Boosting 0.8011000000 GradientBoost	0.78 0.80  Ra	0.80 0.74 0.81 	10000 10000 10000  ampling 520 	FWE f_c]
accuracy macro avg weighted avg  Gradient assif Gradient	0.72 0.84 Boosting 0.8011000000	0.78 0.80 Ra ing Rando	0.80 0.74 0.81 andomOversa 0.60637245 omOversampl	10000 10000 10000 impling 220 ing NONE	FWE f_c
accuracy macro avg weighted avg  Gradient assif Gradient	0.72 0.84  Boosting 0.8011000000  GradientBoost: RandomOversam	0.78 0.80 Ra ing Rando	0.80 0.74 0.81 andomOversa 0.60637245 omOversampl	10000 10000 10000 impling 220 ing NONE	FWE f_c
accuracy macro avg weighted avg  Gradient assif Gradient	0.72 0.84 Boosting 0.8011000000 GradientBoost: RandomOversam	0.78 0.80 Ra ing Rando pling Mod recall	0.80 0.74 0.81 andomOversa 0.60637245 omOversampl	10000 10000 10000  mmpling 520  ing NONE entBoosting support	FWE f_c]
accuracy macro avg weighted avg Gradient assif Gradient SVALUATING:	0.72 0.84  Boosting 0.8011000000  GradientBoost: RandomOversam  precision 0.93	0.78 0.80  Ra  ing Rando oling Mod recall 0.82	0.80 0.74 0.81 andomOversa 0.60637245 omOversamplel: Gradie f1-score	10000 10000 10000  mmpling 520  ing NONE entBoosting support 7963	FWE f_c
accuracy macro avg weighted avg Gradient assif EVALUATING: Description:	0.72 0.84 Boosting 0.8011000000 GradientBoost: RandomOversamp precision 0.93 0.51	0.78 0.80  Ra  ing Rando oling Mod recall 0.82	0.80 0.74 0.81 andomOversa 0.60637245  omOversamplel: Gradie f1-score 0.87 0.61	10000 10000 10000 impling 520 ing NONE entBoosting support 7963 2037	FWE f_c
accuracy macro avg weighted avg Gradient assif EVALUATING: Description:	0.72 0.84  Boosting 0.8011000000  GradientBoost: RandomOversample precision 0.93 0.51	0.78 0.80 Rading Rando pling Mod recall 0.82 0.75	0.80 0.74 0.81 andomOversa 0.60637245 0.60637245 comOversamplel: Gradie f1-score 0.87 0.61	10000 10000 10000 10000 impling 520 ing NONE entBoosting support 7963 2037	FWE f_c]
accuracy macro avg weighted avg  Gradient assif CVALUATING: Description:	0.72 0.84 Boosting 0.8011000000 GradientBoost: RandomOversamp precision 0.93 0.51	0.78 0.80 Rando ling Rando pling Mod recall 0.82 0.75	0.80 0.74 0.81 0.60637245 0.60637245 0.60637245 0.60637245 0.60637245 0.60637245 0.60637245 0.60637245 0.60637245 0.60637245 0.60637245	10000 10000 10000 10000 impling 520 ing NONE entBoosting support 7963 2037 10000 10000	FWE f_c
accuracy macro avg weighted avg  Gradient assif CVALUATING: Description:	0.72 0.84 Boosting 0.8011000000 GradientBoost: RandomOversamp precision 0.93 0.51	0.78 0.80 Rando ling Rando pling Mod recall 0.82 0.75	0.80 0.74 0.81 0.60637245 0.60637245 0.60637245 0.60637245 0.60637245 0.60637245 0.60637245 0.60637245 0.60637245 0.60637245 0.60637245	10000 10000 10000 10000 impling 520 ing NONE entBoosting support 7963 2037 10000 10000	FWE f_c
accuracy macro avg weighted avg Gradient assif EVALUATING: Description:  0 1 accuracy macro avg	0.72 0.84 Boosting 0.8011000000 GradientBoost: RandomOversamp precision 0.93 0.51	0.78 0.80 Rando ling Rando pling Mod recall 0.82 0.75	0.80 0.74 0.81 0.60637245 0.60637245 0.60637245 0.60637245 0.60637245 0.60637245 0.60637245 0.60637245 0.60637245 0.60637245 0.60637245	10000 10000 10000 10000 impling 520 ing NONE entBoosting support 7963 2037 10000 10000	FWE f_c]
accuracy macro avg weighted avg Gradient assif EVALUATING: Description:  0 1 accuracy macro avg weighted avg	0.72 0.84  Boosting 0.8011000000  GradientBoost: RandomOversample precision 0.93 0.51  0.72 0.84	0.78 0.80 Rading Rando oling Mod recall 0.82 0.75	0.80 0.74 0.81 0.81 0.60637245 0.60637245 0.60637245 0.61 0.87 0.61 0.80 0.74 0.81	10000 10000 10000 10000 10000 impling 520 	FWE f_cl
accuracy macro avg weighted avg  Gradient assif  EVALUATING: Description:   accuracy macro avg weighted avg  Gradient	0.72 0.84 Boosting 0.8011000000 GradientBoost: RandomOversamp precision 0.93 0.51	0.78 0.80 Rading Rando oling Mod recall 0.82 0.75 0.78 0.80	0.80 0.74 0.81 0.60637245 0.60637245 0.60637245 0.60637245 0.61 0.80 0.74 0.81	10000 10000 10000 10000  mmpling 520  ing NONE entBoosting support 7963 2037 10000 10000 10000	FWE f_cl

	GradientBoost NONE Model: G	_		Threshold	0.8
besci iption.	precision		_	support	
0	0.88	0.97	0.92	7963	
1	0.78	0.47	0.59	2037	
accuracy			0.86	10000	
macro avg					
weighted avg	0.86	0.86	0.85	10000	
Gradient	Boosting			NONE	Variance Threshol
	0.8648000000		0.5857843		
	GradientBoost NONE Model: G	_		ndomForest	
·	precision		_	support	
0	0.88	0.97	0.92	7963	
1	0.78	0.47	0.58	2037	
accuracy			0.86	10000	
macro avg		0.72	0.75	10000	
weighted avg	0.86	0.86	0.85	10000	
Gradient	•		0 500007	NONE	RFECV RandomF
Gradient	Boosting 0.8639000000		0.5823872		RFECV RandomF
	•		0.5823872		RFECV RandomF
orest EVALUATING:	0.8639000000  GradientBoost	ing NONE	FWE f_cla	2353	RFECV RandomF
orest EVALUATING:	0.8639000000  GradientBoost NONE Model: G	ing NONE	FWE f_cla	2353  assif	RFECV RandomF
orest EVALUATING:	0.8639000000  GradientBoost	ing NONE	FWE f_cla	2353  assif	RFECV RandomF
orest EVALUATING:	0.8639000000  GradientBoost NONE Model: G precision	ing NONE	FWE f_cla oosting f1-score	2353  assif support	RFECV RandomF
orest EVALUATING: Description:	0.8639000000 GradientBoost NONE Model: G precision 0.88	ing NONE radientB recall	FWE f_classociated for the following f1-score 0.92	2353  assif support 7963	RFECV RandomF
orest EVALUATING: Description:	0.8639000000 GradientBoost NONE Model: G precision 0.88 0.77	ing NONE radientB recall 0.96	FWE f_classing f1-score 0.92 0.59	2353 	RFECV RandomF
orest EVALUATING: Description:  0 1	0.8639000000 GradientBoost NONE Model: G precision 0.88 0.77	ing NONE radientB recall 0.96 0.47	FWE f_clamosting f1-score 0.92 0.59	2353 	RFECV RandomF
orest EVALUATING: Description:	0.8639000000 GradientBoost NONE Model: G precision 0.88 0.77	ing NONE radientB recall 0.96 0.47	FWE f_cla oosting f1-score 0.92 0.59 0.86 0.75	2353 	RFECV RandomF
orest EVALUATING: Description:  0 1 accuracy macro avg weighted avg	0.8639000000 GradientBoost NONE Model: G precision  0.88 0.77  0.82 0.86	ing NONE radientB recall 0.96 0.47 0.72 0.86	FWE f_classociations fl-score 0.92 0.59 0.86 0.75 0.85	2353 	RFECV RandomF
orest EVALUATING: Description:  0 1 accuracy macro avg weighted avg	0.8639000000 GradientBoost NONE Model: G precision  0.88 0.77  0.82 0.86	ing NONE radientB recall 0.96 0.47 0.72 0.86	FWE f_classociations fl-score 0.92 0.59 0.86 0.75 0.85	2353 	
orest EVALUATING: Description:  0 1 accuracy macro avg weighted avg	0.8639000000 GradientBoost NONE Model: G precision  0.88 0.77  0.82 0.86  Boosting	ing NONE radientB recall 0.96 0.47 0.72 0.86	FWE f_cla oosting f1-score 0.92 0.59 0.86 0.75 0.85	2353 	
orest EVALUATING: Description:  0 1 accuracy macro avg weighted avg Gradient	0.8639000000 GradientBoost NONE Model: G precision  0.88 0.77  0.82 0.86	ing NONE radientB recall 0.96 0.47 0.72 0.86	FWE f_classos fl-score 0.92 0.59 0.86 0.75 0.85	2353 assif support 7963 2037 10000 10000 10000	
orest EVALUATING: Description:  0 1 accuracy macro avg weighted avg Gradient	0.8639000000 GradientBoost NONE Model: G precision  0.88 0.77  0.82 0.86	ing NONE radientB recall 0.96 0.47 0.72 0.86	FWE f_classos fl-score 0.92 0.59 0.86 0.75 0.85	2353 assif support 7963 2037 10000 10000 10000	FWE f_cl
orest	0.8639000000 GradientBoost NONE Model: G precision  0.88 0.77  0.82 0.86	ing NONE iradientB recall 0.96 0.47 0.72 0.86	FWE f_clasosting f1-score 0.92 0.59 0.86 0.75 0.85	2353 assif support 7963 2037 10000 10000 10000	FWE f_cl
orest	0.8639000000 GradientBoost NONE Model: G precision  0.88 0.77  0.82 0.86  Boosting 0.8639000000 GradientBoost NONE Model: G	ing NONE radientB recall 0.96 0.47 0.72 0.86	FWE f_clasosting f1-score 0.92 0.59 0.86 0.75 0.85	2353 assif support 7963 2037 10000 10000 10000	FWE f_cl
orest	0.8639000000 GradientBoost NONE Model: G precision  0.88 0.77  0.82 0.86	ing NONE radientB recall 0.96 0.47 0.72 0.86	FWE f_clasosting f1-score 0.92 0.59 0.86 0.75 0.85	2353 assif support 7963 2037 10000 10000 10000	FWE f_cl
orest	0.8639000000 GradientBoost NONE Model: G precision  0.88 0.77  0.82 0.86  Boosting 0.8639000000 GradientBoost NONE Model: G precision	ing NONE radientB recall 0.96 0.47 0.72 0.86	FWE f_clasosting f1-score 0.92 0.59 0.86 0.75 0.85	2353	FWE f_cl

	0.83 0.86			10000	
	3oosting 0.8648000000			NONE 37	
	RandomForest smote Model:	RandomFo	rest		
	precision	recall	†1-score	support	
0 1			0.90 0.62		
accuracy macro avg weighted avg			0.76		
	omForest				Variance Threshol
	0.8357000000	)	0.6185279		
	RandomForest			Forest	
Description:	smote Model:			cuppont	
	precision	recarr	11-Score	Support	
0			0.90		
1	0.59	0.66	0.62	2037	
accuracy			0.84	10000	
macro avg		0.77		10000	
weighted avg	0.84	0.84	0.84	10000	
Rando	omForest 0.8372000000 			smote 148 	RFECV RandomF
	RandomForest smote Model:	RandomFo	rest		
	precision	recall	f1-score	support	
0	0.91	0.88	0.89	7963	
1		0.68			
accuracy			0.84	10000	
accuracy macro avg	0.75	0.78			
weighted avg		0.84		10000	
<b></b>			· <b></b>	<b></b>	<b>_</b>

<del>-</del>	

FWE f_c	smote	0.6261343		omForest 0.8352000000	Rando assif
				RandomForest	
				<pre>smote Model:</pre>	Description:
	support	f1-score	recall	precision	
	7963	0.90	0.88	0.91	0
	2037	0.62	0.65	0.59	1
	10000	0.84			accuracy
			0.77	0.75	macro avg
	10000	0.84	0.84		veighted avg
	smote			omForest	
	77 	0.61852797		0.8357000000 	IONE 
					·
	reshold 0.8		-	RandomForest adasyn Model:	
	support			precision	escription.
				·	
			0.86		0
	2037	0.61	0.68	0.55	1
	10000	0.82			accuracy
	10000	0.75	0.77	0.73	macro avg
	10000	0.83	0.82	0.84	veighted avg
Variance Thresho	adasyn			omForest	Rando
				0.8221000000	1 0.8
	mForest			${\tt RandomForest}$	
				adasyn Model:	Description:
	support	†1-score	recall	precision	
			0 86	0.91	0
	7963	0.88	0.00		
		0.88 0.61			1
	2037	0.61			_
	2037 10000	0.61 0.82	0.69	0.55	accuracy macro avg
	2037 10000 10000	0.61 0.82 0.75	<ul><li>0.69</li><li>0.77</li></ul>	<ul><li>0.55</li><li>0.73</li></ul>	accuracy macro avg
	2037 10000 10000	0.61 0.82 0.75	<ul><li>0.69</li><li>0.77</li></ul>	<ul><li>0.55</li><li>0.73</li></ul>	accuracy macro avg
RFECV Random	2037 10000 10000	0.61 0.82 0.75	<ul><li>0.69</li><li>0.77</li></ul>	<ul><li>0.55</li><li>0.73</li></ul>	accuracy macro avg weighted avg

 ${\tt EVALUATING:} \quad {\tt RandomForest \ adasyn \ FWE \ f\_classif}$ 

Docenintion	adasyn Modol.	Pandom F	onost		
Description.	adasyn Model: precision			support	
	p. 002020		555. 5	Juppo. C	
0		0.84			
1	0.54	0.71	0.61	2037	
266119261			0 02	10000	
accuracy macro avg		0.79	0.82 0.75		
weighted avg					
weighted avg	0.04	0.02	0.05	10000	
	omForest		0.640044	adasyn	FWE f_cl
assif	0.8170000000			1624	
EVALUATING:	RandomForest	adasyn N	ONE		
	adasyn Model:	-			
•	precision			support	
0			0.88		
1	0.55	0.68	0.61	2037	
accuracy			0 82	10000	
accuracy macro avg		0.77		10000	
weighted avg					
	omForest		0 64046636	adasyn	
NONE	0.8221000000		0.61046638	-	
NONE				-	
NONE 	0.8221000000 				0.8
NONE EVALUATING:	0.8221000000	svmsmote	Variance		0.8
NONE EVALUATING:	0.8221000000  RandomForest	svmsmote 1: Rando	Variance mForest	393 Threshold	0.8
NONE EVALUATING: Description:	0.8221000000  RandomForest svmsmote Mode precision	svmsmote l: Rando recall	Variance mForest f1-score	393 Threshold support	0.8
NONE EVALUATING: Description:	0.8221000000  RandomForest svmsmote Mode precision  0.91	svmsmote l: Rando recall 0.87	Variance mForest f1-score 0.89	Threshold support 7963	0.8
NONE EVALUATING: Description:	0.8221000000  RandomForest svmsmote Mode precision  0.91	svmsmote l: Rando recall	Variance mForest f1-score 0.89	Threshold support 7963	0.8
NONE EVALUATING: Description:	0.8221000000  RandomForest symsmote Mode precision  0.91 0.58	svmsmote l: Rando recall 0.87	Variance mForest f1-score 0.89 0.62	Threshold support 7963 2037	0.8
NONE EVALUATING: Description:  0 1 accuracy	0.8221000000  RandomForest symsmote Mode precision  0.91 0.58	svmsmote l: Rando recall 0.87 0.68	Variance mForest f1-score 0.89 0.62	Threshold support 7963 2037	0.8
NONE EVALUATING: Description:  0 1 accuracy macro avg	0.8221000000  RandomForest symsmote Mode precision  0.91 0.58	svmsmote l: Rando recall 0.87 0.68	Variance mForest f1-score 0.89 0.62 0.83 0.76	Threshold support 7963 2037 10000	0.8
NONE EVALUATING: Description:  0 1 accuracy macro avg	0.8221000000  RandomForest symsmote Mode precision  0.91 0.58	svmsmote l: Rando recall 0.87 0.68	Variance mForest f1-score 0.89 0.62 0.83 0.76	Threshold support 7963 2037 10000	0.8
NONE EVALUATING: Description:  0 1 accuracy macro avg weighted avg	0.8221000000  RandomForest symsmote Mode precision  0.91 0.58	svmsmote 1: Rando recall 0.87 0.68 0.78 0.83	Variance mForest f1-score 0.89 0.62 0.83 0.76 0.84	Threshold support 7963 2037 10000 10000	
NONE	0.8221000000  RandomForest symsmote Mode precision  0.91	svmsmote 1: Rando recall 0.87 0.68 0.78 0.83	Variance mForest f1-score 0.89 0.62 0.83 0.76 0.84	Threshold support 7963 2037 10000 10000	
NONE EVALUATING: Description:  0 1 accuracy macro avg weighted avg	0.8221000000  RandomForest svmsmote Mode precision  0.91 0.58  0.75 0.85	svmsmote 1: Rando recall 0.87 0.68 0.78 0.83	Variance mForest f1-score 0.89 0.62 0.83 0.76 0.84	Threshold support 7963 2037 10000 10000	
NONE EVALUATING: Description:  0 1 accuracy macro avg weighted avg	RandomForest svmsmote Mode precision  0.91 0.58  0.75 0.85	svmsmote 1: Rando recall 0.87 0.68 0.78 0.83	Variance mForest f1-score 0.89 0.62 0.83 0.76 0.84	Threshold support 7963 2037 10000 10000 10000	
NONE EVALUATING: Description:  0 1 accuracy macro avg weighted avg	RandomForest svmsmote Mode precision  0.91 0.58  0.75 0.85	svmsmote 1: Rando recall 0.87 0.68 0.78 0.83	Variance mForest f1-score 0.89 0.62 0.83 0.76 0.84	Threshold support 7963 2037 10000 10000 10000	Variance Threshol
NONE EVALUATING: Description:  0 1 accuracy macro avg weighted avg Rando d 0.8	RandomForest svmsmote Mode precision  0.91 0.58  0.75 0.85	svmsmote 1: Rando recall 0.87 0.68 0.78 0.83	Variance mForest f1-score  0.89 0.62  0.83 0.76 0.84	Threshold support 7963 2037 10000 10000 10000	Variance Threshol
NONE	RandomForest svmsmote Mode precision  0.91 0.58  0.75 0.85  OMForest 0.8331000000  RandomForest svmsmote Mode	svmsmote 1: Rando recall 0.87 0.68 0.78 0.83	Variance mForest f1-score 0.89 0.62 0.83 0.76 0.84	Threshold support 7963 2037 10000 10000 10000	Variance Threshol
NONE	0.8221000000	svmsmote 1: Rando recall 0.87 0.68 0.78 0.83	Variance mForest f1-score 0.89 0.62 0.83 0.76 0.84	Threshold support 7963 2037 10000 10000 10000	Variance Threshol
NONE	RandomForest svmsmote Mode precision  0.91 0.58  0.75 0.85  omForest 0.8331000000  RandomForest svmsmote Mode precision	svmsmote 1: Rando recall 0.87 0.68 0.78 0.83 svmsmote 1: Rando recall	Variance mForest f1-score  0.89 0.62  0.83 0.76 0.84  RFECV Rar mForest f1-score	Threshold support 7963 2037 10000 10000 10000 svmsmote 5190 adomForest support	Variance Threshol
NONE	0.8221000000	svmsmote 1: Rando recall 0.87 0.68 0.78 0.83 svmsmote 1: Rando recall	Variance mForest f1-score  0.89 0.62  0.83 0.76 0.84  RFECV Rar mForest f1-score  0.89	Threshold support 7963 2037 10000 10000 10000 svmsmote 6190 ndomForest support 7963	Variance Threshol

accuracy macro avg weighted avg	0.75 0.85	0.83	0.84	10000	
 Rando orest	omForest 0.8327000000				RFECV RandomF
	RandomForest svmsmote Mode	l: Rando	mForest		
	precision	recall	f1-score	support	
0 1	0.92 0.57		0.89 0.63		
accuracy macro avg weighted avg	0.74	0.78 0.83	0.83 0.76 0.84	10000	
	omForest		SV		FWE f_cl
assif	0.8321000000		0.62829311	.49	_
EVALUATING:	RandomForest	svmsmote	NONE		
Description:	svmsmote Mode				
	precision	recall	†1-score	support	
0	0.91	0.87	0.89	7963	
1	0.58	0.68	0.62	2037	
accuracy			0 83	10000	
macro avg	0.75	0.78		10000	
weighted avg				10000	
Rando NONE	omForest 0.8331000000 			rmsmote 90 	
	RandomForest smotetomek Mo			• Threshold	0.8
besci ipcion.			f1-score	support	
0			0.89		
1	0.58	0.65	0.62	2037	
accuracy			0.83	10000	
macro avg	0.75	0.77			
weighted avg				10000	

d 0.8	omForest 0.8349000000 		0.6177355	tetomek 869 	Variance Threshol
	RandomForest			andomForest	
Description:	smotetomek Mo precision			cuppont	
	precision	recarr	f1-score	Support	
0	0.91	0.88	0.89	7963	
1	0.59	0.66	0.62	2037	
accuracy		0 77	0.84 0.76		
macro avg weighted avg					
weighted avg					
	omForest			tetomek	RFECV RandomF
	0.8352000000				
 EVALUATING:	RandomForest	cmototon	nok EWE f c	laccif	
	smotetomek Mo		_	105511	
	precision			support	
0	0.91	0.88	0.90	7963	
1		0.68			
accuracy		0.70	0.84		
macro avg weighted avg		0.78			
weighted avg	0.03	0.04	0.04	10000	
Rand	omForest		smo		FWE f_cl
assif	0.8365000000				
	Dan dam Fanaat		al NONE		
	RandomForest smotetomek Mo				
beset iperom.	precision			support	
0	0.91	0.88	0.89	7963	
1		0.65			
accuracy			0.83	10000	
macro avg		0.77			
weighted avg					
Rand	omForest		smo	tetomek	
			0.61773558		

EVALUATING: RandomForest RandomOversampling Variance Threshold 0.8 Description: RandomOversampling Model: RandomForest

	precisi	on recall	f1-score	support	
	0 0.9	92 0.84	0.88	7963	
		54 0.72			
accurac	V		0.82	10000	
macro av	g 0.	73 0.78	0.75	10000	
weighted av	g 0.	84 0.82	0.83	10000	
		F 00000			Variance Threshol
		rest RandomO\ ersampling Mo			domForest
besci ipcion		on recall			
		92 0.84 54 0.71			
	1 0.	54 0.71	0.01	2037	
accurac			0.82		
macro av	g 0.	73 0.78	0.75	10000	
weighted av	g 0.	84 0.82	0.83	10000	
 Pan	domEonost	F	PandomOvons	ampling	RFECV RandomF
		00000		. •	KFECV KAHUOIIIF
 EVALUATING:	RandomFo	rest RandomOv	/Arsamnling	r FWF f cla	ccif
		ersampling Mo		_	3311
	precisi	on recall	f1-score	support	
	0 0.9	92 0.84	0.88	7963	
	1 0.		0.62		
accurac	-	72 0 79	0.82		
macro av weighted av	_	73 0.78 84 0.82			
6	6				
 Ran	domForest	F	RandomOvers	ampling	FWE f_cl
assif	0.81610		0.6158345		
 FVALUATING:	RandomFo	rest RandomOv	/ersampling	NONE	
		ersampling Mo			
	precisi	on recall	11 30010	F	
	·				
	·	92 0.84		7963	
	0 0.9	92 0.84	0.88	7963	

	0.73 0.84				
Rando NONE	omForest 0.8163000000		RandomOvers 0.61383224		
	RandomForest NONE Model: Ra			shold 0.8	
	precision			support	
0	0.92	0.83	0.88	7963	
1	0.53	0.72	0.61	2037	
accuracy			0.81	10000	
macro avg					
weighted avg	0.84	0.81	0.82	10000	
	omForest				Variance Threshol
d 0.8	0.8110000000		0.60788383	174 	
	D   E	NONE DE	:CV D		
	RandomForest NONE Model: Ra			orest	
•	precision			support	
0	0.92	0.83	0.88	7963	
1	0.53	0.72	0.61	2037	
accuracy			0.81	10000	
	0.72				
weighted avg	0.84	0.81	0.82	10000	
Rando	omForest			NONE	RFECV RandomF
orest	0.8119000000				
	RandomForest NONE Model: Ra		_		
besci ipcion.	precision			support	
0	0 93	a 82	0.87	7963	
1	0.52				
accuracy			0.81	10000	
macro avg		0.78			
weighted avg	0.84	0.81	0.82	10000	

RandomForest NONE FWE  $f_cl$ 

assif	0.8062000000		0.6088009	689	
EVALUATING:	RandomForest	NONE NON	IE		
Description:	NONE Model: R	andomFor	est		
-	precision	recall	f1-score	support	
0			0.88		
1	0.53	0.72	0.61	2037	
2661102614			Ω 01	10000	
accuracy	0.72	0 78			
_	0.84				
werphicea avb	0.01	0.01	0.02	10000	
	omForest			NONE	
NONE	0.8110000000		0.60788381	74	
EVALUATING:	SVC smote Var	iance Th	reshold 0.8	8	
	smote Model:				
	precision		f1-score	support	
				• •	
0	0.92	0.82	0.87	7963	
1	0.50	0.72	0.59	2037	
accuracy				10000	
macro avg					
weighted avg	0.84	0.80	0.81	10000	
	SVC			smote	Variance Threshol
d 0.8	0.7985000000	)	0.5936680		
	SVC smote RFE		mForest		
Description:	smote Model:		C4		
	precision	recall	+1-score	support	
0	0.92	0.82	0.87	7963	
1		0.72	0.59		
_	0.50	0.72	0.33	2037	
accuracy			0.80	10000	
macro avg		0.77		10000	
weighted avg					
	SVC			smote	RFECV RandomF
		1	0.59395163		
orest	0.7986000000				
orest	0.7986000000				
	0.7986000000  SVC smote FWE				
EVALUATING:		f_class			

0	0.92	0.82	0.87	7963	
1	0.51	0.73	0.60	2037	
accuracy			0.80	10000	
macro avg	0.71	0.77	0.73	10000	
weighted avg	0.84	0.80	0.81	10000	
assif	SVC 0.8006000000		0.5991154	smote 001	FWE f_cl
	CVC amata NON				
	SVC smote NON smote Model:				
beset iperon.	precision		f1-score	support	
0	0.92	0.82	0.87	7963	
1			0.59	2037	
accuracy			0.80	10000	
	0.71	0.77			
weighted avg	0.84	0.80	0.81	10000	
NONE	SVC 0.7985000000		0.59366807	smote	
	SVC adasyn Va		hreshold 0	.8	
vescription:	adasyn Model: precision		f1-score	support	
а	0.93	0.77	0.84	7963	
1		0.77		2037	
accuracy			0.77	10000	
macro avg	0.69	0.77	0.71	10000	
weighted avg	0.83	0.77	0.79	10000	
	SVC			adasyn	Variance Threshol
d 0.8	0.7668000000		0.5724239	091	
	SVC adasyn RF adasyn Model:		lomForest		
Desci iperoil.	precision		f1-score	support	
0	0.93	0.77	0.84	7963	
1		0.77		2037	
accuracy			0.77	10000	
macro avg		0.77		10000	

weighted avg	0.83	0.77	0.79	10000	
	SVC			adasyn	RFECV RandomF
orest	0.7668000000		0.5724239	091 	
	SVC adasyn FW	_	sif		
Description:	adasyn Model:		C4		
	precision	recall	f1-score	support	
0	0.93	0.76	0.84	7963	
1			0.58		
accuracy				10000	
macro avg					
weighted avg	0.84	0.77	0.79	10000	
	SVC			adasyn	FWE f_cl
assif	0.7672000000		0.5784136	183	
EVALUATING:	SVC adasyn NO	NE			
	adasyn Model:				
·	precision		f1-score	support	
_					
0			0.84		
1	0.46	0.77	0.57	2037	
accuracy			0.77	10000	
macro avg		0.77		10000	
	0.83	0.77	0.79	10000	
	SVC			adasyn	
NONE	0.7668000000		0.57242390	-	
	SVC symsmote		Threshold	0.8	
Description:	svmsmote Mode precision		f1_scono	support	
	precision	recarr	11-20016	Support	
0	0.92	0.81	0.87	7963	
1			0.60		
accuracy			0.80		
	0.71				
weighted avg	0.84	0.80	0.81	10000	
	SVC			vmsmote	Variance Threshol
d 0.8	0.7984000000		0.5985663	082	

<b>EVALUATING:</b>	SVC svmsmote	RFECV Ra	ndomForest		
Description:	svmsmote Mode	1: SVC			
•	precision		f1-score	support	
	•				
0	0.92	0.81	0.87	7963	
1	0.50	0.74	0.60	2037	
accuracy				10000	
macro avg	0.71	0.78	0.73	10000	
weighted avg	0.84	0.80	0.81	10000	
	SVC				RFECV RandomF
orest	0.7991000000		0.5997210	600	
	SVC svmsmote				
Description:	svmsmote Mode				
	precision	recall	f1-score	support	
0					
1	0.50	0.74	0.60	2037	
accuracy			0.80	10000	
macro avg		0.78	0.73	10000	
weighted avg	0.84	0.80	0.81	10000	
	SVC			vmsmote	FWE f_cl
assif	0.7979000000		0.6001978	239	
	61.6				
	SVC symsmote				
Description:	svmsmote Mode		<b>C</b> 4		
	precision	recall	f1-score	support	
	0.00	0.01		70.53	
0					
1	0.50	0.74	0.60	2037	
			0.00	10000	
accuracy		0.70	0.80		
macro avg					
weighted avg	0.84	0.80	0.81	10000	
	CVC		-	·memete	
NONE	SVC			vmsmote	
NONE	0.7984000000		0.59856630	04	
EVALUATING.	SVC smotetome	k Vanian	nce Throche	1400	
				IU 0.0	
Description	smotetomek Mo	מפוי לאיי			

precision recall f1-score support

	0	0.92	0.82	0.87	7963	
	1	0.50	0.72	0.59	2037	
accui	racv			0.80	10000	
macro	_	0.71	0.77	0.73	10000	
	_	0.84	0.77	0.75	10000	
weighted	avg	0.04	0.00	0.81	10000	
		SVC				Variance Threshol
d 0.8		0.7978000000	)	0.5928312	525	
EVALUATI	NG:	SVC smotetome	k RFECV	RandomFore	st	
Descript:	ion:	smotetomek Mc	odel: SVC			
		precision	recall	f1-score	support	
		•				
	0	0.92	0.82	0.87	7963	
	1	0.50	0.72	0.59	2037	
	_		• • • •	0.22	_00,	
accui	racv			0 80	10000	
macro	-		0 77	0.00	10000	
	_					
weighted	avg	0.83	0.00	0.01	10000	
		SVC				RFECV RandomF
		0.7973000000				
 EVALUATIN	NG:	SVC smotetome	ek FWE f_	classif		
		SVC smotetomes				
			del: SVC		support	
		smotetomek Mc	del: SVC		support	
	ion:	smotetomek Mc precision	odel: SVC recall	f1-score		
	ion:	smotetomek Mc precision 0.92	odel: SVC recall 0.82	f1-score 0.87	7963	
	ion: 0	smotetomek Mc precision 0.92	odel: SVC recall 0.82	f1-score		
Descript:	ion: 0 1	smotetomek Mc precision 0.92	odel: SVC recall 0.82	f1-score 0.87 0.60	7963 2037	
Descript:	ion: 0 1 racy	smotetomek Mo precision 0.92 0.50	odel: SVC recall 0.82 0.73	f1-score 0.87 0.60 0.80	7963 2037 10000	
Descript: accumacro	ion:  0 1 racy avg	smotetomek Mo precision  0.92  0.50	odel: SVC recall 0.82 0.73	f1-score 0.87 0.60 0.80 0.73	7963 2037 10000 10000	
Descript:	ion:  0 1 racy avg	smotetomek Mo precision  0.92  0.50	odel: SVC recall 0.82 0.73	f1-score 0.87 0.60 0.80 0.73	7963 2037 10000	
Descript: accumacro	ion:  0 1 racy avg avg	smotetomek Mo precision  0.92  0.50  0.71  0.84	0.82 0.73 0.77 0.80	f1-score 0.87 0.60 0.80 0.73 0.81	7963 2037 10000 10000 10000	
Descript: accumacro	ion:  0 1 racy avg avg	smotetomek Mo precision  0.92  0.50  0.71  0.84	0.82 0.73 0.77 0.80	f1-score 0.87 0.60 0.80 0.73 0.81	7963 2037 10000 10000 10000	
Descript: accumacro	ion:  0 1 racy avg avg	smotetomek Mo precision  0.92 0.50  0.71 0.84	0.82 0.73 0.77 0.80	f1-score 0.87 0.60 0.80 0.73 0.81	7963 2037 10000 10000 10000	
accum macro weighted	ion:  0 1 racy avg avg	smotetomek Morprecision  0.92 0.50  0.71 0.84	0.82 0.73 0.77 0.80	f1-score  0.87  0.60  0.80  0.73  0.81	7963 2037 10000 10000 10000	 FWE f_cl
accur macro weighted	ion:  0 1 racy avg avg	smotetomek Mc precision  0.92 0.50  0.71 0.84  SVC 0.79860000000	0.82 0.73 0.77 0.80	f1-score  0.87 0.60  0.80 0.73 0.81	7963 2037 10000 10000 10000	FWE f_cl
accur macro weighted	ion:  0 1 racy avg avg	smotetomek Mc precision  0.92 0.50  0.71 0.84  SVC 0.79860000000	0.82 0.73 0.77 0.80	f1-score  0.87 0.60  0.80 0.73 0.81	7963 2037 10000 10000 10000	
accur macro weighted	ion:  0 1 racy avg avg	smotetomek Mc precision  0.92 0.50  0.71 0.84  SVC 0.79860000000	0.82 0.73 0.77 0.80	f1-score  0.87 0.60  0.80 0.73 0.81	7963 2037 10000 10000 10000	FWE f_cl
accummacro weighted	ion:  0 1  racy avg avg	smotetomek Mo precision  0.92 0.50  0.71 0.84  SVC 0.79860000000	0.82 0.73 0.77 0.80	f1-score  0.87 0.60  0.80 0.73 0.81  smo	7963 2037 10000 10000 10000	FWE f_cl
accummacro weighted	ion:  0 1  racy avg avg	smotetomek Mo precision  0.92 0.50  0.71 0.84  SVC 0.7986000000	0.82 0.73 0.77 0.80	f1-score  0.87 0.60  0.80 0.73 0.81  smo	7963 2037 10000 10000 10000	FWE f_cl
accummacro weighted	ion:  0 1  racy avg avg	smotetomek Mo precision  0.92 0.50  0.71 0.84  SVC 0.79860000000	odel: SVC recall 0.82 0.73 0.77 0.80	f1-score  0.87 0.60  0.80 0.73 0.81	7963 2037 10000 10000 10000 tetomek	FWE f_cl
accummacro weighted	ion:  0 1  racy avg avg	smotetomek Mc precision  0.92 0.50  0.71 0.84  SVC 0.7986000000000000000000000000000000000000	odel: SVC recall 0.82 0.73 0.77 0.80	f1-score  0.87 0.60  0.80 0.73 0.81	7963 2037 10000 10000 10000 tetomek	FWE f_cl
accummacro weighted	ion:  0 1 racy avg avg	smotetomek Morprecision  0.92 0.50  0.71 0.84  SVC 0.7986000000000000000000000000000000000000	odel: SVC recall 0.82 0.73 0.77 0.80	f1-score  0.87 0.60  0.80 0.73 0.81  smo 0.59574465 f1-score	7963 2037 10000 10000 10000 tetomek 809 	FWE f_cl
accummacro weighted	ion:  0 1  racy avg avg	smotetomek Morprecision  0.92 0.50  0.71 0.84  SVC 0.7986000000000000000000000000000000000000	odel: SVC recall 0.82 0.73 0.77 0.80 	f1-score  0.87 0.60  0.80 0.73 0.81  smo 0.59574468 f1-score 0.87	7963 2037 10000 10000 10000 tetomek 809 support 7963	FWE f_cl
accummacro weighted	ion:  0 1 racy avg avg	smotetomek Morprecision  0.92 0.50  0.71 0.84  SVC 0.7986000000000000000000000000000000000000	odel: SVC recall 0.82 0.73 0.77 0.80	f1-score  0.87 0.60  0.80 0.73 0.81  smo 0.59574468 f1-score 0.87	7963 2037 10000 10000 10000 tetomek 809 support 7963	FWE f_cl
accummacro weighted assif EVALUATIN Descript:	ion:  0 1 racy avg avg ion:	smotetomek Morprecision  0.92 0.50  0.71 0.84  SVC 0.7986000000000000000000000000000000000000	odel: SVC recall 0.82 0.73 0.77 0.80 	f1-score  0.87 0.60  0.80 0.73 0.81  smo 0.59574466 f1-score  0.87 0.59	7963 2037 10000 10000 10000 	FWE f_cl
accummacro weighted	ion:  0 1 racy avg avg ion:	smotetomek Morprecision  0.92 0.50  0.71 0.84  SVC 0.7986000000000000000000000000000000000000	0.82 0.73 0.77 0.80 0.77 0.80 0.77 0.80 0.72	f1-score  0.87 0.60  0.80 0.73 0.81  smo 0.59574465 f1-score  0.87 0.59 0.80	7963 2037 10000 10000 10000 	FWE f_cl
accummacro weighted assif EVALUATIN Descript:	ion:  0 1 racy avg avg NG: ion:  0 1 racy avg	smotetomek Morprecision  0.92 0.50  0.71 0.84  SVC 0.7986000000000000000000000000000000000000	0.82 0.73 0.77 0.80 0.77 0.80 0.77 0.80 0.72	f1-score  0.87 0.60  0.80 0.73 0.81  smo 0.59574468 f1-score  0.87 0.59 0.80 0.73	7963 2037 10000 10000 10000 	FWE f_cl

	SVC			tetomek		
NONE	0.7978000000 	0	.592831252	25 		
	SVC RandomOver			Threshold	0.8	
Description:	RandomOversamprecision	_		support		
	p. 002020		555. 5	эмрро. с		
0	0.93					
1	0.50	0.75	0.60	2037		
accuracy			0.79	10000		
macro avg	0.71	0.78				
weighted avg				10000		
	CVC	D -				Vanianas Thurshall
d 0.8	0.7942000000		0.59615384	. •		Variance Threshol
 EVALUATING:	SVC RandomOver	ncampling	DEECV Day	ndomEonost		
	RandomOversam			idollirorest		
beset iperon.	precision			support		
	·					
0			0.86			
1	0.50	0.75	0.60	2037		
accuracy			0.79	10000		
macro avg	0.71	0.78	0.73			
weighted avg	0.84	0.79	0.81	10000		
	SVC	D o	ndomOvone:	amaling		RFECV RandomF
orest	0.7934000000		ndomOversa 0.59506076	. •		KFECV Kalluoliir
	SVC RandomOver		_	assif		
Description:	RandomOversamprecision	_		sunnort		
	precision	rccuii	11 30010	Suppor C		
0	0.93	0.81	0.86	7963		
1	0.50	0.75	0.60	2037		
accuracy			a ga	10000		
macro avg	0.71	0.78		10000		
weighted avg						
	SVC		ndomOversa			FWE f_cl
assif	0.7950000000		0.59976573	165		

EVALUATING:	SVC RandomOve	rsamplin	g NONE		
	RandomOversam				
	precision			support	
0	0.93	0 81	0.86	7963	
1		0.75		2037	
_	0.50	0.75	0.00	2037	
accuracy			0.80	10000	
macro avg		0.78	0.73		
_					
weighted avg	0.84	0.00	0.81	10000	
	SVC	В	andomOvana:	amalina	
NONE			andomOvers		
NONE	0.7973000000		0.599960528	59	
	CVC NONE V				
	SVC NONE Vari		esnoia 0.8		
Description:	NONE Model: S		Ca		
	precision	recall	†1-score	support	
_					
0			0.92		
1	0.82	0.38	0.52	2037	
accuracy			0.86		
macro avg	0.84	0.68	0.72	10000	
weighted avg	0.85	0.86	0.84	10000	
	SVC			NONE	Variance Threshol
d 0.8	SVC 0.8569000000		0.5196374		Variance Threshol
d 0.8			0.5196374		Variance Threshol
d 0.8			0.5196374		Variance Threshol
					Variance Threshol
EVALUATING:	0.8569000000	V Random			Variance Threshol
EVALUATING:	0.8569000000  SVC NONE RFEC	 V Random VC	Forest	522	Variance Threshol
EVALUATING:	0.8569000000 SVC NONE RFEC NONE Model: S	 V Random VC	Forest	522	Variance Threshol
EVALUATING:	0.8569000000 SVC NONE RFEC NONE Model: S precision	 V Random VC	Forest f1-score	support	Variance Threshol
EVALUATING: Description:	0.8569000000  SVC NONE RFEC  NONE Model: S  precision  0.86	V Random VC recall	Forest f1-score 0.92	support 7963	Variance Threshol
EVALUATING: Description:	0.8569000000  SVC NONE RFEC  NONE Model: S  precision  0.86	V Random VC recall 0.98	Forest f1-score 0.92	support 7963	Variance Threshol
EVALUATING: Description:	0.8569000000  SVC NONE RFEC  NONE Model: S  precision  0.86	V Random VC recall 0.98	Forest f1-score 0.92	support 7963 2037	Variance Threshol
EVALUATING: Description:	0.8569000000 SVC NONE RFEC NONE Model: S precision 0.86 0.82	V Random VC recall 0.98 0.38	Forest  f1-score  0.92 0.52  0.86	support 7963 2037	Variance Threshol
EVALUATING: Description:  0 1 accuracy macro avg	0.8569000000  SVC NONE RFEC NONE Model: S precision  0.86 0.82	V Random VC recall 0.98 0.38	Forest  f1-score  0.92  0.52  0.86  0.72	support 7963 2037 10000 10000	Variance Threshol
EVALUATING: Description:  0 1 accuracy macro avg	0.8569000000 SVC NONE RFEC NONE Model: S precision 0.86 0.82	V Random VC recall 0.98 0.38	Forest  f1-score  0.92  0.52  0.86  0.72	support 7963 2037 10000 10000	Variance Threshol
EVALUATING: Description:  0 1 accuracy macro avg	0.8569000000  SVC NONE RFEC NONE Model: S precision  0.86 0.82	V Random VC recall 0.98 0.38	Forest  f1-score  0.92  0.52  0.86  0.72	support 7963 2037 10000 10000	Variance Threshol
EVALUATING: Description:  0 1 accuracy macro avg	0.8569000000  SVC NONE RFEC NONE Model: S precision  0.86 0.82	V Random VC recall 0.98 0.38	Forest  f1-score  0.92  0.52  0.86  0.72	support 7963 2037 10000 10000	Variance Threshol
EVALUATING: Description:  0 1 accuracy macro avg	0.8569000000  SVC NONE RFEC  NONE Model: S  precision  0.86  0.82  0.84  0.85	V Random VC recall 0.98 0.38	Forest  f1-score  0.92  0.52  0.86  0.72	support 7963 2037 10000 10000	
EVALUATING: Description:  0 1 accuracy macro avg weighted avg	0.8569000000  SVC NONE RFEC NONE Model: S precision  0.86 0.82  0.84 0.85	V Random VC recall 0.98 0.38 0.68 0.86	Forest  f1-score  0.92 0.52  0.86 0.72 0.84	support 7963 2037 10000 10000	Variance Threshol
EVALUATING: Description:  0 1 accuracy macro avg	0.8569000000  SVC NONE RFEC  NONE Model: S  precision  0.86  0.82  0.84  0.85	V Random VC recall 0.98 0.38 0.68 0.86	Forest  f1-score  0.92  0.52  0.86  0.72	support 7963 2037 10000 10000	
EVALUATING: Description:  0 1 accuracy macro avg weighted avg	0.8569000000  SVC NONE RFEC NONE Model: S precision  0.86 0.82  0.84 0.85	V Random VC recall 0.98 0.38 0.68 0.86	Forest  f1-score  0.92 0.52  0.86 0.72 0.84	support 7963 2037 10000 10000	
EVALUATING: Description:  0 1 accuracy macro avg weighted avg  orest	0.8569000000  SVC NONE RFEC  NONE Model: S precision  0.86 0.82  0.84 0.85  SVC 0.8575000000	V Random VC recall 0.98 0.38 0.68 0.86	f1-score 0.92 0.52 0.86 0.72 0.84	support 7963 2037 10000 10000	
EVALUATING: Description:  0 1 accuracy macro avg weighted avg  orest EVALUATING:	0.8569000000  SVC NONE RFEC NONE Model: S precision  0.86 0.82  0.84 0.85  SVC 0.8575000000	V Random VC recall 0.98 0.38 0.68 0.86	f1-score 0.92 0.52 0.86 0.72 0.84	support 7963 2037 10000 10000	
EVALUATING: Description:  0 1 accuracy macro avg weighted avg  orest EVALUATING:	0.8569000000  SVC NONE RFEC NONE Model: S precision  0.86 0.82  0.84 0.85  SVC 0.8575000000  SVC NONE FWE NONE Model: S	V Random VC recall 0.98 0.38 0.68 0.86	Forest  f1-score  0.92 0.52  0.86 0.72 0.84	support 7963 2037 10000 10000 10000	
EVALUATING: Description:  0 1 accuracy macro avg weighted avg  orest EVALUATING:	0.8569000000  SVC NONE RFEC NONE Model: S precision  0.86 0.82  0.84 0.85  SVC 0.8575000000	V Random VC recall 0.98 0.38 0.68 0.86	Forest  f1-score  0.92 0.52  0.86 0.72 0.84	support 7963 2037 10000 10000 10000	
EVALUATING: Description:  0 1 accuracy macro avg weighted avg  orest EVALUATING:	0.8569000000  SVC NONE RFEC NONE Model: S precision  0.86 0.82  0.84 0.85  SVC 0.8575000000  SVC NONE FWE NONE Model: S	V Random VC recall 0.98 0.38 0.68 0.86	Forest  f1-score  0.92 0.52  0.86 0.72 0.84	support 7963 2037 10000 10000 10000 NONE 121 support	

	2037	0.53	0.39	0.82	1	
		0.86				accu
	10000			0.84	•	macro
	10000	0.84	0.86	0.85	avg	weighted
FWE f_cl	NONE			SVC		
	75 	).52747252 		581000000 	0.	assif 
				NONE NONE	NG. SVC	 Εναι παττ
			_	Model: SV		
	support	1-score		cision		<b>,</b> -
	7963	0.92	0.98	0.86	0	
	2037	0.52	0.38	0.82	1	
	10000	0.86			racy	accu
	10000	0.72	0.68	0.84	avg	macro
	10000	0.84	0.86	0.85	avg	weighted
	NONE			SVC		
	2	519637462		69000000	0.8	NONE
Variance Threshol	smote			ing	ientBoos	Grad
DEFOV D   F		62337046		411000000 ·		
RFECV RandomF	smote	.62681590		1ng 433000000	ientBoos	Grad orest
FWE f_cl	os smote	.02001590			.ں ientBoos	
TWL T_CI		62272727		340000000		assif
	smote				ientBoos	Grad
	9	623370466		11000000	0.8	NONE
Variance Threshol	adasyn			_	ientBoos	
RFECV RandomF		62576687		414000000		d 0.8
KFECV Kalluollir	adasyn 43	.62638659		417000000	ientBoos a	orest
FWE f_cl	adasyn				ientBoos	
_	•	.61692408		239000000		assif
	adasyn			_	ientBoos	
		625766871		14000000 ·		NONE
Variance Threshol	msmote 04	sv 63022508.		ing 390000000	ientBoos	Grad d 0.8
RFECV RandomF	msmote				ientBoos	
THE LEVE TRAITED III		).62750518		383000000		orest
FWE f_cl	msmote	sv		ing	ientBoos	Grad
	80	.62903939		324000000		assif
	msmote			_	ientBoos	
Vanianaa Thaash-1		630225080		90000000		NONE
Variance Threshol	etomek 24	smoτ 62661955.		1ng 415000000	ientBoos a	Grad d 0.8
		5.02001933 Smot			.ں ientBoos	
RFFCV RandomF	CIUNICA			-··o		Ji uu
RFECV RandomF		62773029		432000000		orest
RFECV RandomF FWE f_cl		62773029				orest

	dientBoosting	smotetomek	
NONE Gra	0.8415000000 dientBoosting	0.6266195524 RandomOversampling	Variance Threshol
	0.8023000000 dientBoosting	0.6065671642 RandomOversampling	RFECV RandomF
orest	0.8043000000	0.6077370214	
	dientBoosting 0.8011000000	RandomOversampling	FWE f_cl
assif	dientBoosting	0.6063724520 RandomOversampling	
NONE	0.8023000000	0.6062537343	
Gra	dientBoosting	NONE	Variance Threshol
d 0.8	0.8648000000	0.5857843137	
Gra	dientBoosting	NONE	RFECV RandomF
orest	0.8639000000	0.5823872353	
	dientBoosting	NONE	FWE f_cl
assif	0.8639000000 dientBoosting	0.5851874429 NONE	
NONE	0.864800000	0.5857843137	
NONE	RandomForest	smote	Variance Threshol
d 0.8	0.8357000000	0.6185279777	
	RandomForest	smote	RFECV RandomF
orest	0.8372000000	0.6226240148	
	RandomForest	smote	FWE f_cl
assif	0.8352000000	0.6261343013	
NONE	RandomForest	smote	
NONE	0.8357000000	0.6185279777	Vaniance Threshel
d 0.8	RandomForest 0.8221000000	adasyn 0.6104663893	Variance Threshol
u 0.0	RandomForest	adasyn	RFECV RandomF
orest	0.8205000000	0.6086766950	n zev nandom
	RandomForest	adasyn	FWE f_cl
assif	0.8170000000	0.6129441624	_
	RandomForest	adasyn	
NONE	0.8221000000	0.6104663893	
	RandomForest	svmsmote	Variance Threshol
d 0.8	0.8331000000	0.6248595190	DEECV DandomE
orest	RandomForest 0.8327000000	svmsmote 0.6248037677	RFECV RandomF
OI ESC	RandomForest	svmsmote	FWE f_cl
assif	0.8321000000	0.6282931149	
	RandomForest	svmsmote	
NONE	0.8331000000	0.6248595190	
	RandomForest	smotetomek	Variance Threshol
d 0.8	0.8349000000	0.6177355869	
	RandomForest	smotetomek	RFECV RandomF
orest	0.8352000000	0.6185185185	THE 4 cl
assif	RandomForest 0.8365000000	smotetomek 0.6283246192	FWE f_cl
a3311	RandomForest	smotetomek	
NONE	0.8349000000	0.6177355869	
	RandomForest	RandomOversampling	Variance Threshol
d 0.8	0.8175000000	0.6160319798	
	RandomForest	RandomOversampling	RFECV RandomF
orest	0.8172000000	0.6140202703	
<u>:</u> C	RandomForest	RandomOversampling	FWE f_cl
assif	0.8161000000 RandomForest	0.6158345519	
	Natioumifultest	RandomOversampling	

NONE	0.8163000000	0.6138322472	
	RandomForest	NONE	Variance Threshol
d 0.8	0.8110000000 RandomForest	0.6078838174 NONE	RFECV RandomF
orest	0.8119000000	0.6103169671	KFECV KAHUOIIIF
0. 00 0	RandomForest	NONE	FWE f_cl
assif	0.8062000000	0.6088009689	
	RandomForest	NONE	
NONE	0.8110000000 SVC	0.6078838174 smote	Variance Threshol
d 0.8	0.7985000000	0.5936680782	variance infestion
u. 0.0	SVC	smote	RFECV RandomF
orest	0.7986000000	0.5939516129	
	SVC	smote	FWE f_cl
assif	0.8006000000	0.5991154001	
NONE	SVC 0.7985000000	smote 0.5936680782	
NONL	5.738300000 SVC	adasyn	Variance Threshol
d 0.8	0.7668000000	0.5724239091	va. 14.1.66 11.1 63.1.61
	SVC	adasyn	RFECV RandomF
orest	0.7668000000	0.5724239091	
	SVC	adasyn	FWE f_cl
assif	0.7672000000	0.5784136183	
NONE	SVC 0.766800000	adasyn 0.5724239091	
NONL	SVC	svmsmote	Variance Threshol
d 0.8	0.7984000000	0.5985663082	
	SVC	svmsmote	RFECV RandomF
orest	0.7991000000	0.5997210600	
<b>:</b> C	SVC	svmsmote	FWE f_cl
assif	0.7979000000 SVC	0.6001978239 svmsmote	
NONE	0.798400000	0.5985663082	
	SVC	smotetomek	Variance Threshol
d 0.8	0.7978000000	0.5928312525	
	SVC	smotetomek	RFECV RandomF
orest	0.7973000000	0.5915776748	EUE C -1
assif	SVC 0.7986000000	smotetomek 0.5957446809	FWE f_cl
a3311	5VC	smotetomek	
NONE	0.7978000000	0.5928312525	
	SVC	RandomOversampling	Variance Threshol
d 0.8	0.7942000000	0.5961538462	
	SVC	RandomOversampling	RFECV RandomF
orest	0.7934000000 SVC	0.5950607605 RandomOversampling	FWE f_cl
assif	0.7950000000	0.5997657165	LML 1_CI
	SVC	RandomOversampling	
NONE	0.7973000000	0.5999605289	
	SVC	NONE	Variance Threshol
d 0.8	0.8569000000	0.5196374622	DEECH D. L. T.
orest	SVC 0.8575000000	NONE 0.5235707121	RFECV RandomF
OI ESL	8.837300000 SVC	0.3233707121 NONE	FWE f_cl
assif	0.8581000000	0.5274725275	
	SVC	NONE	

```
In [90]: all_f1s = []
    for model_desc, sampler_desc, selector_desc, acc, f1 in all_results:
        all_f1s.append(f1)
    sorted_indices = np.argsort(all_f1s)[::-1]
    print("%20s %22s %25s %10s %10s" % ('ALGORITHM', 'SAMPLING', 'FEATURE-SELECTIO N', 'ACC', 'F1'))
    print("%20s %22s %25s %10s %10s" % ('-'*20, '-'*22, '-'*25, '-'*10, '-'*10))
    for i in sorted_indices:
        model_desc, sampler_desc, selector_desc, acc, f1 = all_results[i]
        print("%20s %22s %25s %10.5f %10.5f" % (model_desc, sampler_desc, selector_desc, acc, f1))
```

ACC	ALGORITHM F1	SAMPLING	FEATURE-SELECTION	
900	GradientBoosting 0.63023	svmsmote	Variance Threshold 0.8	0.83
	GradientBoosting	svmsmote	NONE	0.83
900	0.63023 GradientBoosting	svmsmote	RFECV RandomForest	0.83
850	0.62933 GradientBoosting	svmsmote	FWE f_classif	0.83
240	0.62904 RandomForest	smotetomek	FWE f_classif	0.83
650				
210	RandomForest 0.62829	svmsmote	FWE f_classif	0.83
150	GradientBoosting 0.62662	smotetomek	Variance Threshold 0.8	0.84
	GradientBoosting	smotetomek	NONE	0.84
150	0.62662 GradientBoosting	smotetomek	RFECV RandomForest	0.84
140	0.62630 RandomForest	smote	FWE f_classif	0.83
520	0.62613 GradientBoosting	adasyn	- RFECV RandomForest	0.84
160	0.62589	-		
140	GradientBoosting 0.62577	adasyn	NONE	0.84
140	GradientBoosting 0.62577	adasyn	Variance Threshold 0.8	0.84
580	GradientBoosting 0.62511	smotetomek	FWE f_classif	0.83
	RandomForest	svmsmote	Variance Threshold 0.8	0.83
310	0.62486 RandomForest	svmsmote	NONE	0.83
310	0.62486 GradientBoosting	smote	RFECV RandomForest	0.84
190	0.62348			
110	GradientBoosting 0.62337	smote	NONE	0.84
110	GradientBoosting 0.62337	smote	Variance Threshold 0.8	0.84
	RandomForest	svmsmote	RFECV RandomForest	0.83
200	0.62281 GradientBoosting	smote	FWE f_classif	0.83
400	0.62273 RandomForest	smotetomek	RFECV RandomForest	0.83
640	0.62059			
630	RandomForest 0.62027	smote	RFECV RandomForest	0.83
940	RandomForest 0.61883	RandomOversampling	RFECV RandomForest	0.81
	RandomForest	smote	NONE	0.83
570	0.61853 RandomForest	smote	Variance Threshold 0.8	0.83
570	0.61853 RandomForest	smotetomek	NONE	0.83
		55 CC COCK	.10112	

400	0 (1774				
490	0.61774 RandomFo	orest	smotetomek	Variance Threshold 0.8	0.83
490	0.61774 GradientBoos	sting	adasyn	FWE f_classif	0.82
390	0.61692	SCING	adasyn	1 WL 1_C143311	0.02
400	RandomForest 0 0.61395		RandomOversampling	FWE f_classif	0.81
	RandomFo	orest	adasyn	RFECV RandomForest	0.82
360	0.61299 RandomFo	nnast	adasyn	FWE f_classif	0.81
700	0.61294	טו פגנ	auasyn	LMT I_CTG2211	0.81
	RandomFo	orest	RandomOversampling	Variance Threshold 0.8	0.81
520	0.61193 RandomFo	orest	RandomOversampling	NONE	0.81
560	0.61113				0.02
	RandomFo	orest	adasyn	NONE	0.82
210	0.61047		- 4	Vaniana Thuashald O.O.	0.00
210	RandomFo 0.61047	orest	adasyn	Variance Threshold 0.8	0.82
210	RandomFo	nrest	NONE	RFECV RandomForest	0.81
260	0.60926	51 636	NONE	KI LEV Kandomi of esc	0.01
	RandomFo	orest	NONE	FWE f_classif	0.80
620	0.60880			_	
	GradientBoo	sting	RandomOversampling	NONE	0.80
410	0.60875		DandamOvanaamalina	DEECV DawdowEowest	0.00
490	GradientBoos 0.60815	sting	RandomOversampling	RFECV RandomForest	0.80
470	RandomFo	orest	NONE	Variance Threshold 0.8	0.81
100	0.60788				
	RandomFo	orest	NONE	NONE	0.81
100	0.60788	. •	D   0   1:	ENE C 1	0.00
280	GradientBoos 0.60686	sting	RandomOversampling	FWE f_classif	0.80
200	GradientBoos	sting	RandomOversampling	Variance Threshold 0.8	0.80
220	0.60487				0.00
		SVC	RandomOversampling	NONE	0.79
760	0.60063				
700	0 60020	SVC	svmsmote	FWE f_classif	0.79
790	0.60020	SVC	smote	FWE f_classif	0.80
060	0.59912	340	Sillote	FWL I_CIASSII	0.80
		SVC	svmsmote	NONE	0.79
840	0.59857				
		SVC	svmsmote	Variance Threshold 0.8	0.79
840	0.59857	61.46	D   0   1'	51.5 6 1 16	. =0
440	0.59797	SVC	RandomOversampling	FWE f_classif	0.79
440	0.55757	SVC	svmsmote	RFECV RandomForest	0.79
790	0.59717	3.0	371113111000	in zer nangem er ese	0.,5
		SVC	RandomOversampling	RFECV RandomForest	0.79
570	0.59696				
060	0 50574	SVC	smotetomek	FWE f_classif	0.79
860	0.59574	SVC	RandomOversampling	Variance Threshold 0.8	0.79
290	0.59559	300	Manaomover Sampiting	variance in conora 0.0	0.75
-		SVC	smote	NONE	0.79
850	0.59367				

		SVC	smote	Variance Threshold 0.8	0.79
850	0.59367	SVC	smotetomek	Variance Threshold 0.8	0.79
780	0.59283				
		SVC	smotetomek	NONE	0.79
780	0.59283	6) (6		DEFOV D   -	0.70
760	0.59243	SVC	smotetomek	RFECV RandomForest	0.79
700	0.33243	SVC	smote	RFECV RandomForest	0.79
720	0.59179	3.0	31110 CC	in zev nanaom or ese	0.,,
	GradientBoos	sting	NONE	Variance Threshold 0.8	0.86
480	0.58578				
	GradientBoos	sting	NONE	NONE	0.86
480	0.58578				
	GradientBoos	sting	NONE	FWE f_classif	0.86
390	0.58519				
	GradientBoos	sting	NONE	RFECV RandomForest	0.86
400	0.58154				
	0 57044	SVC	adasyn	FWE f_classif	0.76
720	0.57841	61.16		v · <del>-</del> 11100	0.76
600	0 57242	SVC	adasyn	Variance Threshold 0.8	0.76
680	0.57242	SVC	adasun	NONE	0.76
680	0.57242	SVC	adasyn	NONE	0.76
080	0.37242	SVC	adasyn	RFECV RandomForest	0.76
600	0.57127	310	adasyn	Ri Eev Randomi of ese	0.70
	013.1=1	SVC	NONE	FWE f_classif	0.85
810	0.52747			_	
		SVC	NONE	RFECV RandomForest	0.85
720	0.52336				
		SVC	NONE	Variance Threshold 0.8	0.85
690	0.51964				
		SVC	NONE	NONE	0.85
690	0.51964				

Here we can see that applying variance-threshold feature selection and no feature selection gave the exact same results. This is because no feature was eliminated.

```
In [99]: | selectors = [
             ('NONE', None),
         1
         for i in np.arange(0, 1.01, 0.05):
             selectors.append(('Variance Threshold %f' % i, lambda: VarianceThreshold(i
         * (1-i)),))
         model = lambda: GradientBoostingClassifier(random_state=0, max_depth=3, min_sa
         mples_split=42, n_estimators=100)
         all descs2 = ""
         all_results2 = []
         for selector_desc, selector in selectors:
             model_desc = "GradientBoosting"
             sampler_desc = "svmsmote"
             print('-'*80)
             print("EVALUATING: ", model_desc, sampler_desc, selector_desc)
             acc, f1 = model_compare_3(scaled_X.to_numpy(), y.to_numpy(),\
                                       SVMSMOTE(random_state=0), 'svmsmote',\
                                       model, 'gradient-boosting',\
                                       selector, selector_desc)
             print("%20s %30s %30s %20.10f %20.10f" % (model_desc, sampler_desc, select
         or desc, acc, f1))
             all_descs2 += "%20s %30s %30s %20.10f %20.10f" % (model_desc, sampler_desc
         , selector desc, acc, f1)
             all_descs2 += '\n'
             all_results2.append((model_desc, sampler_desc, selector_desc, acc, f1,))
             print(all_descs2)
```

	C 1: 15 1		. NONE			
	GradientBoost symsmote Mode	_		nσ		
besch iption.	precision					
0	0.91	0.88	0.90	7963		
1	0.59	0.67	0.63	2037		
			0.04	10000		
accuracy macro avg		0.79		10000		
	0.85					
mergineed avg	0.03	0.01	0.01	20000		
Gradient	Boosting 0.8390000000		_	vmsmote		
NONE				004 		
	GradientBoost	_			old 0.0000	900
Description:	svmsmote Mode	_		_		
	precision	recall	†1-score	support		
0	0.91	0.88	0.90	7963		
1			0.63			
accuracy				10000		
macro avg						
weighted avg	0.85	0.84	0.84	10000		
Gradient	_				Variance	Threshold 0.0
00000	0.8390000000		0.6302250	1804		
EVALUATING:	GradientBoost	ing svms	mote Varia	nce Thresh	old 0.050	900
	svmsmote Mode	_				
	precision	recall	f1-score	support		
0	0.91	0.88	0.90	7963		
1	0.59	0.67		2037		
_	0.33	0.07	0.03	2037		
accuracy			0.84	10000		
macro avg	0.75	0.78	0.76	10000		
weighted avg	0.85	0.84	0.84	10000		
Gradient	Boosting		S	vmsmote	Variance	Threshold 0.0
50000	0.8390000000		0.6302250	804		
 FVALIIATTNG•	GradientBoost	ing syms	mote Varia	nce Thrach	old a 1aa	300
	svmsmote Mode				OIG 0.100	
	precision	_		-		

precision recall f1-score support

	_					
	0		0.88			
	1	0.59	0.67	0.63	2037	
accurac	У			0.84		
macro av	g	0.75	0.78	0.76	10000	
weighted av	g	0.85	0.84	0.84	10000	
Gradien	tΒ	oosting		sv	msmote	Variance Threshold 0.1
00000		0.8390000000		0.63022508	804	
EVALUATING:		GradientBoost	ing svms	mote Varian	ice Thresh	old 0.150000
		svmsmote Mode	_			
		precision	_		_	
		p. cc2520		. 2 500. 0	зарро. с	
	0	a 91	0 88	0.90	7963	
	1			0.63		
	_	0.55	0.07	0.03	2037	
				0.04	10000	
accurac		o ==			10000	
macro av		0.75				
weighted av	g	0.85	0.84	0.84	10000	
 Cdi	<b>+</b> D					Vanianas Thuashald O 1
		oosting				Variance Threshold 0.1
50000		0.8390000000		0.63022508	304	
		GradientBoost	_			old 0.200000
		svmsmote Mode	l: gradi	ent-boostin	ng	old 0.200000
			l: gradi	ent-boostin	ng	old 0.200000
		svmsmote Mode	l: gradi	ent-boostin	ng	old 0.200000
Description		svmsmote Mode precision	l: gradi recall	ent-boostin	ng support	old 0.200000
Description	:	svmsmote Mode precision 0.91	l: gradi recall 0.88	ent-boostin f1-score	ng support	old 0.200000
Description	0	svmsmote Mode precision 0.91	l: gradi recall 0.88	ent-boostin f1-score 0.90	support 7963	old 0.200000
Description	): 0 1	svmsmote Mode precision 0.91	l: gradi recall 0.88	ent-boostin f1-score 0.90	7963 2037	old 0.200000
Description accurac	0 1	svmsmote Mode precision 0.91 0.59	1: gradi recall 0.88 0.67	ent-boostin f1-score 0.90 0.63 0.84	7963 2037	old 0.200000
Description  accurac  macro av	0 1 y	svmsmote Mode precision 0.91 0.59	1: gradi recall 0.88 0.67 0.78	ent-boostin f1-score 0.90 0.63 0.84 0.76	7963 2037 10000	old 0.200000
Description accurac	0 1 y	svmsmote Mode precision 0.91 0.59	1: gradi recall 0.88 0.67 0.78	ent-boostin f1-score 0.90 0.63 0.84 0.76	7963 2037 10000	old 0.200000
accurac macro av weighted av	0 1 1 7 8	svmsmote Mode precision 0.91 0.59 0.75 0.85	1: gradi recall 0.88 0.67 0.78 0.84	ent-boostin f1-score 0.90 0.63 0.84 0.76 0.84	7963 2037 10000 10000	old 0.200000
accurac macro av weighted av	0 1 1 7 8	svmsmote Mode precision 0.91 0.59 0.75 0.85	1: gradi recall 0.88 0.67 0.78 0.84	ent-boostin f1-score 0.90 0.63 0.84 0.76 0.84	7963 2037 10000 10000	
accurac macro av weighted av	0 1 ygg	svmsmote Mode precision 0.91 0.59 0.75 0.85	1: gradi recall 0.88 0.67 0.78 0.84	ent-boostin f1-score 0.90 0.63 0.84 0.76 0.84	19 support 7963 2037 10000 10000	
accurac macro av weighted av	0 1 7 7 7 8 7 8	symsmote Mode precision  0.91 0.59  0.75 0.85	1: gradi recall 0.88 0.67 0.78 0.84	ent-boostin f1-score 0.90 0.63 0.84 0.76 0.84	7963 2037 10000 10000 10000	
accurac macro av weighted av Gradien 00000	0 1 7 7 7 8 7 8 7 8	0.91 0.59 0.75 0.85 0.00sting 0.8390000000	1: gradi recall 0.88 0.67 0.78 0.84	ent-boostin f1-score 0.90 0.63 0.84 0.76 0.84	7963 2037 10000 10000 10000	Variance Threshold 0.2
accurac macro av weighted av Gradien 00000	0 1 7 7 7 8 7 8 7 8	0.91 0.59 0.75 0.85 0.00sting 0.8390000000	1: gradi recall 0.88 0.67 0.78 0.84	ent-boostin f1-score 0.90 0.63 0.84 0.76 0.84	7963 2037 10000 10000 10000	
accurac macro av weighted av Gradien 00000	0 1 'g 'g	0.91 0.59 0.75 0.85 	1: gradi recall 0.88 0.67 0.78 0.84	ent-boostin f1-score 0.90 0.63 0.84 0.76 0.84	7963 2037 10000 10000 10000	Variance Threshold 0.2
accurac macro av weighted av Gradien 00000	0 1 ygg 	0.91 0.59 0.75 0.85 	1: gradi recall 0.88 0.67 0.78 0.84	ent-boostin f1-score 0.90 0.63 0.84 0.76 0.84 sv 0.63022508	7963 2037 10000 10000 10000 	Variance Threshold 0.2
accurac macro av weighted av Gradien 00000	0 1 ygg 	ovmsmote Mode precision  0.91 0.59  0.75 0.85	1: gradi recall 0.88 0.67 0.78 0.84	ent-boostin f1-score 0.90 0.63 0.84 0.76 0.84 sv 0.63022508 mote Varian ent-boostin	7963 2037 10000 10000 10000	Variance Threshold 0.2
accurac macro av weighted av Gradien 00000	0 1 ygg 	0.91 0.59 0.75 0.85 	1: gradi recall 0.88 0.67 0.78 0.84	ent-boostin f1-score 0.90 0.63 0.84 0.76 0.84 sv 0.63022508 mote Varian ent-boostin	7963 2037 10000 10000 10000	Variance Threshold 0.2
accurac macro av weighted av Gradien 00000 EVALUATING:	0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	osting 0.8390000000 0.8390000000 0.83900000000000000000000000000000000000	1: gradi     recall	ent-boostin f1-score 0.90 0.63 0.84 0.76 0.84 sv 0.63022508 mote Varian ent-boostin f1-score	rg support 7963 2037 10000 10000 10000	Variance Threshold 0.2
accurac macro av weighted av Gradien 00000 EVALUATING:	0 1 ygg 	ovmsmote Mode precision  0.91 0.59  0.75 0.85   oosting 0.8390000000  GradientBoost svmsmote Mode precision  0.91	1: gradi recall 0.88 0.67 0.78 0.84 ing svms 1: gradi recall 0.88	ent-boostin f1-score 0.90 0.63 0.84 0.76 0.84 sv 0.63022508 mote Varian ent-boostin f1-score	7963 2037 10000 10000 10000 	Variance Threshold 0.2
accurac macro av weighted av Gradien 00000	0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	ovmsmote Mode precision  0.91 0.59  0.75 0.85   oosting 0.8390000000  GradientBoost svmsmote Mode precision  0.91	1: gradi     recall	ent-boostin f1-score 0.90 0.63 0.84 0.76 0.84 sv 0.63022508 mote Varian ent-boostin f1-score	7963 2037 10000 10000 10000 	Variance Threshold 0.2
accurac macro av weighted av Gradien 00000	0 1 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	ovmsmote Mode precision  0.91 0.59  0.75 0.85   oosting 0.8390000000  GradientBoost svmsmote Mode precision  0.91	1: gradi recall 0.88 0.67 0.78 0.84 ing svms 1: gradi recall 0.88	ent-boostin f1-score 0.90 0.63 0.84 0.76 0.84 sv 0.63022508 mote Varian ent-boostin f1-score	7963 2037 10000 10000 10000 	Variance Threshold 0.2
accurac macro av weighted av Gradien 00000 EVALUATING:	0 1 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	ovmsmote Mode precision  0.91 0.59  0.75 0.85   oosting 0.8390000000  GradientBoost svmsmote Mode precision  0.91	1: gradi recall 0.88 0.67 0.78 0.84 ing svms 1: gradi recall 0.88	ent-boostin f1-score 0.90 0.63 0.84 0.76 0.84 	7963 2037 10000 10000 10000 	Variance Threshold 0.2
accurac macro av weighted av Gradien 00000 EVALUATING: Description	0 1	osting 0.8390000000 0.8390000000 0.83900000000000000000000000000000000000	1: gradi recall 0.88 0.67 0.78 0.84 ing svms 1: gradi recall 0.88 0.67	ent-boostin f1-score 0.90 0.63 0.84 0.76 0.84 	7963 2037 10000 10000 10000 	Variance Threshold 0.2
accurac macro av weighted av Gradien 00000 EVALUATING:	: 0 1	ovmsmote Mode precision  0.91 0.59  0.75 0.85   oosting 0.8390000000  GradientBoost symsmote Mode precision  0.91 0.59  0.75	1: gradi recall 0.88 0.67 0.78 0.84 ing svms 1: gradi recall 0.88 0.67	ent-boostin f1-score 0.90 0.63 0.84 0.76 0.84 sv 0.63022508 mote Varian ent-boostin f1-score 0.90 0.63 0.84 0.76	7963 2037 10000 10000 10000 10000 	Variance Threshold 0.2

Gradient 50000	Boosting 0.8390000000		s\ 0.63022508		Variance	Threshold 0.2
 EVALUATING:	GnadiantRoost:	ina syms	moto Vaniar	oco Thnoch	old 0 3000	200
	GradientBoost: svmsmote Model	_			1010 0.300k	900
eser ipeion.	precision	_		-		
	p. cc1510	· ccull	. 2 500. 0	зарро. с		
0	0.91	0.88	0.90	7963		
1	0.59	0.67	0.63	2037		
accuracy			0.84	10000		
macro avg	0.75	0.78	0.76	10000		
weighted avg	0.85	0.84	0.84	10000		
 C					\/	Th
Gradient 10000	0.8390000000		s\ 0.63022508		variance	Threshold 0.
	C	•			-14 0 250	200
	GradientBoost	_			1010 0.3500	<i>9</i> 00
pescription:	svmsmote Mode:					
	precision	recarr	T1-Score	Support		
0	a 91	0 88	0.90	7963		
1	0.59					
-	0.55	0.07	0.03	2037		
accuracy			0.84	10000		
macro avg	0.75	0.78				
weighted avg	0.85	0.84	0.84	10000		
 Gradient	Boosting		S۱	/msmote	Variance	Threshold 0.
50000	0.8390000000		0.63022508	304		
	GradientBoost:	•			old 0.4000	900
Description:	svmsmote Mode	_		•		
	precision	recall	f1-score	support		
0	A 01	ρ 00	0.90	7062		
0 1			0.90 0.63			
1	ود.ه	0.07	6.03	2037		
accuracy			0 Q1	10000		
macro avg	0.75	0 72		10000		
veighted avg						
· · ·						
					\/	Thus a hald O
Gradient	BOOSTING		S۱	/msmote	variance	Threshold 0.

------

---

EVALUATING:	GradientBoost	ing svms	mote Variar	ce Thresh	old 0.450000
	svmsmote Mode	_			
·	precision	_		-	
0	0.91	0.88	0.90	7963	
1	0.59		0.63		
_	0,000	• • • • • • • • • • • • • • • • • • • •	0.00		
accuracy			0.84	10000	
macro avg		0.78			
weighted avg		0.84	0.84	10000	
Gradient	_				Variance Threshold 0.4
50000	0.8390000000			804	
FVALUATING.	GradientBoost:	ing syms	mote Variar	ce Thresh	old 0 500000
	svmsmote Model	•			014 0.300000
эсэс: _рс_с	precision	_		•	
0	0.91	0.88	0.90	7963	
1	0.59	0.67	0.63	2037	
accuracy			0.84		
macro avg		0.78			
weighted avg	0.85	0.84	0.84	10000	
 GradientI	Boosting		SV	rmsmote	Variance Threshold 0.5
Gradient 00000			sv 0.63022508	rmsmote	
Gradient 00000	Boosting 0.8390000000		sv 0.63022508	rmsmote	
Gradient 00000 	Boosting 0.8390000000		sv 0.63022508 	msmote 04	Variance Threshold 0.5
Gradient 00000  EVALUATING:	Boosting  0.8390000000  GradientBoost: svmsmote Model	ing svms l: gradi	sv 0.63022508  mote Variar ent-boostir	rmsmote 804  ice Thresh	Variance Threshold 0.5
Gradient 00000  EVALUATING:	Boosting 0.8390000000GradientBoost	ing svms l: gradi	sv 0.63022508  mote Variar ent-boostir	rmsmote 804  ice Thresh	Variance Threshold 0.5
Gradient 00000  EVALUATING: Description:	Boosting 0.8390000000 GradientBoosts svmsmote Modes precision	ing svms l: gradi recall	sv 0.63022508  mote Variar ent-boostir f1-score	rmsmote 804  ice Thresh ig support	Variance Threshold 0.5
Gradient 00000  EVALUATING: Description:	Boosting 0.8390000000 GradientBoost: svmsmote Model precision 0.91	ing svms l: gradi recall 0.88	sv 0.63022508  mote Variar ent-boostir f1-score 0.90	rmsmote 804  ace Thresh g support 7963	Variance Threshold 0.5
Gradient 00000  EVALUATING: Description:	Boosting 0.8390000000 GradientBoosts svmsmote Modes precision	ing svms l: gradi recall 0.88	sv 0.63022508  mote Variar ent-boostir f1-score 0.90	rmsmote 804  ace Thresh g support 7963	Variance Threshold 0.5
Gradient 00000  EVALUATING: Description:	Boosting 0.8390000000 GradientBoosts symsmote Modes precision 0.91 0.59	ing svms l: gradi recall 0.88	0.63022508  mote Variar ent-boostir f1-score 0.90 0.63	rmsmote 304 dee Thresh g support 7963 2037	Variance Threshold 0.5
Gradientl 00000 EVALUATING: Description:	Boosting 0.8390000000 GradientBoosts symsmote Modes precision 0.91 0.59	ing svms l: gradi recall 0.88 0.67	8 v 0.63022508 	rmsmote 04  ice Thresh g support 7963 2037	Variance Threshold 0.5
Gradient  00000 EVALUATING: Description:  0 1 accuracy macro avg	Boosting 0.8390000000 GradientBoost: svmsmote Model precision 0.91 0.59	ing svms l: gradi recall 0.88 0.67	0.63022508 mote Variar ent-boostir f1-score  0.90 0.63  0.84 0.76	rmsmote 804 ace Thresh g support 7963 2037 10000 10000	Variance Threshold 0.5
Gradient  00000 EVALUATING: Description:  0 1 accuracy macro avg	Boosting 0.8390000000 GradientBoosts symsmote Modes precision 0.91 0.59	ing svms l: gradi recall 0.88 0.67	8 v 0.63022508 	rmsmote 804 ace Thresh g support 7963 2037 10000 10000	Variance Threshold 0.5
Gradient  00000 EVALUATING: Description:  0 1 accuracy macro avg	Boosting 0.8390000000 GradientBoost: svmsmote Model precision 0.91 0.59	ing svms l: gradi recall 0.88 0.67	0.63022508 mote Variar ent-boostir f1-score  0.90 0.63  0.84 0.76	rmsmote 804 ace Thresh g support 7963 2037 10000 10000	Variance Threshold 0.5
Gradient  00000 EVALUATING: Description:  0 1 accuracy macro avg	Boosting 0.8390000000 GradientBoost: svmsmote Model precision 0.91 0.59	ing svms l: gradi recall 0.88 0.67	0.63022508 mote Variar ent-boostir f1-score  0.90 0.63  0.84 0.76	rmsmote 804 ace Thresh g support 7963 2037 10000 10000	Variance Threshold 0.5
Gradient  00000 EVALUATING: Description:  0 1 accuracy macro avg	GradientBoostisymsmote Model precision  0.91 0.59	ing svms l: gradi recall 0.88 0.67	0.63022508 mote Variar ent-boostir f1-score  0.90 0.63  0.84 0.76 0.84	7msmote 804 1ce Thresh 1g support 7963 2037 10000 10000	Variance Threshold 0.5
Gradient  00000 EVALUATING: Description:  0 1 accuracy macro avg weighted avg	GradientBoostisymsmote Model precision  0.91 0.59	ing svms l: gradi recall 0.88 0.67 0.78 0.84	0.63022508 mote Variar ent-boostir f1-score 0.90 0.63 0.84 0.76 0.84	rmsmote 304 ace Thresh g support 7963 2037 10000 10000	Variance Threshold 0.5
Gradient  00000 EVALUATING: Description:  0 1 accuracy macro avg weighted avg Gradient	Boosting  0.8390000000  GradientBoost: svmsmote Mode: precision  0.91  0.59  0.75  0.85	ing svms l: gradi recall 0.88 0.67 0.78 0.84	0.63022508 mote Variar ent-boostir f1-score 0.90 0.63 0.84 0.76 0.84	rmsmote 304 	Variance Threshold 0.5  old 0.550000  Variance Threshold 0.5
Gradientl 00000 EVALUATING: Description:  0 1 accuracy macro avg weighted avg Gradientl 50000	GradientBoostis symsmote Model precision  0.91 0.59  0.75 0.85  Boosting 0.8390000000	ing svms l: gradi recall 0.88 0.67 0.78 0.84	0.63022508 mote Variar ent-boostir f1-score  0.90 0.63  0.84 0.76 0.84  SN 0.63022508	rmsmote 804 	Variance Threshold 0.5  old 0.550000  Variance Threshold 0.5
Gradientl 00000 EVALUATING: Description:  0 1 accuracy macro avg weighted avg Gradientl 50000 EVALUATING:	GradientBoost: svmsmote Mode: precision  0.91 0.59  0.75 0.85  Boosting 0.8390000000  GradientBoost:	ing svms l: gradi recall 0.88 0.67 0.78 0.84	0.63022508 mote Variar ent-boostir f1-score 0.90 0.63 0.84 0.76 0.84	rmsmote 304  ace Thresh g support 7963 2037 10000 10000 10000	Variance Threshold 0.5  old 0.550000  Variance Threshold 0.5
Gradientl 00000 EVALUATING: Description:  0 1 accuracy macro avg weighted avg Gradientl 50000 EVALUATING:	GradientBoost: svmsmote Mode: precision  0.91 0.59  0.75 0.85  Boosting 0.8390000000  GradientBoost: svmsmote Mode:	ing svms l: gradi recall 0.88 0.67 0.78 0.84	mote Variar ent-boostir f1-score  0.90 0.63  0.84 0.76 0.84	msmote dot ce Thresh support 7963 2037 10000 10000 10000	Variance Threshold 0.5  old 0.550000  Variance Threshold 0.5
Gradientl 00000 EVALUATING: Description:  0 1 accuracy macro avg weighted avg Gradientl 50000 EVALUATING:	GradientBoost: svmsmote Mode: precision  0.91 0.59  0.75 0.85  Boosting 0.8390000000  GradientBoost:	ing svms l: gradi recall 0.88 0.67 0.78 0.84	mote Variar ent-boostir f1-score  0.90 0.63  0.84 0.76 0.84	msmote dot ce Thresh support 7963 2037 10000 10000 10000	Variance Threshold 0.5  old 0.550000  Variance Threshold 0.5
Gradientl 00000 EVALUATING: Description:  0 1 accuracy macro avg weighted avg Gradientl 50000 EVALUATING:	GradientBoost: svmsmote Mode: precision  0.91 0.59  0.75 0.85  Boosting 0.8390000000  GradientBoost: svmsmote Mode:	ing svms l: gradi recall 0.88 0.67 0.78 0.84	mote Variar ent-boostin f1-score  0.63022508  0.63  0.84  0.76  0.84   SV  0.63022508  mote Variar ent-boostin f1-score	msmote dot ce Thresh support 7963 2037 10000 10000 10000	Variance Threshold 0.5  old 0.550000  Variance Threshold 0.5

1	0.59	0.67	0.63	2037		
	0.75 0.85		0.76			
weighted avg	0.05	0.04	0.04	10000		
C	D+				Novi	Thursday 0.6
	Boosting 0.8390000000			304	variance	Threshold 0.6
					110 650	
	GradientBoost symsmote Mode	_			010 0.650	<i>3</i> 00
•	precision					
0	0.91	0.88	0.90	7963		
1	0.59	0.67	0.63	2037		
accuracy			0 8/1	10000		
	0.75	0.78				
	0.85					
Gradient					Variance	Threshold 0.6
50000	0.8390000000		0.63022508	804		
	GradientBoost				old 0.700	900
	svmsmote Mode	l: gradi	ent-boosti	ng	old 0.700	<b>300</b>
		l: gradi	ent-boosti	ng	old 0.700	900
Description:	svmsmote Mode precision 0.91	l: gradi recall 0.88	ent-boostin f1-score 0.90	ng support 7963	old 0.700	900
Description:	svmsmote Mode precision 0.91	l: gradi recall 0.88	ent-boostin f1-score	ng support 7963	old 0.700	900
Description:	svmsmote Mode precision 0.91 0.59	l: gradi recall 0.88	ent-boostin f1-score 0.90	ng support 7963 2037	old 0.700	900
Description:  0 1 accuracy macro avg	symsmote Mode precision 0.91 0.59	1: gradi recall 0.88 0.67	ent-boostin f1-score 0.90 0.63 0.84 0.76	ng support 7963 2037 10000 10000	old 0.700	900
Description: 0 1 accuracy	symsmote Mode precision 0.91 0.59	1: gradi recall 0.88 0.67	ent-boostin f1-score 0.90 0.63 0.84 0.76	ng support 7963 2037 10000 10000	old 0.700	900
Description:  0 1 accuracy macro avg weighted avg	symsmote Mode precision 0.91 0.59	1: gradi recall 0.88 0.67 0.78 0.84	ent-boostin f1-score 0.90 0.63 0.84 0.76 0.84	19000 10000		
Description:  0 1 accuracy macro avg weighted avg	svmsmote Mode precision  0.91 0.59  0.75 0.85	1: gradi recall 0.88 0.67 0.78 0.84	ent-boostin f1-score 0.90 0.63 0.84 0.76 0.84	ng support 7963 2037 10000 10000		
Description:  0 1 accuracy macro avg weighted avg  Gradient	svmsmote Mode precision  0.91 0.59  0.75 0.85	1: gradi recall 0.88 0.67 0.78 0.84	ent-boostin f1-score 0.90 0.63 0.84 0.76 0.84	rig support 7963 2037 10000 10000 10000		
Description:  0 1 accuracy macro avg weighted avg  Gradient	svmsmote Mode precision  0.91 0.59  0.75 0.85	1: gradi recall 0.88 0.67 0.78 0.84	ent-boostin f1-score 0.90 0.63 0.84 0.76 0.84	rig support 7963 2037 10000 10000 10000		
Description:  0 1 accuracy macro avg weighted avg  Gradientl 00000	svmsmote Mode precision  0.91 0.59  0.75 0.85  Boosting 0.8390000000	1: gradi recall 0.88 0.67 0.78 0.84	ent-boostin f1-score 0.90 0.63 0.84 0.76 0.84	19 support 7963 2037 10000 10000 10000 wmsmote 804	Variance	Threshold 0.7
Description:  0 1 accuracy macro avg weighted avg  Gradient 00000 EVALUATING:	svmsmote Mode precision  0.91 0.59  0.75 0.85  Boosting 0.8390000000  GradientBoost svmsmote Mode	1: gradi recall 0.88 0.67 0.78 0.84	ent-boostin f1-score 0.90 0.63 0.84 0.76 0.84	ng support 7963 2037 10000 10000 10000 vmsmote 804 nce Thresh	Variance	Threshold 0.7
Description:  0 1 accuracy macro avg weighted avg  Gradient 00000 EVALUATING:	symsmote Mode precision  0.91 0.59  0.75 0.85  Boosting 0.8390000000 GradientBoost	1: gradi recall 0.88 0.67 0.78 0.84	ent-boostin f1-score 0.90 0.63 0.84 0.76 0.84	ng support 7963 2037 10000 10000 10000 vmsmote 804 nce Thresh	Variance	Threshold 0.7
Description:  0 1 accuracy macro avg weighted avg  Gradient 00000 EVALUATING:	svmsmote Mode precision  0.91 0.59  0.75 0.85  Boosting 0.8390000000 GradientBoost svmsmote Mode precision	1: gradi     recall	ent-boostin f1-score 0.90 0.63 0.84 0.76 0.84	support 7963 2037 10000 10000 10000 vmsmote 804 nce Thresh	Variance	Threshold 0.7
Description:  0 1 accuracy macro avg weighted avg Gradient 00000 EVALUATING: Description:	svmsmote Mode precision  0.91 0.59  0.75 0.85  Boosting 0.8390000000 GradientBoost svmsmote Mode precision 0.91	1: gradi recall 0.88 0.67 0.78 0.84 ing svms 1: gradi recall 0.88	ent-boosting f1-score  0.90 0.63  0.84 0.76 0.84  sv 0.63022508 mote Variangent-boosting f1-score	support 7963 2037 10000 10000 10000 vmsmote 804 nce Thresh ng support 7963	Variance	Threshold 0.7
Description:  0 1 accuracy macro avg weighted avg  Gradient 00000 EVALUATING: Description:	svmsmote Mode precision  0.91 0.59  0.75 0.85  Boosting 0.8390000000  GradientBoost svmsmote Mode precision  0.91 0.59	1: gradi recall 0.88 0.67 0.78 0.84 ing svms 1: gradi recall 0.88	ent-boosting f1-score  0.90	ng support 7963 2037 10000 10000 10000 	Variance	Threshold 0.7
Description:  0 1 accuracy macro avg weighted avg  Gradientl 00000 EVALUATING: Description:  0 1 accuracy	svmsmote Mode precision  0.91 0.59  0.75 0.85  Boosting 0.8390000000 GradientBoost svmsmote Mode precision  0.91 0.59	1: gradi recall 0.88 0.67 0.78 0.84 ing svms 1: gradi recall 0.88 0.67	ent-boosting f1-score  0.90	rig support 7963 2037 10000 10000 10000 	Variance	Threshold 0.7
Description:  0 1 accuracy macro avg weighted avg  Gradient 00000 EVALUATING: Description:	svmsmote Mode precision  0.91 0.59  0.75 0.85  Boosting 0.8390000000 GradientBoost svmsmote Mode precision  0.91 0.59  0.75	1: gradi recall 0.88 0.67 0.78 0.84 ing svms 1: gradi recall 0.88	ent-boosting f1-score  0.90	rig support 7963 2037 10000 10000 10000 	Variance	Threshold 0.7

Gradient 50000	Boosting 0.8390000000				Variance	Threshold	0.7
 EVALUATING:	GradientBoosti	ng syms	mote Varian	ce Thresh	old 0 8000	200	
	svmsmote Mode	_			0.000	500	
•	precision	_		•			
0			0.90				
1	0.59	0.67	0.63	2037			
accuracy			0.84	10000			
macro avg		0.78					
weighted avg	0.85	0.84	0.84	10000			
Gradient	Boosting		sv	msmote	Variance	Threshold	0.8
00000	0.8390000000		0.63022508	04			
 EVALUATING:	GradientBoosti	na syms	mote Varian	ca Thrach	old 0 8500	200	
	svmsmote Mode				OIU 6.656	900	
beset iperon.	precision	_		-			
	p. 55225						
0	0.91	0.88	0.90				
1	0.59	0.67	0.63	2037			
			0.04	10000			
accuracy macro avg		0 70		10000 10000			
weighted avg							
	0.02						
	D + +				\/	Th	0 0
Gradient 50000	0.8390000000				variance	Threshold	0.8
	GradientBoosti				old 0.9000	900	
Description:	svmsmote Model	_		•			
	precision	recall	f1-score	support			
0	a 91	0 88	0.90	7963			
1			0.63				
_	0.55	0.07	0.05	2037			
accuracy			0.84	10000			
macro avg				10000			
weighted avg	0.85	0.84	0.84	10000			
Gradient	Boosting		SV	msmote	Variance	Threshold	0.9
00000	0.8390000000						

---

EVALUATING: GradientBoosting symsmote Variance Threshold 0.950000 Description: symsmote Model: gradient-boosting

		GradientBoost	_			old 0.9500	900
Descripti	ion:	svmsmote Mode	_		_		
		precision	recall	†1-score	support		
	0	0.91	0.88	0.90	7963		
	1	0.59	0.67	0.63	2037		
accur	_			0.84			
macro	_						
weighted	avg	0.85	0.84	0.84	10000		
		Boosting 0.8390000000				Variance	Threshold 0.9
	uc.	Condinate	:		Thus ah	-14 1 000	200
		GradientBoost: svmsmote Model	_			010 1.0000	<i>3</i> 00
besci ipci	LUII.	precision	_		-		
		·					
	0			0.90			
	1	0.59	0.67	0.63	2037		
accur	racv			0.84	10000		
		0.75	0.78				
weighted	avg	0.85	0.84	0.84	10000		
 C:d:	: 4- 1					Vaniana	Thursdald 1 O
Grad:		Boosting 0.8390000000				variance	Threshold 1.0
		Boosting			/msmote		
NONE	LCITCE	0.8390000000		0.630225086			
Gradi	ientE	Boosting		S۱	/msmote	Variance	Threshold 0.0
00000		0.8390000000		0.63022508	304		
	ientE	Boosting			/msmote	Variance	Threshold 0.0
50000		0.8390000000		0.63022508			
	ient	Boosting			/msmote	Variance	Threshold 0.1
00000 Gradi	i ant [	0.8390000000 Boosting		0.63022508	304 /msmote	Vaniance	Threshold 0.1
50000	Lenci	0.8390000000		0.63022508		vai Tailce	1111 E31101u 0.1
	ientE	Boosting			/msmote	Variance	Threshold 0.2
00000		0.8390000000		0.63022508	304		
Gradi	ientE	Boosting		S۱	/msmote	Variance	Threshold 0.2
50000		0.8390000000		0.63022508	304		
	ient	Boosting		_	/msmote	Variance	Threshold 0.3
00000	: an+r	0.839000000		0.63022508	_	Vaniance	Throshold 0 2
50000	renti	Boosting 0.8390000000		0.63022508	/msmote	variance	Threshold 0.3
	ientF	Boosting			/msmote	Variance	Threshold 0.4
00000		0.8390000000		0.63022508		vai zaiice	2511014 011
	ientE	Boosting			/msmote	Variance	Threshold 0.4
50000		0.8390000000		0.63022508	304		
	ientE	Boosting		_	/msmote	Variance	Threshold 0.5
00000	•	0.8390000000		0.63022508		., .	TI 13165
Gradi	ıentE	Boosting		S۱	/msmote	variance	Threshold 0.5

```
50000
              0.8390000000
                                   0.6302250804
    GradientBoosting
                                            svmsmote
                                                        Variance Threshold 0.6
              0.8390000000
                                   0.6302250804
00000
    GradientBoosting
                                            svmsmote
                                                        Variance Threshold 0.6
              0.8390000000
50000
                                   0.6302250804
                                                        Variance Threshold 0.7
    GradientBoosting
                                            svmsmote
00000
              0.8390000000
                                   0.6302250804
                                                        Variance Threshold 0.7
    GradientBoosting
                                            svmsmote
              0.8390000000
                                   0.6302250804
50000
                                            svmsmote
                                                        Variance Threshold 0.8
    GradientBoosting
              0.8390000000
00000
                                   0.6302250804
    GradientBoosting
                                                        Variance Threshold 0.8
                                            svmsmote
              0.8390000000
50000
                                   0.6302250804
                                                        Variance Threshold 0.9
    GradientBoosting
                                            svmsmote
              0.8390000000
                                   0.6302250804
00000
                                                        Variance Threshold 0.9
    GradientBoosting
                                            svmsmote
50000
              0.8390000000
                                   0.6302250804
                                                        Variance Threshold 1.0
    GradientBoosting
                                            svmsmote
              0.8390000000
                                   0.6302250804
00000
```

```
In [180]:
          from imblearn.ensemble import RUSBoostClassifier
          def search_rus_boost(X, y):
              rusbc = RUSBoostClassifier(random state=0)
              p = [('rusbc', rusbc,)]
              p = imblearn.pipeline.Pipeline(p)
              param_grid = {
                   'rusbc n estimators': [1000, 10000],
                   'rusbc__learning_rate': [1, 0.1, 0.01, 0.001],
                   'rusbc__algorithm': ['SAMME', 'SAMME.R'],
              }
              scorer = make_scorer(metrics.f1_score)
              cv1 = StratifiedKFold(n_splits=6, random_state=10, shuffle=True)
              search = GridSearchCV(p, param grid, n jobs=7, cv=cv1, scoring=scorer)
              search.fit(X, y)
              print("GRADIENT BOOSTING")
              print(f"Best Params = {search.best params }")
              print(f"Best score = {search.best_score_}")
              print('-' * 120)
              return search.best estimator
          rusboost = search_rus_boost(scaled_X.to_numpy(), y.to_numpy())
```

```
GRADIENT BOOSTING
Best Params = {'rusbc__algorithm': 'SAMME.R', 'rusbc__learning_rate': 0.001,
'rusbc__n_estimators': 10000}
Best score = 0.5837437094271175
```

```
In [181]: from imblearn.ensemble import RUSBoostClassifier
          def search_rus_boost(X, y):
              rusbc = RUSBoostClassifier(random state=0)
              p = [('rusbc', rusbc,)]
              p = imblearn.pipeline.Pipeline(p)
              param_grid = {
                  'rusbc__n_estimators': [10000, 100000],
                  'rusbc__learning_rate': [0.001, 0.0001],
                  'rusbc__algorithm': ['SAMME.R'],
              }
              scorer = make_scorer(metrics.f1_score)
              cv1 = StratifiedKFold(n_splits=6, random_state=10, shuffle=True)
              search = GridSearchCV(p, param_grid, n_jobs=7, cv=cv1, scoring=scorer)
              search.fit(X, y)
              print("GRADIENT BOOSTING")
              print(f"Best Params = {search.best_params_}")
              print(f"Best score = {search.best_score_}")
              print('-' * 120)
              return search.best estimator
          rusboost = search_rus_boost(scaled_X.to_numpy(), y.to_numpy())
          GRADIENT BOOSTING
          Best Params = {'rusbc__algorithm': 'SAMME.R', 'rusbc__learning_rate': 0.001,
          'rusbc n estimators': 100000}
          Best score = 0.5847328429438647
```

In [ ]: