

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from tqdm import tqdm
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.model_selection import StratifiedKFold
from sklearn.tree import DecisionTreeClassifier
from sklearn import metrics
from sklearn.metrics import confusion_matrix
import xgboost as xgb
import warnings
warnings.filterwarnings('ignore')
import plotly.offline as offline
import plotly.graph_objs as go
offline.init_notebook_mode()
```

```
In [2]: csv = pd.read_csv('musk_csv.csv')
```

```
In [3]: csv.shape
```

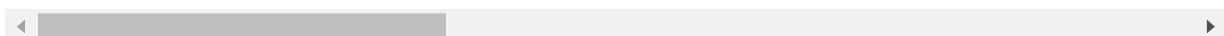
```
Out[3]: (6598, 170)
```

```
In [4]: csv.describe()
```

```
Out[4]:
```

| | ID | f1 | f2 | f3 | f4 | f5 | f |
|-------|------------|-------------|-------------|-------------|-------------|-------------|-------------|
| count | 6598.00000 | 6598.000000 | 6598.000000 | 6598.000000 | 6598.000000 | 6598.000000 | 6598.000000 |
| mean | 3299.50000 | 58.945135 | -119.128524 | -73.146560 | -0.628372 | -103.533495 | 18.359800 |
| std | 1904.82287 | 53.249007 | 90.813375 | 67.956235 | 80.444617 | 64.387559 | 80.59365 |
| min | 1.00000 | -31.000000 | -199.000000 | -167.000000 | -114.000000 | -118.000000 | -183.00000 |
| 25% | 1650.25000 | 37.000000 | -193.000000 | -137.000000 | -70.000000 | -117.000000 | -28.00000 |
| 50% | 3299.50000 | 44.000000 | -149.000000 | -99.000000 | -25.000000 | -117.000000 | 33.00000 |
| 75% | 4948.75000 | 53.000000 | -95.000000 | -19.000000 | 42.000000 | -116.000000 | 74.00000 |
| max | 6598.00000 | 292.000000 | 95.000000 | 81.000000 | 161.000000 | 325.000000 | 200.00000 |

8 rows × 168 columns



```
In [5]: csv.head()
```

Out[5]:

| | ID | molecule_name | conformation_name | f1 | f2 | f3 | f4 | f5 | f6 | f7 | ... | f158 | f159 | f1 |
|---|----|---------------|-------------------|----|------|------|-----|------|----|----|-----|------|------|----|
| 0 | 1 | MUSK-211 | 211_1+1 | 46 | -108 | -60 | -69 | -117 | 49 | 38 | ... | -308 | 52 | |
| 1 | 2 | MUSK-211 | 211_1+10 | 41 | -188 | -145 | 22 | -117 | -6 | 57 | ... | -59 | -2 | |
| 2 | 3 | MUSK-211 | 211_1+11 | 46 | -194 | -145 | 28 | -117 | 73 | 57 | ... | -134 | -154 | |
| 3 | 4 | MUSK-211 | 211_1+12 | 41 | -188 | -145 | 22 | -117 | -7 | 57 | ... | -60 | -4 | |
| 4 | 5 | MUSK-211 | 211_1+13 | 41 | -188 | -145 | 22 | -117 | -7 | 57 | ... | -60 | -4 | |

5 rows × 170 columns

```
In [6]: csv['class'].value_counts()
```

Out[6]: 0 5581
1 1017
Name: class, dtype: int64

Data is Highly imbalance. Approx Ratio is 5:1

```
In [7]: # checking weather any value is missing or not
for col in csv.columns:
    print(csv[col].isnull().value_counts())
```

```
False    6598
Name: ID, dtype: int64
False    6598
Name: molecule_name, dtype: int64
False    6598
Name: conformation_name, dtype: int64
False    6598
Name: f1, dtype: int64
False    6598
Name: f2, dtype: int64
False    6598
Name: f3, dtype: int64
False    6598
Name: f4, dtype: int64
False    6598
Name: f5, dtype: int64
False    6598
Name: f6, dtype: int64
False    6598
Name: f7, dtype: int64
False    6598
Name: f8, dtype: int64
False    6598
Name: f9, dtype: int64
False    6598
Name: f10, dtype: int64
False    6598
Name: f11, dtype: int64
False    6598
Name: f12, dtype: int64
False    6598
Name: f13, dtype: int64
False    6598
Name: f14, dtype: int64
False    6598
Name: f15, dtype: int64
False    6598
Name: f16, dtype: int64
False    6598
Name: f17, dtype: int64
False    6598
Name: f18, dtype: int64
False    6598
Name: f19, dtype: int64
False    6598
Name: f20, dtype: int64
False    6598
Name: f21, dtype: int64
False    6598
Name: f22, dtype: int64
False    6598
Name: f23, dtype: int64
False    6598
Name: f24, dtype: int64
False    6598
Name: f25, dtype: int64
False    6598
```

```
Name: f26, dtype: int64
False    6598
Name: f27, dtype: int64
False    6598
Name: f28, dtype: int64
False    6598
Name: f29, dtype: int64
False    6598
Name: f30, dtype: int64
False    6598
Name: f31, dtype: int64
False    6598
Name: f32, dtype: int64
False    6598
Name: f33, dtype: int64
False    6598
Name: f34, dtype: int64
False    6598
Name: f35, dtype: int64
False    6598
Name: f36, dtype: int64
False    6598
Name: f37, dtype: int64
False    6598
Name: f38, dtype: int64
False    6598
Name: f39, dtype: int64
False    6598
Name: f40, dtype: int64
False    6598
Name: f41, dtype: int64
False    6598
Name: f42, dtype: int64
False    6598
Name: f43, dtype: int64
False    6598
Name: f44, dtype: int64
False    6598
Name: f45, dtype: int64
False    6598
Name: f46, dtype: int64
False    6598
Name: f47, dtype: int64
False    6598
Name: f48, dtype: int64
False    6598
Name: f49, dtype: int64
False    6598
Name: f50, dtype: int64
False    6598
Name: f51, dtype: int64
False    6598
Name: f52, dtype: int64
False    6598
Name: f53, dtype: int64
False    6598
Name: f54, dtype: int64
```

```
False    6598
Name: f55, dtype: int64
False    6598
Name: f56, dtype: int64
False    6598
Name: f57, dtype: int64
False    6598
Name: f58, dtype: int64
False    6598
Name: f59, dtype: int64
False    6598
Name: f60, dtype: int64
False    6598
Name: f61, dtype: int64
False    6598
Name: f62, dtype: int64
False    6598
Name: f63, dtype: int64
False    6598
Name: f64, dtype: int64
False    6598
Name: f65, dtype: int64
False    6598
Name: f66, dtype: int64
False    6598
Name: f67, dtype: int64
False    6598
Name: f68, dtype: int64
False    6598
Name: f69, dtype: int64
False    6598
Name: f70, dtype: int64
False    6598
Name: f71, dtype: int64
False    6598
Name: f72, dtype: int64
False    6598
Name: f73, dtype: int64
False    6598
Name: f74, dtype: int64
False    6598
Name: f75, dtype: int64
False    6598
Name: f76, dtype: int64
False    6598
Name: f77, dtype: int64
False    6598
Name: f78, dtype: int64
False    6598
Name: f79, dtype: int64
False    6598
Name: f80, dtype: int64
False    6598
Name: f81, dtype: int64
False    6598
Name: f82, dtype: int64
False    6598
```

```
Name: f83, dtype: int64
False    6598
Name: f84, dtype: int64
False    6598
Name: f85, dtype: int64
False    6598
Name: f86, dtype: int64
False    6598
Name: f87, dtype: int64
False    6598
Name: f88, dtype: int64
False    6598
Name: f89, dtype: int64
False    6598
Name: f90, dtype: int64
False    6598
Name: f91, dtype: int64
False    6598
Name: f92, dtype: int64
False    6598
Name: f93, dtype: int64
False    6598
Name: f94, dtype: int64
False    6598
Name: f95, dtype: int64
False    6598
Name: f96, dtype: int64
False    6598
Name: f97, dtype: int64
False    6598
Name: f98, dtype: int64
False    6598
Name: f99, dtype: int64
False    6598
Name: f100, dtype: int64
False   6598
Name: f101, dtype: int64
False   6598
Name: f102, dtype: int64
False   6598
Name: f103, dtype: int64
False   6598
Name: f104, dtype: int64
False   6598
Name: f105, dtype: int64
False   6598
Name: f106, dtype: int64
False   6598
Name: f107, dtype: int64
False   6598
Name: f108, dtype: int64
False   6598
Name: f109, dtype: int64
False   6598
Name: f110, dtype: int64
False   6598
Name: f111, dtype: int64
```

```
False    6598
Name: f112, dtype: int64
False    6598
Name: f113, dtype: int64
False    6598
Name: f114, dtype: int64
False    6598
Name: f115, dtype: int64
False    6598
Name: f116, dtype: int64
False    6598
Name: f117, dtype: int64
False    6598
Name: f118, dtype: int64
False    6598
Name: f119, dtype: int64
False    6598
Name: f120, dtype: int64
False    6598
Name: f121, dtype: int64
False    6598
Name: f122, dtype: int64
False    6598
Name: f123, dtype: int64
False    6598
Name: f124, dtype: int64
False    6598
Name: f125, dtype: int64
False    6598
Name: f126, dtype: int64
False    6598
Name: f127, dtype: int64
False    6598
Name: f128, dtype: int64
False    6598
Name: f129, dtype: int64
False    6598
Name: f130, dtype: int64
False    6598
Name: f131, dtype: int64
False    6598
Name: f132, dtype: int64
False    6598
Name: f133, dtype: int64
False    6598
Name: f134, dtype: int64
False    6598
Name: f135, dtype: int64
False    6598
Name: f136, dtype: int64
False    6598
Name: f137, dtype: int64
False    6598
Name: f138, dtype: int64
False    6598
Name: f139, dtype: int64
False    6598
```

```
Name: f140, dtype: int64
False    6598
Name: f141, dtype: int64
False    6598
Name: f142, dtype: int64
False    6598
Name: f143, dtype: int64
False    6598
Name: f144, dtype: int64
False    6598
Name: f145, dtype: int64
False    6598
Name: f146, dtype: int64
False    6598
Name: f147, dtype: int64
False    6598
Name: f148, dtype: int64
False    6598
Name: f149, dtype: int64
False    6598
Name: f150, dtype: int64
False    6598
Name: f151, dtype: int64
False    6598
Name: f152, dtype: int64
False    6598
Name: f153, dtype: int64
False    6598
Name: f154, dtype: int64
False    6598
Name: f155, dtype: int64
False    6598
Name: f156, dtype: int64
False    6598
Name: f157, dtype: int64
False    6598
Name: f158, dtype: int64
False    6598
Name: f159, dtype: int64
False    6598
Name: f160, dtype: int64
False    6598
Name: f161, dtype: int64
False    6598
Name: f162, dtype: int64
False    6598
Name: f163, dtype: int64
False    6598
Name: f164, dtype: int64
False    6598
Name: f165, dtype: int64
False    6598
Name: f166, dtype: int64
False    6598
Name: class, dtype: int64
```

```
In [8]: Y = csv['class']
```

```
In [9]: X = csv.drop(['class','molecule_name','conformation_name','ID'],axis=1)
```

```
In [10]: # splitting the data  
x_train, x_test, y_train, y_test = train_test_split(X,Y,test_size=.2,stratify=Y)
```

```
In [11]: def GridSearchCV(x_train,y_train,classifier,folds):
    ''' Takes input data , XGB classifier and number of crossvalidation returns
    the List of train and validation
        auc score calculated by average auc for each fold '''
    # defining variables for storing auc score of training and cross-validation d
    ata
    train_auc_scores = []
    cv_auc_scores     = []

    # iteration over each pair of hyper-parameter
    for base_learner in [150,250,350,500]:
        for depth in [1,2,3,4,5]:
            # defining temporary variables for storing auc_score of training and cros_vali
            dation data for each fold
            train_auc_score_for_avg      = []
            cv_auc_score_for_avg         = []
            # creating indices for each fold and iteration over each fold
            folds_indices    = StratifiedKFold(n_splits=folds,shuffle=True)
            for train_ind, cv_ind in folds_indices.split(x_train, y_train):
                train_x = x_train.iloc[train_ind]
                train_y = y_train.iloc[train_ind]
                cv_x   = x_train.iloc[cv_ind]
                cv_y   = y_train.iloc[cv_ind]
            # Assigning hyper-parameter to model or classifier and fitting it on training d
            ata
                classifier.max_depth = depth
                classifier.n_estimators = base_learner
                classifier.fit(train_x,train_y)
            # Predicting log_probabilities for training and cv data and calculating roc_auc
            _score and storing it in temporary variables
                predictions = classifier.predict(train_x)
                fpr, tpr, thresholds = metrics.roc_curve(train_y,predictions,
                pos_label=1)
                train_auc_score_for_avg.append(metrics.auc(fpr, tpr) )
                predictions = classifier.predict(cv_x)
                fpr, tpr, thresholds = metrics.roc_curve(cv_y,predictions, pos
                _label=1)
                cv_auc_score_for_avg.append( metrics.auc(fpr, tpr))
            # Averaging the roc_auc_score values of all folds and storing it in final vari
            able
                train_auc_scores.append(np.mean(train_auc_score_for_avg))
                cv_auc_scores.append(np.mean(cv_auc_score_for_avg))

    # Returing the values
    return train_auc_scores, cv_auc_scores
```

```
In [12]: def plot_curve(train_auc,cv_auc):
    """ 3-D plot """
    X = [150,150,150,150,150,250,250,250,250,250,350,350,350,350,350,500,500,500,500]
    Y = [1,2,3,4,5,1,2,3,4,5,1,2,3,4,5,1,2,3,4,5]
    trace1 = go.Scatter3d(x=X,y=Y,z=train_auc, name = 'train')
    trace2 = go.Scatter3d(x=X,y=Y,z=cv_auc, name = 'Cross validation')
    data = [trace1, trace2]
    layout = go.Layout(scene = dict(
        xaxis = dict(title='n_estimators'),
        yaxis = dict(title='max_depth'),
        zaxis = dict(title='AUC')))
    fig = go.Figure(data=data, layout=layout)
    offline.iplot(fig, filename='3d-scatter-colorscale')
```

```
In [46]: # PERforming hyerr param tunning  
t_sc_1,cv_sc_1= GridSearchCV(x_train,y_train,xgb.XGBRegressor(n_jobs=-1),3)
```

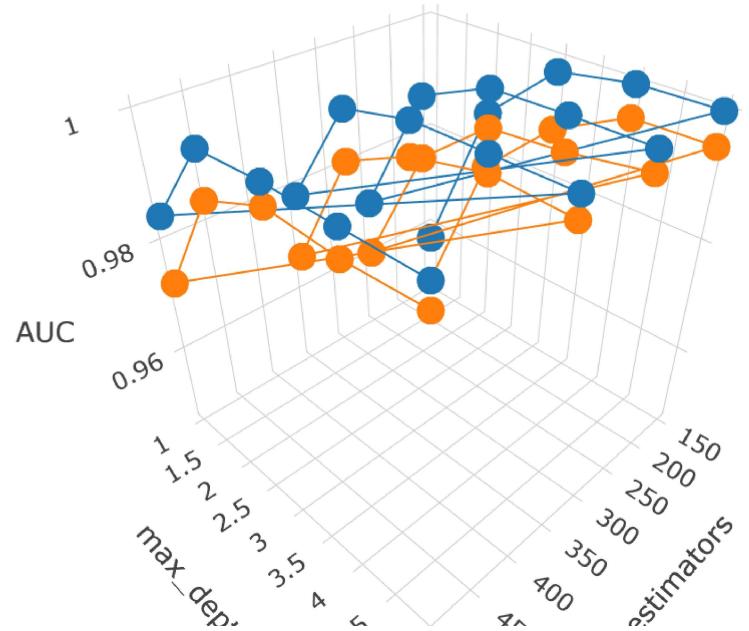
```
[00:23:29] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:square_error.  
[00:23:33] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:square_error.  
[00:23:36] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:square_error.  
[00:23:38] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:square_error.  
[00:23:41] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:square_error.  
[00:23:43] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:square_error.  
[00:23:46] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:square_error.  
[00:23:49] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:square_error.  
[00:23:52] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:square_error.  
[00:23:56] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:square_error.  
[00:23:59] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:square_error.  
[00:24:03] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:square_error.  
[00:24:08] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:square_error.  
[00:24:12] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:square_error.  
[00:24:17] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:square_error.  
[00:24:22] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:square_error.  
[00:24:25] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:square_error.  
[00:24:28] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:square_error.  
[00:24:31] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:square_error.
```

```
[00:24:35] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:24:39] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:24:43] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:24:48] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:24:53] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:24:59] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:25:05] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:25:12] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:25:19] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:25:27] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:25:35] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:25:43] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:25:47] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:25:51] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:25:55] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:26:02] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:26:07] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:26:12] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:26:20] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.
```

```
[00:26:28] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:26:35] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:26:44] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:26:54] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:27:03] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:27:17] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:27:31] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:27:43] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:27:48] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:27:54] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:28:01] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:28:10] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:28:19] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:28:27] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:28:38] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:28:48] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:28:59] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:29:12] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.  
[00:29:25] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squared_error.
```

```
[00:29:38] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squarederror.  
[00:29:54] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squarederror.  
[00:30:11] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squarederror.
```

In [51]: `# plot for aucs
plot_curve(t_sc_1, cv_sc_1)`



BEST Hyperparam

1. n_estimator:350
2. max_depth :4

```
In [13]: # Defining classifier with best hyper param found  
gbdt_classifier = xgb.XGBRegressor(n_estimators=350,max_depth=4,n_jobs=-1)
```

```
In [14]: # Training  
gbdt_classifier.fit(x_train,y_train)
```

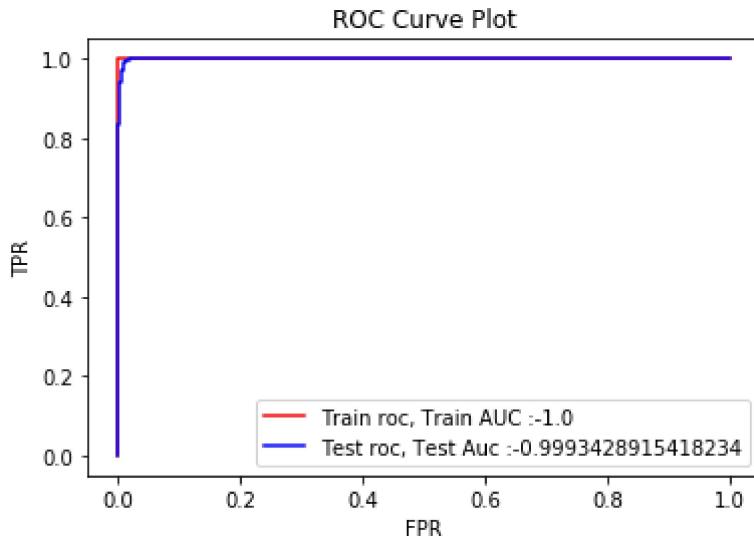
[00:49:56] WARNING: C:/Jenkins/workspace/xgboost-win64_release_0.90/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squarederror.

```
Out[14]: XGBRegressor(base_score=0.5, booster='gbtree', colsample_bylevel=1,  
colsample_bynode=1, colsample_bytree=1, gamma=0,  
importance_type='gain', learning_rate=0.1, max_delta_step=0,  
max_depth=4, min_child_weight=1, missing=None, n_estimators=350,  
n_jobs=-1, nthread=None, objective='reg:linear', random_state=0,  
reg_alpha=0, reg_lambda=1, scale_pos_weight=1, seed=None,  
silent=None, subsample=1, verbosity=1)
```

```
In [15]: # predicting probabilities for class Label 1  
prediction_train = gbdt_classifier.predict(x_train)  
prediction_test = gbdt_classifier.predict(x_test)
```

```
In [16]: # calculating fpr and tpr values for train and test both  
train_fpr, train_tpr, train_thresholds = metrics.roc_curve(y_train,prediction_train)  
test_fpr, test_tpr, test_thresholds = metrics.roc_curve(y_test,prediction_test)
```

```
In [17]: # Roc curve
plt.plot(train_fpr,train_tpr,color='r',label='Train roc, Train AUC :-'+str(np.trapz(train_tpr,train_fpr)))
plt.plot(test_fpr,test_tpr,color='b',label='Test roc, Test Auc :-'+str(np.trapz(test_tpr,test_fpr)))
plt.title('ROC Curve Plot')
plt.xlabel('FPR')
plt.ylabel('TPR')
plt.legend()
plt.show()
```



```
In [18]: # Finding threshold value which yields maximum auc score
threshold=train_thresholds[list(train_tpr-train_fpr).index(max(train_tpr-train_fpr))]
```

```
In [19]: # converting prediction probability into class by taking threshold as 0.5
pred_class_train = [ 0 if prediction_train[i] < threshold else 1 for i in range(len(prediction_train))]
pred_class_test = [ 0 if prediction_test[i] < threshold else 1 for i in range(len(prediction_test))]
```

```
In [20]: # calculating values for confusion matrix
tn_1,fp_1,fn_1,tp_1=confusion_matrix (y_test,pred_class_test).ravel()
# Printing Confusion Matrix
print(f'''Confusion Matrix
```

| | | Predicted | | |
|----------------------------|---|-----------|--------|-----|
| | | 0 | 1 | |
| A c t u a l | 0 | {tn_1} | {fp_1} | |
| | 1 | {fn_1} | {tp_1} | |
| | | | | ... |
| | | | |) |
| | | | | |
| | | | | |

Confusion Matrix

| | | Predicted | | |
|----------------------------|---|-----------|-----|--|
| | | 0 | 1 | |
| A c t u a l | 0 | 1104 | 13 | |
| | 1 | 2 | 201 | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |