In [1]:

```
import dill
#dill.dump_session('notebook_env.db')
#dill.load_session('notebook_env.db')

#https://stackoverflow.com/questions/34342155/how-to-pickle-or-store-jupyter-ipython-no
tebook-session-for-later
#https://www.reddit.com/r/IPython/comments/6reiqp/how_can_i_save_and_load_the_state_of_
the_kernel/dl6f2yn/
```

In [2]:

```
import pandas as pd
import numpy as np

import warnings
import math
import datetime
```

In [3]:

```
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import confusion_matrix

from sklearn import tree
from sklearn.neighbors import KNeighborsClassifier
from sklearn import svm
from sklearn.neural_network import MLPClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.dummy import DummyClassifier

import matplotlib as mt

import pickle

warnings.filterwarnings('ignore')
```

In [4]:

```
from imblearn.over_sampling import SMOTE
#from imblearn.under_sampling import NeighbourhoodCleaningRule
from imblearn.under_sampling import RandomUnderSampler
```

In [5]:

```
%matplotlib notebook
```

In [6]:

```
i_clf_tree  = 1
i_clf_knn   = 2
i_clf_svm   = 3
i_clf_mlp   = 4
i_clf_naive = 5
i_clf_dummy = 6
nomes_algs = ['Tree', 'KNN', 'SVM', 'MLP', 'Naive', 'Dummy']

def ObterAlgoritmoClf(tp_algoritmo):
    if tp_algoritmo == i_clf_tree:
        return tree.DecisionTreeClassifier(criterion='gini', max_depth=5)
    elif tp_algoritmo == i_clf_knn:
        return KNeighborsClassifier(n_neighbors=3)
    elif tp_algoritmo == i_clf_svm:
        return svm.SVC(C=1.0, kernel='sigmoid')
    elif tp_algoritmo == i_clf_mlp:
        return MLPClassifier(hidden_layer_sizes=100, activation='relu')
    elif tp_algoritmo == i_clf_naive:
        return GaussianNB()
    elif tp_algoritmo == i_clf_dummy:
        return DummyClassifier(strategy='prior')
    else:
        return None

def ObterMatrizConfusao(clf, features_teste, target_teste):
    cls_predict = clf.predict(features_teste)
    mat_conf = confusion_matrix(target_teste, cls_predict)
    score = clf.score(features_teste, cls_predict)
    return mat_conf

def PrepararLista(lista):
    listaCopy = lista.copy()
    modelo = {  }
    for col in list(lista.columns.values):
        print(col)
        if (lista.dtypes[col] == 'object'):
            col_le = LabelEncoder()

            col_labels = col_le.fit_transform(lista[col])
            col_mapping = {index: label for index, label in enumerate(col_le.classes_)}

            modelo[col] = col_mapping
            listaCopy[col] = col_labels
        else:
            modelo[col] =  'tipado'

    return modelo, listaCopy

def DM_Calcular(mod_matriz_confusao, ret):
    matriz_confusao = mod_matriz_confusao

    VP = matriz_confusao[0,0]
    FP = matriz_confusao[0,1]
    FN = matriz_confusao[1,0]
    VN = matriz_confusao[1,1]

    sensibilidade  = VP/(VP+FN)*100
    especificidade = VN/(VN+FP)*100
    if np.isnan(sensibilidade):
```

```
        sensibilidade = 0
    if np.isnan(especificidade):
        especificidade = 0

    acuracia        = ((VP+VN)/(VP+VN+FP+FN))*100
    #fi              = (VP*VN - FP*FN) / Math.sqrt((VP + FP)*(VP + FN)*(VN + FP)*(VN + F
N))
    VPP             = (VP/(VP+FP))*100
    VPN             = (VN/(VN+FN))*100
    eficiencia      = (sensibilidade+especificidade)/2
    totalP          = VP+FP
    totalN          = VN+FN
    totalVPFN       = VP+FN
    totalFPVN       = FP+VN
    totalG          = VP+FP+FN+VN

    if ret == 'sens':
        return round(sensibilidade, 2)
    elif ret == 'esp':
        return round(especificidade, 2)
    elif ret == 'acur':
        return round(acuracia, 2)
    elif ret == 'VPP':
        return round(VPP, 2)
    elif ret == 'VPN':
        return round(VPN, 2)
    elif ret == 'efic':
        return round(eficiencia, 2)
    elif ret == 'totP':
        return totalP
    elif ret == 'totN':
        return totalN
    elif ret == 'totG':
        return totalG
    else:
        return -1
```

In [7]:

```
colsExibir = ['id_alg', 'algoritmo', 'acur', 'sens', 'esp', 'efic', 'VPP', 'VPN', 'TExe
c' ]
metr_alg = ['sens', 'esp', 'acur', 'VPP', 'VPN', 'efic', 'totP', 'totN', 'totG' ]
cols = ['id_alg', 'algoritmo', 'acur', 'sens', 'esp', 'efic', 'VPP', 'VPN', 'mat_conf',
'AlgBin', 'TExec' ]
colsExibirMin = ['id_alg', 'algoritmo', 'acur', 'sens', 'esp', 'efic', 'TExec' ]
```

In [8]:

```
def ExibirMedidas(X_train, X_test, y_train, y_test):
    algs = []
    for idx in range(1, 7):
        dtIni = datetime.datetime.now()
        print(dtIni)

        d = dict(id_alg=idx, algoritmo=nomes_algs[idx - 1])
        print(d)

        clf = ObterAlgoritmoClf(idx)
        # analisar o resultado depois retirar filtro
        #if idx != i_clf_tree:
        #    continue

        clf.fit(X_train, y_train)
        d['AlgBin'] = clf

        mat_conf = ObterMatrizConfusao(clf, X_test, y_test)
        for mtr in metr_alg:
            d[mtr] = DM_Calcular(mat_conf, mtr)
        d['mat_conf'] = mat_conf

        dtFim = datetime.datetime.now()
        dtDiff = dtFim - dtIni

        d['TExec'] = round(dtDiff.total_seconds() / 60, 2)

        print(clf)
        algs.append(d)

    dfAlg = pd.DataFrame(algs)
    dfAlg = dfAlg.fillna(0)

    print(datetime.datetime.now())

    return dfAlg[cols]

def ReExibirMedidas(dfAlg2, X_test, y_test):
    lstdfAlg = dfAlg2.T.to_dict()
    algs = []
    print('ReExibirMedidas')
    for idx in range(1, 7):
        row = lstdfAlg[idx-1]

        print(row['algoritmo'])
        clf = row['AlgBin']
        dtIni = datetime.datetime.now()
        print(dtIni)

        mat_conf = ObterMatrizConfusao(clf, X_test, y_test)
        for mtr in metr_alg:
            row[mtr] = DM_Calcular(mat_conf, mtr)

        #d['Modelo'] = pickle.dumps(clf)

        print(clf)
        dtFim = datetime.datetime.now()
        dtDiff = dtFim - dtIni
```

```python
        row['TExec'] = round(dtDiff.total_seconds() / 60, 2)

        algs.append(row)

    dfAlg = pd.DataFrame(algs)
    dfAlg = dfAlg.fillna(0)

    return dfAlg[cols]
```

In [9]:

```python
def GerarResampling(tipo, X, y):
    if tipo == 'over':
        sm = SMOTE(random_state=42)
        return sm.fit_resample(X, y)
    elif tipo == 'under':
        #cnn = NeighbourhoodCleaningRule()
        cnn = RandomUnderSampler(random_state=42)
        return cnn.fit_resample(X, y)
```

In [10]:

```python
def ExibirDesbanciamento(y):
    dfDesb = pd.DataFrame({'col': y})
    ig_verd = dfDesb['col'][ dfDesb['col'] == 1 ].size
    ig_fals = dfDesb['col'][ dfDesb['col'] == 0 ].size
    lstGrf = []
    dicv = dict()
    dicv['Atr'] = 'Certificado'
    dicv['Qtd'] = ig_verd
    lstGrf.append(dicv)
    dicf = dict()
    dicf['Atr'] = 'Não Certificado'
    dicf['Qtd'] = ig_fals
    lstGrf.append(dicf)

    if ig_verd > ig_fals:
        print(round(ig_fals / ig_verd, 2))
    else:
        print(round(ig_verd / ig_fals, 2))

    return pd.DataFrame(lstGrf)
```

In [11]:

```python
path_arq_dir = r'D:\Dados\bstoll\Documents\SuperOneNotes\jupyter-notebook\mooc-dataset'
path_arq =  path_arq_dir + '\\' + r'HMXPC13_DI_v2_5-14-14.csv'

df = pd.read_csv(path_arq)
```

In [12]:

```
df.head(5)
```

Out[12]:

	course_id	userid_DI	registered	viewed	explored	certified	fina
0	HarvardX/CB22x/2013_Spring	MHxPC130442623	1	0	0	0	
1	HarvardX/CS50x/2012	MHxPC130442623	1	1	0	0	
2	HarvardX/CB22x/2013_Spring	MHxPC130275857	1	0	0	0	
3	HarvardX/CS50x/2012	MHxPC130275857	1	0	0	0	
4	HarvardX/ER22x/2013_Spring	MHxPC130275857	1	0	0	0	

In [13]:

```
df.describe()
```

Out[13]:

	registered	viewed	explored	certified	YoB	neve
count	641138.0	641138.000000	641138.000000	641138.000000	544533.000000	441987.000
mean	1.0	0.624299	0.061899	0.027587	1985.253279	431.008
std	0.0	0.484304	0.240973	0.163786	8.891814	1516.116
min	1.0	0.000000	0.000000	0.000000	1931.000000	1.000
25%	1.0	0.000000	0.000000	0.000000	1982.000000	3.000
50%	1.0	1.000000	0.000000	0.000000	1988.000000	24.000
75%	1.0	1.000000	0.000000	0.000000	1991.000000	158.000
max	1.0	1.000000	1.000000	1.000000	2013.000000	197757.000

In [14]:

```
df.dtypes
```

Out[14]:

```
course_id              object
userid_DI              object
registered              int64
viewed                  int64
explored                int64
certified               int64
final_cc_cname_DI      object
LoE_DI                 object
YoB                   float64
gender                 object
grade                  object
start_time_DI          object
last_event_DI          object
nevents               float64
ndays_act             float64
nplay_video           float64
nchapters             float64
nforum_posts            int64
roles                 float64
incomplete_flag       float64
dtype: object
```

In [15]:

```
classe = 'certified'
print(classe)
```

```
certified
```

In [16]:

```
#df['gender'] = df['gender'].fillna('')
#df['roles']  = df['roles'].fillna(0)
#df['YoB']    = df['YoB'].fillna(0)
```

In [17]:

```
df['Q_ndays_act']    = pd.qcut(df['ndays_act'],   4,duplicates='drop')
df['Q_nplay_video']  = pd.qcut(df['nplay_video'], 4,duplicates='drop')
df['Q_nchapters']    = pd.qcut(df['nchapters'],   4,duplicates='drop')
df['Q_nforum_posts'] = pd.qcut(df['nforum_posts'],4,duplicates='drop')
```

In [18]:

```
del df['nevents']
del df['ndays_act']
del df['nplay_video']
del df['nchapters']
del df['nforum_posts']

df = df.rename(columns=
          {'Q_nevents':      'nevents',
           'Q_ndays_act':    'ndays_act',
           'Q_nplay_video':  'nplay_video',
           'Q_nchapters':    'nchapters',
           'Q_nforum_posts': 'nforum_posts',
          })

#start_time_DI
#last_event_DI
#del df['Unnamed']
```

In [19]:

```
df['start_year'] = pd.DatetimeIndex(df['start_time_DI']).year
df['start_month'] = pd.DatetimeIndex(df['start_time_DI']).month

df['last_e_year'] = pd.DatetimeIndex(df['last_event_DI']).year
df['last_e_month'] = pd.DatetimeIndex(df['last_event_DI']).month
```

In [20]:

```
del df['start_time_DI']
del df['last_event_DI']
del df['userid_DI']
#del df['Unnamed: 0']
```

In [21]:

```
# diminuir a eficiencia dos algoritmos
del df['nplay_video']
del df['nforum_posts']
del df['nchapters']
del df['ndays_act']
del df['incomplete_flag']
del df['grade']
del df['explored']
#del df['age']
```

In [22]:

```
for col in df.columns:
    if col == classe:
        next
    df[col] = df[col].fillna(0)
    print('=====================================================================')
    print(col)
    print(df[col].unique())
```

```
======================================================================
course_id
['HarvardX/CB22x/2013_Spring' 'HarvardX/CS50x/2012'
 'HarvardX/ER22x/2013_Spring' 'HarvardX/PH207x/2012_Fall'
 'HarvardX/PH278x/2013_Spring' 'MITx/6.002x/2012_Fall'
 'MITx/6.002x/2013_Spring' 'MITx/14.73x/2013_Spring'
 'MITx/2.01x/2013_Spring' 'MITx/3.091x/2012_Fall'
 'MITx/3.091x/2013_Spring' 'MITx/6.00x/2012_Fall' 'MITx/6.00x/2013_Spring'
 'MITx/7.00x/2013_Spring' 'MITx/8.02x/2013_Spring'
 'MITx/8.MReV/2013_Summer']
======================================================================
registered
[1]
======================================================================
viewed
[0 1]
======================================================================
certified
[0 1]
======================================================================
final_cc_cname_DI
['United States' 'France' 'Unknown/Other' 'Mexico' 'Australia' 'India'
 'Canada' 'Russian Federation' 'Other South Asia'
 'Other North & Central Amer., Caribbean' 'Other Europe' 'Other Oceania'
 'Japan' 'Other Africa' 'Colombia' 'Germany'
 'Other Middle East/Central Asia' 'Poland' 'Indonesia' 'Other East Asia'
 'Bangladesh' 'China' 'United Kingdom' 'Ukraine' 'Spain' 'Greece'
 'Pakistan' 'Brazil' 'Nigeria' 'Egypt' 'Other South America' 'Portugal'
 'Philippines' 'Morocco']
======================================================================
LoE_DI
[0 'Secondary' "Bachelor's" "Master's" 'Doctorate' 'Less than Secondary']
======================================================================
YoB
[   0. 2012. 1987. 1968. 1989. 1978. 1993. 1988. 1981. 1980. 1991. 1977.
 1992. 1990. 1986. 1984. 1982. 1983. 1979. 1994. 1967. 1969. 1985. 1971.
 1973. 1974. 1995. 1972. 1976. 1965. 1963. 1964. 1975. 1955. 1944. 1966.
 1957. 1997. 2000. 1960. 1970. 1996. 1959. 1961. 1953. 1952. 1956. 1962.
 1958. 1999. 1945. 2011. 1954. 1947. 1948. 1998. 1950. 1949. 1951. 1940.
 1936. 1941. 1942. 2010. 2008. 2002. 1937. 2001. 1946. 1939. 1938. 2009.
 1943. 1935. 2007. 2003. 1931. 1934. 2013.]
======================================================================
gender
[0 'm' 'f' 'o']
======================================================================
roles
[0.]
======================================================================
start_year
[2012 2013]
======================================================================
start_month
[12 10  2  9  1  6  7  3  8  4  5 11]
======================================================================
last_e_year
[2013.    0. 2012.]
======================================================================
last_e_month
[11.  0.  5.  3.  6.  8.  1. 12.  4.  2.  7.  9. 10.]
```

In [23]:

```
#======================================================================
#
# exibir dados agrupados para ver corelação do atributo com a classe
#
#======================================================================

for col in df.columns:
    if col == classe:
        next

    dfTmpGrp = df.groupby([col, classe]).agg(['count'])
    print(dfTmpGrp)
    print('======================================================================')
```

\

course_id	certified	registered count	viewed count	final_cc_cname_DI count
HarvardX/CB22x/2013_Spring	0	29618	29618	29618
	1	384	384	384
HarvardX/CS50x/2012	0	168334	168334	168334
	1	1287	1287	1287
HarvardX/ER22x/2013_Spring	0	55060	55060	55060
	1	2346	2346	2346
HarvardX/PH207x/2012_Fall	0	39750	39750	39750
	1	1842	1842	1842
HarvardX/PH278x/2013_Spring	0	38891	38891	38891
	1	711	711	711
MITx/14.73x/2013_Spring	0	25785	25785	25785
	1	2085	2085	2085
MITx/2.01x/2013_Spring	0	5418	5418	5418
	1	247	247	247
MITx/3.091x/2012_Fall	0	13583	13583	13583
	1	632	632	632
MITx/3.091x/2013_Spring	0	6001	6001	6001
	1	138	138	138
MITx/6.002x/2012_Fall	0	39061	39061	39061
	1	1750	1750	1750
MITx/6.002x/2013_Spring	0	21642	21642	21642
	1	593	593	593
MITx/6.00x/2012_Fall	0	64254	64254	64254
	1	2477	2477	2477
MITx/6.00x/2013_Spring	0	56462	56462	56462
	1	1253	1253	1253
MITx/7.00x/2013_Spring	0	20186	20186	20186
	1	823	823	823
MITx/8.02x/2013_Spring	0	30226	30226	30226
	1	822	822	822
MITx/8.MReV/2013_Summer	0	9180	9180	9180
	1	297	297	297

course_id	certified	LoE_DI count	YoB count	gender count	roles count \
HarvardX/CB22x/2013_Spring	0	29618	29618	29618	29618
	1	384	384	384	384
HarvardX/CS50x/2012	0	168334	168334	168334	168334
	1	1287	1287	1287	1287
HarvardX/ER22x/2013_Spring	0	55060	55060	55060	55060
	1	2346	2346	2346	2346
HarvardX/PH207x/2012_Fall	0	39750	39750	39750	39750
	1	1842	1842	1842	1842
HarvardX/PH278x/2013_Spring	0	38891	38891	38891	38891
	1	711	711	711	711
MITx/14.73x/2013_Spring	0	25785	25785	25785	25785
	1	2085	2085	2085	2085
MITx/2.01x/2013_Spring	0	5418	5418	5418	5418
	1	247	247	247	247
MITx/3.091x/2012_Fall	0	13583	13583	13583	13583
	1	632	632	632	632
MITx/3.091x/2013_Spring	0	6001	6001	6001	6001
	1	138	138	138	138
MITx/6.002x/2012_Fall	0	39061	39061	39061	39061
	1	1750	1750	1750	1750
MITx/6.002x/2013_Spring	0	21642	21642	21642	21642

```
                             1            593     593     593     593
MITx/6.00x/2012_Fall         0          64254   64254   64254   64254
                             1           2477    2477    2477    2477
MITx/6.00x/2013_Spring       0          56462   56462   56462   56462
                             1           1253    1253    1253    1253
MITx/7.00x/2013_Spring       0          20186   20186   20186   20186
                             1            823     823     823     823
MITx/8.02x/2013_Spring       0          30226   30226   30226   30226
                             1            822     822     822     822
MITx/8.MReV/2013_Summer      0           9180    9180    9180    9180
                             1            297     297     297     297

                                       start_year start_month last_e_year
\
                                            count       count       count
course_id                    certified
HarvardX/CB22x/2013_Spring   0              29618       29618       29618
                             1                384         384         384
HarvardX/CS50x/2012          0             168334      168334      168334
                             1               1287        1287        1287
HarvardX/ER22x/2013_Spring   0              55060       55060       55060
                             1               2346        2346        2346
HarvardX/PH207x/2012_Fall    0              39750       39750       39750
                             1               1842        1842        1842
HarvardX/PH278x/2013_Spring  0              38891       38891       38891
                             1                711         711         711
MITx/14.73x/2013_Spring      0              25785       25785       25785
                             1               2085        2085        2085
MITx/2.01x/2013_Spring       0               5418        5418        5418
                             1                247         247         247
MITx/3.091x/2012_Fall        0              13583       13583       13583
                             1                632         632         632
MITx/3.091x/2013_Spring      0               6001        6001        6001
                             1                138         138         138
MITx/6.002x/2012_Fall        0              39061       39061       39061
                             1               1750        1750        1750
MITx/6.002x/2013_Spring      0              21642       21642       21642
                             1                593         593         593
MITx/6.00x/2012_Fall         0              64254       64254       64254
                             1               2477        2477        2477
MITx/6.00x/2013_Spring       0              56462       56462       56462
                             1               1253        1253        1253
MITx/7.00x/2013_Spring       0              20186       20186       20186
                             1                823         823         823
MITx/8.02x/2013_Spring       0              30226       30226       30226
                             1                822         822         822
MITx/8.MReV/2013_Summer      0               9180        9180        9180
                             1                297         297         297

                                       last_e_month
                                              count
course_id                    certified
HarvardX/CB22x/2013_Spring   0                29618
                             1                  384
HarvardX/CS50x/2012          0               168334
                             1                 1287
HarvardX/ER22x/2013_Spring   0                55060
                             1                 2346
HarvardX/PH207x/2012_Fall    0                39750
                             1                 1842
HarvardX/PH278x/2013_Spring  0                38891
```

```
                           1                  711
MITx/14.73x/2013_Spring    0                25785
                           1                 2085
MITx/2.01x/2013_Spring     0                 5418
                           1                  247
MITx/3.091x/2012_Fall      0                13583
                           1                  632
MITx/3.091x/2013_Spring    0                 6001
                           1                  138
MITx/6.002x/2012_Fall      0                39061
                           1                 1750
MITx/6.002x/2013_Spring    0                21642
                           1                  593
MITx/6.00x/2012_Fall       0                64254
                           1                 2477
MITx/6.00x/2013_Spring     0                56462
                           1                 1253
MITx/7.00x/2013_Spring     0                20186
                           1                  823
MITx/8.02x/2013_Spring     0                30226
                           1                  822
MITx/8.MReV/2013_Summer    0                 9180
                           1                  297
====================================================================
                     course_id  viewed final_cc_cname_DI  LoE_DI     YoB
\
                         count   count             count   count   count
registered certified
1          0            623451  623451            623451  623451  623451
           1             17687   17687             17687   17687   17687

                      gender   roles start_year start_month last_e_year  \
                       count   count      count       count       count
registered certified
1          0          623451  623451     623451      623451      623451
           1           17687   17687      17687       17687       17687

                     last_e_month
                            count
registered certified
1          0               623451
           1                17687
====================================================================
                 course_id registered final_cc_cname_DI  LoE_DI     YoB  \
                     count      count             count   count   count
viewed certified
0      0            240876     240876            240876  240876  240876
1      0            382575     382575            382575  382575  382575
       1             17687      17687             17687   17687   17687

                  gender   roles start_year start_month last_e_year  \
                   count   count      count       count       count
viewed certified
0      0          240876  240876     240876      240876      240876
1      0          382575  382575     382575      382575      382575
       1           17687   17687      17687       17687       17687

                 last_e_month
                        count
viewed certified
0      0               240876
```

```
1      0               382575
       1                17687
=====================================================================
                    course_id registered  viewed final_cc_cname_DI  LoE_DI  \
                        count      count   count             count   count
certified certified
0         0            623451     623451  623451            623451  623451
1         1             17687      17687   17687             17687   17687

                        YoB  gender   roles start_year start_month  \
                      count   count   count      count       count
certified certified
0         0          623451  623451  623451     623451      623451
1         1           17687   17687   17687      17687       17687

                    last_e_year last_e_month
                          count        count
certified certified
0         0              623451       623451
1         1               17687        17687
=====================================================================
                                                         course_id registered  vie
wed  \
                                                             count      count   co
unt
final_cc_cname_DI                       certified
Australia                               0                     6223       6223    6
223
                                        1                      196        196
196
Bangladesh                              0                     3148       3148    3
148
                                        1                       34         34
34
Brazil                                  0                    17403      17403   17
403
                                        1                      453        453
453
Canada                                  0                    12405      12405   12
405
                                        1                      333        333
333
China                                   0                     5108       5108    5
108
                                        1                       62         62
62
Colombia                                0                     4601       4601    4
601
                                        1                      202        202
202
Egypt                                   0                     9166       9166    9
166
                                        1                      120        120
120
France                                  0                     4496       4496    4
496
                                        1                      204        204
204
Germany                                 0                     7612       7612    7
612
```

```
                                       1           462        462
462
Greece                                 0          4845       4845    4
845
                                       1           317        317
317
India                                  0         85491      85491   85
491
                                       1          3205       3205    3
205
Indonesia                              0          3314       3314    3
314
                                       1            96         96
96
Japan                                  0          2230       2230    2
230
                                       1            40         40
40
Mexico                                 0          5470       5470    5
470
                                       1           168        168
168
Morocco                                0          3938       3938    3
938
                                       1            28         28
28
...                                                ...        ...
...
Other Middle East/Central Asia         0         17005      17005   17
005
                                       1           320        320
320
Other North & Central Amer., Caribbean 0          4265       4265    4
265
                                       1           169        169
169
Other Oceania                          0           341        341
341
                                       1             5          5
5
Other South America                    0          9624       9624    9
624
                                       1           292        292
292
Other South Asia                       0         12584      12584   12
584
                                       1           408        408
408
Pakistan                               0         10667      10667   10
667
                                       1           157        157
157
Philippines                            0          5293       5293    5
293
                                       1            81         81
81
Poland                                 0          4813       4813    4
813
                                       1           413        413
413
Portugal                               0          2081       2081    2
```

```
081
                                       1              112        112
112
Russian Federation                     0             9797       9797    9
797
                                       1              635        635
635
Spain                                  0             9166       9166    9
166
                                       1              837        837
837
Ukraine                                0             3897       3897    3
897
                                       1              203        203
203
United Kingdom                         0            21261      21261   21
261
                                       1              870        870
870
United States                          0           179858     179858  179
858
                                       1             4382       4382    4
382
Unknown/Other                          0            81968      81968   81
968
                                       1               61         61
61

                                                LoE_DI     YoB  gender
\
                                                 count   count   count
final_cc_cname_DI                      certified
Australia                              0          6223    6223    6223
                                       1           196     196     196
Bangladesh                             0          3148    3148    3148
                                       1            34      34      34
Brazil                                 0         17403   17403   17403
                                       1           453     453     453
Canada                                 0         12405   12405   12405
                                       1           333     333     333
China                                  0          5108    5108    5108
                                       1            62      62      62
Colombia                               0          4601    4601    4601
                                       1           202     202     202
Egypt                                  0          9166    9166    9166
                                       1           120     120     120
France                                 0          4496    4496    4496
                                       1           204     204     204
Germany                                0          7612    7612    7612
                                       1           462     462     462
Greece                                 0          4845    4845    4845
                                       1           317     317     317
India                                  0         85491   85491   85491
                                       1          3205    3205    3205
Indonesia                              0          3314    3314    3314
                                       1            96      96      96
Japan                                  0          2230    2230    2230
                                       1            40      40      40
Mexico                                 0          5470    5470    5470
                                       1           168     168     168
Morocco                                0          3938    3938    3938
```

```
                                          1               28       28       28
...                                                      ...      ...      ...
Other Middle East/Central Asia            0            17005    17005    17005
                                          1              320      320      320
Other North & Central Amer., Caribbean    0             4265     4265     4265
                                          1              169      169      169
Other Oceania                             0              341      341      341
                                          1                5        5        5
Other South America                       0             9624     9624     9624
                                          1              292      292      292
Other South Asia                          0            12584    12584    12584
                                          1              408      408      408
Pakistan                                  0            10667    10667    10667
                                          1              157      157      157
Philippines                               0             5293     5293     5293
                                          1               81       81       81
Poland                                    0             4813     4813     4813
                                          1              413      413      413
Portugal                                  0             2081     2081     2081
                                          1              112      112      112
Russian Federation                        0             9797     9797     9797
                                          1              635      635      635
Spain                                     0             9166     9166     9166
                                          1              837      837      837
Ukraine                                   0             3897     3897     3897
                                          1              203      203      203
United Kingdom                            0            21261    21261    21261
                                          1              870      870      870
United States                             0           179858   179858   179858
                                          1             4382     4382     4382
Unknown/Other                             0            81968    81968    81968
                                          1               61       61       61

                                                       roles start_year  \
                                                       count      count
final_cc_cname_DI                         certified
Australia                                 0             6223       6223
                                          1              196        196
Bangladesh                                0             3148       3148
                                          1               34         34
Brazil                                    0            17403      17403
                                          1              453        453
Canada                                    0            12405      12405
                                          1              333        333
China                                     0             5108       5108
                                          1               62         62
Colombia                                  0             4601       4601
                                          1              202        202
Egypt                                     0             9166       9166
                                          1              120        120
France                                    0             4496       4496
                                          1              204        204
Germany                                   0             7612       7612
                                          1              462        462
Greece                                    0             4845       4845
                                          1              317        317
India                                     0            85491      85491
                                          1             3205       3205
Indonesia                                 0             3314       3314
                                          1               96         96
Japan                                     0             2230       2230
```

```
                                         1              40        40
Mexico                                   0            5470      5470
                                         1             168       168
Morocco                                  0            3938      3938
                                         1              28        28
...                                                    ...       ...
Other Middle East/Central Asia           0           17005     17005
                                         1             320       320
Other North & Central Amer., Caribbean 0             4265      4265
                                         1             169       169
Other Oceania                            0             341       341
                                         1               5         5
Other South America                      0            9624      9624
                                         1             292       292
Other South Asia                         0           12584     12584
                                         1             408       408
Pakistan                                 0           10667     10667
                                         1             157       157
Philippines                              0            5293      5293
                                         1              81        81
Poland                                   0            4813      4813
                                         1             413       413
Portugal                                 0            2081      2081
                                         1             112       112
Russian Federation                       0            9797      9797
                                         1             635       635
Spain                                    0            9166      9166
                                         1             837       837
Ukraine                                  0            3897      3897
                                         1             203       203
United Kingdom                           0           21261     21261
                                         1             870       870
United States                            0          179858    179858
                                         1            4382      4382
Unknown/Other                            0           81968     81968
                                         1              61        61

                                                  start_month last_e_year
\
                                                        count       count
final_cc_cname_DI                        certified
Australia                                0               6223        6223
                                         1                196         196
Bangladesh                               0               3148        3148
                                         1                 34          34
Brazil                                   0              17403       17403
                                         1                453         453
Canada                                   0              12405       12405
                                         1                333         333
China                                    0               5108        5108
                                         1                 62          62
Colombia                                 0               4601        4601
                                         1                202         202
Egypt                                    0               9166        9166
                                         1                120         120
France                                   0               4496        4496
                                         1                204         204
Germany                                  0               7612        7612
                                         1                462         462
Greece                                   0               4845        4845
                                         1                317         317
```

```
India                                  0            85491       85491
                                       1             3205        3205
Indonesia                              0             3314        3314
                                       1               96          96
Japan                                  0             2230        2230
                                       1               40          40
Mexico                                 0             5470        5470
                                       1              168         168
Morocco                                0             3938        3938
                                       1               28          28
...                                                   ...         ...
Other Middle East/Central Asia         0            17005       17005
                                       1              320         320
Other North & Central Amer., Caribbean 0             4265        4265
                                       1              169         169
Other Oceania                          0              341         341
                                       1                5           5
Other South America                    0             9624        9624
                                       1              292         292
Other South Asia                       0            12584       12584
                                       1              408         408
Pakistan                               0            10667       10667
                                       1              157         157
Philippines                            0             5293        5293
                                       1               81          81
Poland                                 0             4813        4813
                                       1              413         413
Portugal                               0             2081        2081
                                       1              112         112
Russian Federation                     0             9797        9797
                                       1              635         635
Spain                                  0             9166        9166
                                       1              837         837
Ukraine                                0             3897        3897
                                       1              203         203
United Kingdom                         0            21261       21261
                                       1              870         870
United States                          0           179858      179858
                                       1             4382        4382
Unknown/Other                          0            81968       81968
                                       1               61          61

                                                last_e_month
                                                       count
final_cc_cname_DI                      certified
Australia                              0                6223
                                       1                 196
Bangladesh                             0                3148
                                       1                  34
Brazil                                 0               17403
                                       1                 453
Canada                                 0               12405
                                       1                 333
China                                  0                5108
                                       1                  62
Colombia                               0                4601
                                       1                 202
Egypt                                  0                9166
                                       1                 120
France                                 0                4496
                                       1                 204
```

```
Germany                               0               7612
                                      1                462
Greece                                0               4845
                                      1                317
India                                 0              85491
                                      1               3205
Indonesia                             0               3314
                                      1                 96
Japan                                 0               2230
                                      1                 40
Mexico                                0               5470
                                      1                168
Morocco                               0               3938
                                      1                 28
...                                                    ...
Other Middle East/Central Asia        0              17005
                                      1                320
Other North & Central Amer., Caribbean 0              4265
                                      1                169
Other Oceania                         0                341
                                      1                  5
Other South America                   0               9624
                                      1                292
Other South Asia                      0              12584
                                      1                408
Pakistan                              0              10667
                                      1                157
Philippines                           0               5293
                                      1                 81
Poland                                0               4813
                                      1                413
Portugal                              0               2081
                                      1                112
Russian Federation                    0               9797
                                      1                635
Spain                                 0               9166
                                      1                837
Ukraine                               0               3897
                                      1                203
United Kingdom                        0              21261
                                      1                870
United States                         0             179858
                                      1               4382
Unknown/Other                         0              81968
                                      1                 61

[68 rows x 11 columns]
======================================================================
                              course_id registered  viewed final_cc_cname_
DI  \
                                  count      count   count             cou
nt
LoE_DI              certified
0                   0            102092     102092  102092            1020
92
                    1              3916       3916    3916              39
16
Bachelor's          0            214898     214898  214898            2148
98
                    1              4870       4870    4870              48
70
```

```
Doctorate           0              12965      12965   12965             129
65
                    1                422        422     422               4
22
Less than Secondary 0              13690      13690   13690             136
90
                    1                402        402     402               4
02
Master's            0             113971     113971  113971            1139
71
                    1               4218       4218    4218              42
18
Secondary           0             165835     165835  165835            1658
35
                    1               3859       3859    3859              38
59

                                 YoB  gender   roles start_year start_mon
th  \
                               count   count   count      count       cou
nt
LoE_DI              certified
0                   0         102092  102092  102092     102092      1020
92
                    1           3916    3916    3916       3916        39
16
Bachelor's          0         214898  214898  214898     214898      2148
98
                    1           4870    4870    4870       4870        48
70
Doctorate           0          12965   12965   12965      12965       129
65
                    1            422     422     422        422         4
22
Less than Secondary 0          13690   13690   13690      13690       136
90
                    1            402     402     402        402         4
02
Master's            0         113971  113971  113971     113971      1139
71
                    1           4218    4218    4218       4218        42
18
Secondary           0         165835  165835  165835     165835      1658
35
                    1           3859    3859    3859       3859        38
59

                              last_e_year last_e_month
                                    count        count
LoE_DI              certified
0                   0              102092       102092
                    1                3916         3916
Bachelor's          0              214898       214898
                    1                4870         4870
Doctorate           0               12965        12965
                    1                 422          422
Less than Secondary 0               13690        13690
                    1                 402          402
Master's            0              113971       113971
                    1                4218         4218
Secondary           0              165835       165835
```

```
                   1                3859         3859
======================================================================
                 course_id registered viewed final_cc_cname_DI LoE_DI gend
er  \
                     count      count  count             count  count   cou
nt
YOB    certified
0.0    0             92737      92737  92737             92737  92737   927
37
       1              3868       3868   3868              3868   3868    38
68
1931.0 0                 7          7      7                 7      7
7
1934.0 0                 5          5      5                 5      5
5
1935.0 0                36         36     36                36     36
36
1936.0 0                42         42     42                42     42
42
       1                 1          1      1                 1      1
1
1937.0 0                60         60     60                60     60
60
       1                 4          4      4                 4      4
4
1938.0 0                72         72     72                72     72
72
       1                 2          2      2                 2      2
2
1939.0 0                80         80     80                80     80
80
       1                 6          6      6                 6      6
6
1940.0 0                91         91     91                91     91
91
       1                 1          1      1                 1      1
1
1941.0 0                94         94     94                94     94
94
       1                 2          2      2                 2      2
2
1942.0 0               193        193    193               193    193     1
93
       1                 3          3      3                 3      3
3
1943.0 0               218        218    218               218    218     2
18
       1                 5          5      5                 5      5
5
1944.0 0               218        218    218               218    218     2
18
       1                 3          3      3                 3      3
3
1945.0 0               211        211    211               211    211     2
11
       1                 5          5      5                 5      5
5
1946.0 0               320        320    320               320    320     3
20
       1                11         11     11                11     11
11
```

```
1947.0 0             451         451    451                  451    451    4
51
       1               9           9      9                    9      9
9
1948.0 0             355         355    355                  355    355    3
55
...                  ...         ...    ...                  ...    ...
...
1992.0 0           36789       36789  36789                36789  36789  367
89
       1            1057        1057   1057                 1057   1057   10
57
1993.0 0           32985       32985  32985                32985  32985  329
85
       1            1014        1014   1014                 1014   1014   10
14
1994.0 0           23411       23411  23411                23411  23411  234
11
       1             510         510    510                  510    510    5
10
1995.0 0           13546       13546  13546                13546  13546  135
46
       1             260         260    260                  260    260    2
60
1996.0 0            7518        7518   7518                 7518   7518   75
18
       1             236         236    236                  236    236    2
36
1997.0 0            3979        3979   3979                 3979   3979   39
79
       1             103         103    103                  103    103    1
03
1998.0 0            1895        1895   1895                 1895   1895   18
95
       1              40          40     40                   40     40
40
1999.0 0             934         934    934                  934    934    9
34
       1              18          18     18                   18     18
18
2000.0 0             334         334    334                  334    334    3
34
       1               6           6      6                    6      6
6
2001.0 0             140         140    140                  140    140    1
40
       1               2           2      2                    2      2
2
2002.0 0              44          44     44                   44     44
44
       1               1           1      1                    1      1
1
2003.0 0              10          10     10                   10     10
10
2007.0 0               6           6      6                    6      6
6
2008.0 0              10          10     10                   10     10
10
2009.0 0               8           8      8                    8      8
8
2010.0 0              17          17     17                   17     17
```

```
17
2011.0 0               34          34      34                34      34
34
2012.0 0              472         472     472               472     472    4
72
2013.0 0               61          61      61                61      61
61

                 roles start_year start_month last_e_year last_e_month
                 count      count       count       count        count
YoB    certified
0.0    0         92737      92737       92737       92737        92737
       1          3868       3868        3868        3868         3868
1931.0 0             7          7           7           7            7
1934.0 0             5          5           5           5            5
1935.0 0            36         36          36          36           36
1936.0 0            42         42          42          42           42
       1             1          1           1           1            1
1937.0 0            60         60          60          60           60
       1             4          4           4           4            4
1938.0 0            72         72          72          72           72
       1             2          2           2           2            2
1939.0 0            80         80          80          80           80
       1             6          6           6           6            6
1940.0 0            91         91          91          91           91
       1             1          1           1           1            1
1941.0 0            94         94          94          94           94
       1             2          2           2           2            2
1942.0 0           193        193         193         193          193
       1             3          3           3           3            3
1943.0 0           218        218         218         218          218
       1             5          5           5           5            5
1944.0 0           218        218         218         218          218
       1             3          3           3           3            3
1945.0 0           211        211         211         211          211
       1             5          5           5           5            5
1946.0 0           320        320         320         320          320
       1            11         11          11          11           11
1947.0 0           451        451         451         451          451
       1             9          9           9           9            9
1948.0 0           355        355         355         355          355
...                ...        ...         ...         ...          ...
1992.0 0         36789      36789       36789       36789        36789
       1          1057       1057        1057        1057         1057
1993.0 0         32985      32985       32985       32985        32985
       1          1014       1014        1014        1014         1014
1994.0 0         23411      23411       23411       23411        23411
       1           510        510         510         510          510
1995.0 0         13546      13546       13546       13546        13546
       1           260        260         260         260          260
1996.0 0          7518       7518        7518        7518         7518
       1           236        236         236         236          236
1997.0 0          3979       3979        3979        3979         3979
       1           103        103         103         103          103
1998.0 0          1895       1895        1895        1895         1895
       1            40         40          40          40           40
1999.0 0           934        934         934         934          934
       1            18         18          18          18           18
2000.0 0           334        334         334         334          334
       1             6          6           6           6            6
2001.0 0           140        140         140         140          140
```

```
       1              2          2           2           2            2
2002.0 0             44         44          44          44           44
       1              1          1           1           1            1
2003.0 0             10         10          10          10           10
2007.0 0              6          6           6           6            6
2008.0 0             10         10          10          10           10
2009.0 0              8          8           8           8            8
2010.0 0             17         17          17          17           17
2011.0 0             34         34          34          34           34
2012.0 0            472        472         472         472          472
2013.0 0             61         61          61          61           61

[147 rows x 11 columns]
======================================================================
                 course_id registered  viewed final_cc_cname_DI  LoE_DI  \
                     count      count   count             count   count
gender certified
0      0             83199      83199   83199             83199   83199
       1              3607       3607    3607              3607    3607
f      0            138563     138563  138563            138563  138563
       1              4232       4232    4232              4232    4232
m      0            401672     401672  401672            401672  401672
       1              9848       9848    9848              9848    9848
o      0                17         17      17                17      17

                       YoB   roles start_year start_month last_e_year  \
                     count   count      count       count       count
gender certified
0      0             83199   83199      83199       83199       83199
       1              3607    3607       3607        3607        3607
f      0            138563  138563     138563      138563      138563
       1              4232    4232       4232        4232        4232
m      0            401672  401672     401672      401672      401672
       1              9848    9848       9848        9848        9848
o      0                17      17         17          17          17

                 last_e_month
                        count
gender certified
0      0                83199
       1                 3607
f      0               138563
       1                 4232
m      0               401672
       1                 9848
o      0                   17
======================================================================
                course_id registered  viewed final_cc_cname_DI  LoE_DI  \
                    count      count   count             count   count
roles certified
0.0   0            623451     623451  623451            623451  623451
      1             17687      17687   17687             17687   17687

                      YoB  gender start_year start_month last_e_year  \
                    count   count      count       count       count
roles certified
0.0   0            623451  623451     623451      623451      623451
      1             17687   17687      17687       17687       17687

                last_e_month
                       count
```

```
roles certified
0.0   0             623451
      1              17687
=====================================================================
                     course_id registered  viewed final_cc_cname_DI  LoE_D
I  \
                         count      count   count             count   coun
t
start_year certified
2012       0            292195     292195  292195            292195  29219
5
           1              8627       8627    8627              8627    862
7
2013       0            331256     331256  331256            331256  33125
6
           1              9060       9060    9060              9060    906
0

                         YoB  gender   roles start_month last_e_year  \
                       count   count   count       count       count
start_year certified
2012       0          292195  292195  292195      292195      292195
           1            8627    8627    8627        8627        8627
2013       0          331256  331256  331256      331256      331256
           1            9060    9060    9060        9060        9060

                     last_e_month
                            count
start_year certified
2012       0               292195
           1                 8627
2013       0               331256
           1                 9060
=====================================================================
                      course_id registered viewed final_cc_cname_DI LoE_DI
\
                          count      count  count             count  count
start_month certified
1           0             75388      75388  75388             75388  75388
            1              2678       2678   2678              2678   2678
2           0             74828      74828  74828             74828  74828
            1              2995       2995   2995              2995   2995
3           0             58007      58007  58007             58007  58007
            1              1528       1528   1528              1528   1528
4           0             30279      30279  30279             30279  30279
            1               457        457    457               457    457
5           0             32080      32080  32080             32080  32080
            1               651        651    651               651    651
6           0             20598      20598  20598             20598  20598
            1               416        416    416               416    416
7           0             35454      35454  35454             35454  35454
            1               969        969    969               969    969
8           0             93369      93369  93369             93369  93369
            1              2148       2148   2148              2148   2148
9           0             64145      64145  64145             64145  64145
            1              2412       2412   2412              2412   2412
10          0             70981      70981  70981             70981  70981
            1              2276       2276   2276              2276   2276
11          0             25097      25097  25097             25097  25097
            1               264        264    264               264    264
12          0             43225      43225  43225             43225  43225
```

```
            1               893        893    893               893    893

                               YoB gender  roles start_year last_e_year last_e_m
onth
                             count  count  count      count       count        c
ount
start_month certified
1           0            75388  75388  75388      75388       75388        7
5388
            1             2678   2678   2678       2678        2678
 2678
2           0            74828  74828  74828      74828       74828        7
4828
            1             2995   2995   2995       2995        2995
 2995
3           0            58007  58007  58007      58007       58007        5
8007
            1             1528   1528   1528       1528        1528
 1528
4           0            30279  30279  30279      30279       30279        3
0279
            1              457    457    457        457         457
457
5           0            32080  32080  32080      32080       32080        3
2080
            1              651    651    651        651         651
651
6           0            20598  20598  20598      20598       20598        2
0598
            1              416    416    416        416         416
416
7           0            35454  35454  35454      35454       35454        3
5454
            1              969    969    969        969         969
969
8           0            93369  93369  93369      93369       93369        9
3369
            1             2148   2148   2148       2148        2148
 2148
9           0            64145  64145  64145      64145       64145        6
4145
            1             2412   2412   2412       2412        2412
 2412
10          0            70981  70981  70981      70981       70981        7
0981
            1             2276   2276   2276       2276        2276
 2276
11          0            25097  25097  25097      25097       25097        2
5097
            1              264    264    264        264         264
264
12          0            43225  43225  43225      43225       43225        4
3225
            1              893    893    893        893         893
893
======================================================================
                      course_id registered  viewed final_cc_cname_DI  LoE_
DI  \
                          count      count   count             count   cou
nt
last_e_year certified
```

```
0.0          0             178942      178942  178942            178942  1789
42
             1                 12          12      12                12
 12
2012.0       0             117866      117866  117866            117866  1178
66
             1                347         347     347               347     3
47
2013.0       0             326643      326643  326643            326643  3266
43
             1              17328       17328   17328             17328   173
28

                             YoB  gender   roles start_year start_month  \
                           count   count   count      count       count
last_e_year certified
0.0          0           178942  178942  178942     178942      178942
             1               12      12      12         12          12
2012.0       0           117866  117866  117866     117866      117866
             1              347     347     347        347         347
2013.0       0           326643  326643  326643     326643      326643
             1            17328   17328   17328      17328       17328

                       last_e_month
                              count
last_e_year certified
0.0          0               178942
             1                   12
2012.0       0               117866
             1                  347
2013.0       0               326643
             1                17328
======================================================================
                          course_id registered  viewed final_cc_cname_DI  LoE
_DI  \
                              count      count   count             count   co
unt
last_e_month certified
0.0           0              178942     178942  178942            178942  178
942
              1                  12         12      12                12
12
1.0           0               24499      24499   24499             24499   24
499
              1                2639       2639    2639              2639    2
639
2.0           0               44229      44229   44229             44229   44
229
              1                1391       1391    1391              1391    1
391
3.0           0               71533      71533   71533             71533   71
533
              1                 589        589     589               589
 589
4.0           0               41145      41145   41145             41145   41
145
              1                 710        710     710               710
 710
5.0           0               42267      42267   42267             42267   42
267
              1                2353       2353    2353              2353    2
```

```
353
6.0          0             33339      33339   33339             33339   33
339
             1              2724       2724    2724              2724    2
724
7.0          0             31254      31254   31254             31254   31
254
             1              2629       2629    2629              2629    2
629
8.0          0             42986      42986   42986             42986   42
986
             1              3528       3528    3528              3528    3
528
9.0          0             27737      27737   27737             27737   27
737
             1               765        765     765               765
 765
10.0         0             46510      46510   46510             46510   46
510
11.0         0             24223      24223   24223             24223   24
223
12.0         0             14787      14787   14787             14787   14
787
             1               347        347     347               347
 347

                           YoB  gender   roles start_year start_month  \
                         count   count   count      count       count
last_e_month certified
0.0          0          178942  178942  178942     178942      178942
             1              12      12      12         12          12
1.0          0           24499   24499   24499      24499       24499
             1            2639    2639    2639       2639        2639
2.0          0           44229   44229   44229      44229       44229
             1            1391    1391    1391       1391        1391
3.0          0           71533   71533   71533      71533       71533
             1             589     589     589        589         589
4.0          0           41145   41145   41145      41145       41145
             1             710     710     710        710         710
5.0          0           42267   42267   42267      42267       42267
             1            2353    2353    2353       2353        2353
6.0          0           33339   33339   33339      33339       33339
             1            2724    2724    2724       2724        2724
7.0          0           31254   31254   31254      31254       31254
             1            2629    2629    2629       2629        2629
8.0          0           42986   42986   42986      42986       42986
             1            3528    3528    3528       3528        3528
9.0          0           27737   27737   27737      27737       27737
             1             765     765     765        765         765
10.0         0           46510   46510   46510      46510       46510
11.0         0           24223   24223   24223      24223       24223
12.0         0           14787   14787   14787      14787       14787
             1             347     347     347        347         347

                       last_e_year
                             count
last_e_month certified
0.0          0              178942
             1                  12
1.0          0               24499
             1                2639
```

```
2.0         0          44229
            1           1391
3.0         0          71533
            1            589
4.0         0          41145
            1            710
5.0         0          42267
            1           2353
6.0         0          33339
            1           2724
7.0         0          31254
            1           2629
8.0         0          42986
            1           3528
9.0         0          27737
            1            765
10.0        0          46510
11.0        0          24223
12.0        0          14787
            1            347
=====================================================================
```

In [24]:

```
df.describe()
```

Out[24]:

	registered	viewed	certified	YoB	roles	start_year
count	641138.0	641138.000000	641138.000000	641138.000000	641138.0	641138.000000
mean	1.0	0.624299	0.027587	1686.120498	0.0	2012.530800
std	0.0	0.484304	0.163786	710.240700	0.0	0.499051
min	1.0	0.000000	0.000000	0.000000	0.0	2012.000000
25%	1.0	0.000000	0.000000	1975.000000	0.0	2012.000000
50%	1.0	1.000000	0.000000	1986.000000	0.0	2013.000000
75%	1.0	1.000000	0.000000	1991.000000	0.0	2013.000000
max	1.0	1.000000	1.000000	2013.000000	0.0	2013.000000

In [25]:

```
data_dum = pd.get_dummies(df, drop_first=False)
#data_dum.head(10).to_csv('novoModelo.csv')

data_dum.shape
```

Out[25]:

```
(641138, 69)
```

In [26]:

```
modelo, dfTrat = PrepararLista(data_dum)
```

```
registered
viewed
certified
YoB
roles
start_year
start_month
last_e_year
last_e_month
course_id_HarvardX/CB22x/2013_Spring
course_id_HarvardX/CS50x/2012
course_id_HarvardX/ER22x/2013_Spring
course_id_HarvardX/PH207x/2012_Fall
course_id_HarvardX/PH278x/2013_Spring
course_id_MITx/14.73x/2013_Spring
course_id_MITx/2.01x/2013_Spring
course_id_MITx/3.091x/2012_Fall
course_id_MITx/3.091x/2013_Spring
course_id_MITx/6.002x/2012_Fall
course_id_MITx/6.002x/2013_Spring
course_id_MITx/6.00x/2012_Fall
course_id_MITx/6.00x/2013_Spring
course_id_MITx/7.00x/2013_Spring
course_id_MITx/8.02x/2013_Spring
course_id_MITx/8.MReV/2013_Summer
final_cc_cname_DI_Australia
final_cc_cname_DI_Bangladesh
final_cc_cname_DI_Brazil
final_cc_cname_DI_Canada
final_cc_cname_DI_China
final_cc_cname_DI_Colombia
final_cc_cname_DI_Egypt
final_cc_cname_DI_France
final_cc_cname_DI_Germany
final_cc_cname_DI_Greece
final_cc_cname_DI_India
final_cc_cname_DI_Indonesia
final_cc_cname_DI_Japan
final_cc_cname_DI_Mexico
final_cc_cname_DI_Morocco
final_cc_cname_DI_Nigeria
final_cc_cname_DI_Other Africa
final_cc_cname_DI_Other East Asia
final_cc_cname_DI_Other Europe
final_cc_cname_DI_Other Middle East/Central Asia
final_cc_cname_DI_Other North & Central Amer., Caribbean
final_cc_cname_DI_Other Oceania
final_cc_cname_DI_Other South America
final_cc_cname_DI_Other South Asia
final_cc_cname_DI_Pakistan
final_cc_cname_DI_Philippines
final_cc_cname_DI_Poland
final_cc_cname_DI_Portugal
final_cc_cname_DI_Russian Federation
final_cc_cname_DI_Spain
final_cc_cname_DI_Ukraine
final_cc_cname_DI_United Kingdom
final_cc_cname_DI_United States
final_cc_cname_DI_Unknown/Other
LoE_DI_0
LoE_DI_Bachelor's
```

```
LoE_DI_Doctorate
LoE_DI_Less than Secondary
LoE_DI_Master's
LoE_DI_Secondary
gender_0
gender_f
gender_m
gender_o
```

In [27]:

```
dfTrat.describe()
```

Out[27]:

	registered	viewed	certified	YoB	roles	start_year
count	641138.0	641138.000000	641138.000000	641138.000000	641138.0	641138.000000
mean	1.0	0.624299	0.027587	1686.120498	0.0	2012.530800
std	0.0	0.484304	0.163786	710.240700	0.0	0.499051
min	1.0	0.000000	0.000000	0.000000	0.0	2012.000000
25%	1.0	0.000000	0.000000	1975.000000	0.0	2012.000000
50%	1.0	1.000000	0.000000	1986.000000	0.0	2013.000000
75%	1.0	1.000000	0.000000	1991.000000	0.0	2013.000000
max	1.0	1.000000	1.000000	2013.000000	0.0	2013.000000

8 rows × 69 columns

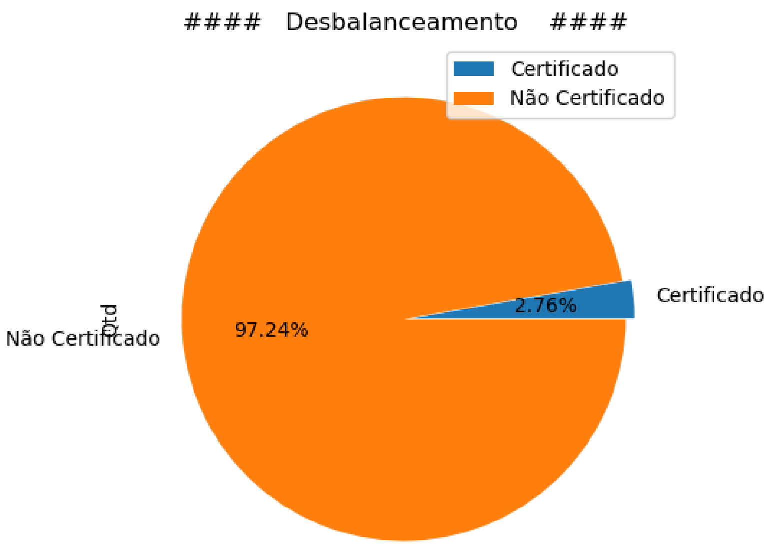
In [28]:

```
val_classes = dfTrat[classe].values
del dfTrat[classe]
```

In [29]:

```
dfDesbOrig = ExibirDesbanciamento(val_classes)
dfDesbOrig.plot.pie(x='Atr', y='Qtd', title='####   Desbalanceamento    ####', labels=d
fDesbOrig['Atr'].values, explode=(0.03, 0.01),autopct = '%1.2f%%', figsize=(8, 5))
dfDesbOrig
```

0.03



Out[29]:

	Atr	Qtd
0	Certificado	17687
1	Não Certificado	623451

In [30]:

```
X_train, X_test, y_train, y_test = train_test_split(dfTrat, val_classes, test_size=0.3,
random_state=42)
X_test_orig = X_test
y_test_orig = y_test
```

In [31]:

```
#======================================================================
#
#                           Obter pesos
#
#======================================================================

clf = ObterAlgoritmoClf(i_clf_tree)
clf.fit(X_train, y_train)

colunas = dfTrat.columns
for i in range(0, len(colunas)):
    print(colunas[i] + '  [ ' + str(round(clf.feature_importances_[i], 3)) + ' ]')
```

```
registered  [ 0.0 ]
viewed  [ 0.093 ]
YoB  [ 0.0 ]
roles  [ 0.0 ]
start_year  [ 0.163 ]
start_month  [ 0.228 ]
last_e_year  [ 0.108 ]
last_e_month  [ 0.217 ]
course_id_HarvardX/CB22x/2013_Spring  [ 0.0 ]
course_id_HarvardX/CS50x/2012  [ 0.124 ]
course_id_HarvardX/ER22x/2013_Spring  [ 0.012 ]
course_id_HarvardX/PH207x/2012_Fall  [ 0.0 ]
course_id_HarvardX/PH278x/2013_Spring  [ 0.0 ]
course_id_MITx/14.73x/2013_Spring  [ 0.04 ]
course_id_MITx/2.01x/2013_Spring  [ 0.0 ]
course_id_MITx/3.091x/2012_Fall  [ 0.0 ]
course_id_MITx/3.091x/2013_Spring  [ 0.0 ]
course_id_MITx/6.002x/2012_Fall  [ 0.012 ]
course_id_MITx/6.002x/2013_Spring  [ 0.0 ]
course_id_MITx/6.00x/2012_Fall  [ 0.0 ]
course_id_MITx/6.00x/2013_Spring  [ 0.0 ]
course_id_MITx/7.00x/2013_Spring  [ 0.0 ]
course_id_MITx/8.02x/2013_Spring  [ 0.0 ]
course_id_MITx/8.MReV/2013_Summer  [ 0.0 ]
final_cc_cname_DI_Australia  [ 0.0 ]
final_cc_cname_DI_Bangladesh  [ 0.0 ]
final_cc_cname_DI_Brazil  [ 0.0 ]
final_cc_cname_DI_Canada  [ 0.0 ]
final_cc_cname_DI_China  [ 0.0 ]
final_cc_cname_DI_Colombia  [ 0.0 ]
final_cc_cname_DI_Egypt  [ 0.0 ]
final_cc_cname_DI_France  [ 0.0 ]
final_cc_cname_DI_Germany  [ 0.0 ]
final_cc_cname_DI_Greece  [ 0.0 ]
final_cc_cname_DI_India  [ 0.003 ]
final_cc_cname_DI_Indonesia  [ 0.0 ]
final_cc_cname_DI_Japan  [ 0.0 ]
final_cc_cname_DI_Mexico  [ 0.0 ]
final_cc_cname_DI_Morocco  [ 0.0 ]
final_cc_cname_DI_Nigeria  [ 0.0 ]
final_cc_cname_DI_Other Africa  [ 0.0 ]
final_cc_cname_DI_Other East Asia  [ 0.0 ]
final_cc_cname_DI_Other Europe  [ 0.0 ]
final_cc_cname_DI_Other Middle East/Central Asia  [ 0.0 ]
final_cc_cname_DI_Other North & Central Amer., Caribbean  [ 0.0 ]
final_cc_cname_DI_Other Oceania  [ 0.0 ]
final_cc_cname_DI_Other South America  [ 0.0 ]
final_cc_cname_DI_Other South Asia  [ 0.0 ]
final_cc_cname_DI_Pakistan  [ 0.0 ]
final_cc_cname_DI_Philippines  [ 0.0 ]
final_cc_cname_DI_Poland  [ 0.0 ]
final_cc_cname_DI_Portugal  [ 0.0 ]
final_cc_cname_DI_Russian Federation  [ 0.0 ]
final_cc_cname_DI_Spain  [ 0.0 ]
final_cc_cname_DI_Ukraine  [ 0.0 ]
final_cc_cname_DI_United Kingdom  [ 0.0 ]
final_cc_cname_DI_United States  [ 0.0 ]
final_cc_cname_DI_Unknown/Other  [ 0.0 ]
LoE_DI_0  [ 0.0 ]
LoE_DI_Bachelor's  [ 0.0 ]
LoE_DI_Doctorate  [ 0.0 ]
```

```
LoE_DI_Less than Secondary  [ 0.0 ]
LoE_DI_Master's  [ 0.0 ]
LoE_DI_Secondary  [ 0.0 ]
gender_0  [ 0.0 ]
gender_f  [ 0.0 ]
gender_m  [ 0.0 ]
gender_o  [ 0.0 ]
```

In [32]:

```
dfAlg = ExibirMedidas(X_train, X_test, y_train, y_test)
dfAlg[colsExibir]
```

```
2019-04-17 15:39:20.079977
{'id_alg': 1, 'algoritmo': 'Tree'}
DecisionTreeClassifier(class_weight=None, criterion='gini', max_depth=5,
            max_features=None, max_leaf_nodes=None,
            min_impurity_decrease=0.0, min_impurity_split=None,
            min_samples_leaf=1, min_samples_split=2,
            min_weight_fraction_leaf=0.0, presort=False, random_state=Non
e,
            splitter='best')
2019-04-17 15:39:22.346298
{'id_alg': 2, 'algoritmo': 'KNN'}
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
           metric_params=None, n_jobs=None, n_neighbors=3, p=2,
           weights='uniform')
2019-04-17 15:46:58.642884
{'id_alg': 3, 'algoritmo': 'SVM'}
SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
  decision_function_shape='ovr', degree=3, gamma='auto_deprecated',
  kernel='sigmoid', max_iter=-1, probability=False, random_state=None,
  shrinking=True, tol=0.001, verbose=False)
2019-04-17 16:14:45.229892
{'id_alg': 4, 'algoritmo': 'MLP'}
MLPClassifier(activation='relu', alpha=0.0001, batch_size='auto', beta_1=
0.9,
       beta_2=0.999, early_stopping=False, epsilon=1e-08,
       hidden_layer_sizes=100, learning_rate='constant',
       learning_rate_init=0.001, max_iter=200, momentum=0.9,
       n_iter_no_change=10, nesterovs_momentum=True, power_t=0.5,
       random_state=None, shuffle=True, solver='adam', tol=0.0001,
       validation_fraction=0.1, verbose=False, warm_start=False)
2019-04-17 16:24:43.393790
{'id_alg': 5, 'algoritmo': 'Naive'}
GaussianNB(priors=None, var_smoothing=1e-09)
2019-04-17 16:24:45.629701
{'id_alg': 6, 'algoritmo': 'Dummy'}
DummyClassifier(constant=None, random_state=None, strategy='prior')
2019-04-17 16:24:45.851322
```

Out[32]:

	id_alg	algoritmo	acur	sens	esp	efic	VPP	VPN	TExec
0	1	Tree	97.20	97.20	0.00	48.60	100.00	0.00	0.04
1	2	KNN	96.81	97.97	40.37	69.17	98.76	29.03	7.60
2	3	SVM	97.20	97.20	0.00	48.60	100.00	0.00	27.78
3	4	MLP	97.20	97.20	0.00	48.60	100.00	0.00	9.97
4	5	Naive	65.73	99.40	6.66	53.03	65.14	86.30	0.04
5	6	Dummy	97.20	97.20	0.00	48.60	100.00	0.00	0.00

In [33]:

```
X_original = dfTrat.values
y_original = val_classes
```

In [34]:

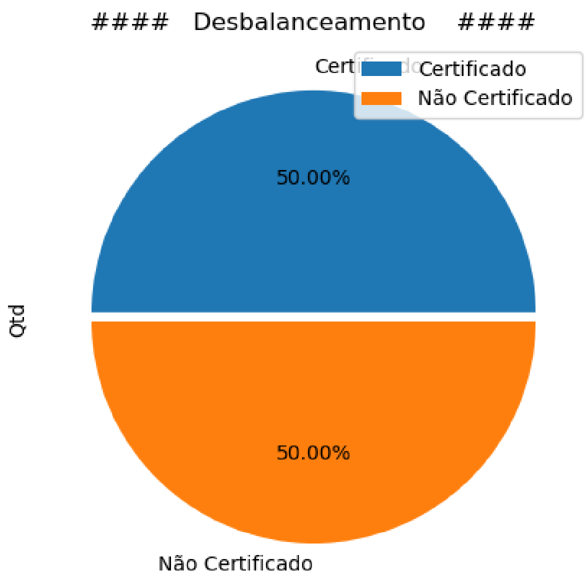
```
#over_sampling
#https://imbalanced-learn.readthedocs.io/en/stable/generated/imblearn.over_sampling.SMO
TE.html

print('over_sampling')
print(datetime.datetime.now())
X_over, y_over = GerarResampling('over', X_original, y_original)
X_train, X_test, y_train, y_test = train_test_split(X_over, y_over, test_size=0.3, rand
om_state=42)

print(datetime.datetime.now())
```

```
over_sampling
2019-04-17 17:07:07.183641
2019-04-17 17:07:17.443654
```

In [35]:

```
dfDesbOver = ExibirDesbanciamento(y_over)
dfDesbOver.plot.pie(x='Atr', y='Qtd', title='####   Desbalanceamento    ####', labels=d
fDesbOver['Atr'].values, explode=(0.03, 0.01),autopct = '%1.2f%%', figsize=(8, 5))
dfDesbOver
```

1.0



Out[35]:

	Atr	Qtd
0	Certificado	623451
1	Não Certificado	623451

In [36]:

```
dfAlgOver = ExibirMedidas(X_over, X_test, y_over, y_test)
dfAlgOver[colsExibir]
```

```
2019-04-17 17:07:52.324262
{'id_alg': 1, 'algoritmo': 'Tree'}
DecisionTreeClassifier(class_weight=None, criterion='gini', max_depth=5,
            max_features=None, max_leaf_nodes=None,
            min_impurity_decrease=0.0, min_impurity_split=None,
            min_samples_leaf=1, min_samples_split=2,
            min_weight_fraction_leaf=0.0, presort=False, random_state=Non
e,
            splitter='best')
2019-04-17 17:08:03.068270
{'id_alg': 2, 'algoritmo': 'KNN'}
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
           metric_params=None, n_jobs=None, n_neighbors=3, p=2,
           weights='uniform')
2019-04-17 17:55:55.605951
{'id_alg': 3, 'algoritmo': 'SVM'}
SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
  decision_function_shape='ovr', degree=3, gamma='auto_deprecated',
  kernel='sigmoid', max_iter=-1, probability=False, random_state=None,
  shrinking=True, tol=0.001, verbose=False)
2019-04-20 01:44:16.690193
{'id_alg': 4, 'algoritmo': 'MLP'}
MLPClassifier(activation='relu', alpha=0.0001, batch_size='auto', beta_1=
0.9,
       beta_2=0.999, early_stopping=False, epsilon=1e-08,
       hidden_layer_sizes=100, learning_rate='constant',
       learning_rate_init=0.001, max_iter=200, momentum=0.9,
       n_iter_no_change=10, nesterovs_momentum=True, power_t=0.5,
       random_state=None, shuffle=True, solver='adam', tol=0.0001,
       validation_fraction=0.1, verbose=False, warm_start=False)
2019-04-20 02:30:59.781103
{'id_alg': 5, 'algoritmo': 'Naive'}
GaussianNB(priors=None, var_smoothing=1e-09)
2019-04-20 02:31:04.307316
{'id_alg': 6, 'algoritmo': 'Dummy'}
DummyClassifier(constant=None, random_state=None, strategy='prior')
2019-04-20 02:31:04.779725
```

Out[36]:

	id_alg	algoritmo	acur	sens	esp	efic	VPP	VPN	TExec
0	1	Tree	89.51	89.47	89.54	89.51	89.57	89.44	0.18
1	2	KNN	96.78	96.56	97.00	96.78	97.02	96.54	47.88
2	3	SVM	49.96	0.00	49.96	24.98	0.00	100.00	3348.35
3	4	MLP	90.63	94.52	87.35	90.94	86.27	94.99	46.72
4	5	Naive	76.60	88.63	70.26	79.45	61.06	92.15	0.08
5	6	Dummy	50.04	50.04	0.00	25.02	100.00	0.00	0.01

In [37]:

```
#from imblearn.under_sampling import RandomUnderSampler
#randCnn = RandomUnderSampler(random_state=42)
```

In [38]:

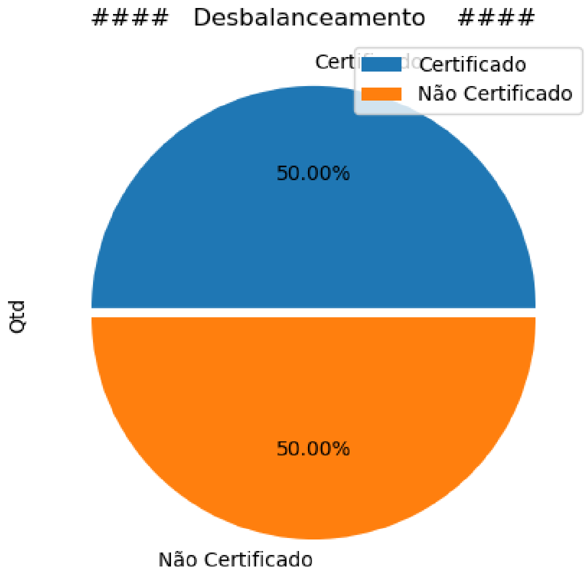
```
#under_sampling
#https://imbalanced-learn.readthedocs.io/en/stable/generated/imblearn.under_sampling.NeighbourhoodCleaningRule.html

print('under_sampling')
print(datetime.datetime.now())
X_under, y_under = GerarResampling('under', X_original, y_original) #original
#X_under, y_under = randCnn.fit_resample(X_original, y_original)
X_train, X_test, y_train, y_test = train_test_split(X_under, y_under, test_size=0.3, random_state=42)
print(datetime.datetime.now())
```

```
under_sampling
2019-04-20 10:13:37.189287
2019-04-20 10:13:38.198081
```

In [39]:

```
dfDesbUnder = ExibirDesbanciamento(y_under)
dfDesbUnder.plot.pie(x='Atr', y='Qtd', title='####   Desbalanceamento    ####', labels=
dfDesbUnder['Atr'].values, explode=(0.03, 0.01),autopct = '%1.2f%%', figsize=(8, 5))
dfDesbUnder
```

1.0



Out[39]:

	Atr	Qtd
0	Certificado	17687
1	Não Certificado	17687

In [40]:

```
dfAlgUnder = ExibirMedidas(X_under, X_test, y_under, y_test)
dfAlgUnder[colsExibir]
```

```
2019-04-20 10:13:46.305790
{'id_alg': 1, 'algoritmo': 'Tree'}
DecisionTreeClassifier(class_weight=None, criterion='gini', max_depth=5,
            max_features=None, max_leaf_nodes=None,
            min_impurity_decrease=0.0, min_impurity_split=None,
            min_samples_leaf=1, min_samples_split=2,
            min_weight_fraction_leaf=0.0, presort=False, random_state=Non
e,
            splitter='best')
2019-04-20 10:13:46.464221
{'id_alg': 2, 'algoritmo': 'KNN'}
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
           metric_params=None, n_jobs=None, n_neighbors=3, p=2,
           weights='uniform')
2019-04-20 10:13:53.853508
{'id_alg': 3, 'algoritmo': 'SVM'}
SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
  decision_function_shape='ovr', degree=3, gamma='auto_deprecated',
  kernel='sigmoid', max_iter=-1, probability=False, random_state=None,
  shrinking=True, tol=0.001, verbose=False)
2019-04-20 10:16:27.248606
{'id_alg': 4, 'algoritmo': 'MLP'}
MLPClassifier(activation='relu', alpha=0.0001, batch_size='auto', beta_1=
0.9,
       beta_2=0.999, early_stopping=False, epsilon=1e-08,
       hidden_layer_sizes=100, learning_rate='constant',
       learning_rate_init=0.001, max_iter=200, momentum=0.9,
       n_iter_no_change=10, nesterovs_momentum=True, power_t=0.5,
       random_state=None, shuffle=True, solver='adam', tol=0.0001,
       validation_fraction=0.1, verbose=False, warm_start=False)
2019-04-20 10:16:31.434311
{'id_alg': 5, 'algoritmo': 'Naive'}
GaussianNB(priors=None, var_smoothing=1e-09)
2019-04-20 10:16:31.554682
{'id_alg': 6, 'algoritmo': 'Dummy'}
DummyClassifier(constant=None, random_state=None, strategy='prior')
2019-04-20 10:16:31.568692
```

Out[40]:

	id_alg	algoritmo	acur	sens	esp	efic	VPP	VPN	TExec
0	1	Tree	88.53	88.65	88.42	88.53	88.37	88.70	0.00
1	2	KNN	93.69	94.17	93.22	93.69	93.14	94.24	0.12
2	3	SVM	50.03	0.00	50.03	25.02	0.00	100.00	2.56
3	4	MLP	65.84	99.53	59.45	79.49	31.79	99.85	0.07
4	5	Naive	75.94	87.88	69.74	78.81	60.15	91.71	0.00
5	6	Dummy	49.97	49.97	0.00	24.98	100.00	0.00	0.00

In [41]:

```
#metr_alg = ['sens', 'esp', 'acur', 'VPP', 'VPN', 'efic', 'totP', 'totN', 'totG' ]
#cols = ['id_alg', 'algoritmo', 'acur', 'sens', 'esp', 'efic', 'VPP', 'VPN', 'mat_con
f', 'AlgBin', 'TExec']

dfAlgUnder2 = ReExibirMedidas(dfAlgUnder, X_test, y_test)
dfAlgUnder2[colsExibir]
```

```
ReExibirMedidas
Tree
2019-04-20 10:35:44.445198
DecisionTreeClassifier(class_weight=None, criterion='gini', max_depth=5,
            max_features=None, max_leaf_nodes=None,
            min_impurity_decrease=0.0, min_impurity_split=None,
            min_samples_leaf=1, min_samples_split=2,
            min_weight_fraction_leaf=0.0, presort=False, random_state=Non
e,
            splitter='best')
KNN
2019-04-20 10:35:44.465225
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
           metric_params=None, n_jobs=None, n_neighbors=3, p=2,
           weights='uniform')
SVM
2019-04-20 10:35:51.003182
SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
  decision_function_shape='ovr', degree=3, gamma='auto_deprecated',
  kernel='sigmoid', max_iter=-1, probability=False, random_state=None,
  shrinking=True, tol=0.001, verbose=False)
MLP
2019-04-20 10:36:48.670322
MLPClassifier(activation='relu', alpha=0.0001, batch_size='auto', beta_1=
0.9,
       beta_2=0.999, early_stopping=False, epsilon=1e-08,
       hidden_layer_sizes=100, learning_rate='constant',
       learning_rate_init=0.001, max_iter=200, momentum=0.9,
       n_iter_no_change=10, nesterovs_momentum=True, power_t=0.5,
       random_state=None, shuffle=True, solver='adam', tol=0.0001,
       validation_fraction=0.1, verbose=False, warm_start=False)
Naive
2019-04-20 10:36:48.714450
GaussianNB(priors=None, var_smoothing=1e-09)
Dummy
2019-04-20 10:36:48.777620
DummyClassifier(constant=None, random_state=None, strategy='prior')
```

Out[41]:

	id_alg	algoritmo	acur	sens	esp	efic	VPP	VPN	TExec
0	1	Tree	88.53	88.65	88.42	88.53	88.37	88.70	0.00
1	2	KNN	93.69	94.17	93.22	93.69	93.14	94.24	0.11
2	3	SVM	50.03	0.00	50.03	25.02	0.00	100.00	0.96
3	4	MLP	65.84	99.53	59.45	79.49	31.79	99.85	0.00
4	5	Naive	75.94	87.88	69.74	78.81	60.15	91.71	0.00
5	6	Dummy	49.97	49.97	0.00	24.98	100.00	0.00	0.00

In [42]:

```
dfAlgOver2 = ReExibirMedidas(dfAlgOver, X_test, y_test)
dfAlgOver2[colsExibir]
```

```
ReExibirMedidas
Tree
2019-04-20 11:50:24.623913
DecisionTreeClassifier(class_weight=None, criterion='gini', max_depth=5,
            max_features=None, max_leaf_nodes=None,
            min_impurity_decrease=0.0, min_impurity_split=None,
            min_samples_leaf=1, min_samples_split=2,
            min_weight_fraction_leaf=0.0, presort=False, random_state=Non
e,
            splitter='best')
KNN
2019-04-20 11:50:24.644495
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
           metric_params=None, n_jobs=None, n_neighbors=3, p=2,
           weights='uniform')
SVM
2019-04-20 11:50:47.275784
SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
  decision_function_shape='ovr', degree=3, gamma='auto_deprecated',
  kernel='sigmoid', max_iter=-1, probability=False, random_state=None,
  shrinking=True, tol=0.001, verbose=False)
MLP
2019-04-20 12:25:42.203951
MLPClassifier(activation='relu', alpha=0.0001, batch_size='auto', beta_1=
0.9,
       beta_2=0.999, early_stopping=False, epsilon=1e-08,
       hidden_layer_sizes=100, learning_rate='constant',
       learning_rate_init=0.001, max_iter=200, momentum=0.9,
       n_iter_no_change=10, nesterovs_momentum=True, power_t=0.5,
       random_state=None, shuffle=True, solver='adam', tol=0.0001,
       validation_fraction=0.1, verbose=False, warm_start=False)
Naive
2019-04-20 12:25:42.366897
GaussianNB(priors=None, var_smoothing=1e-09)
Dummy
2019-04-20 12:25:42.430069
DummyClassifier(constant=None, random_state=None, strategy='prior')
```

Out[42]:

	id_alg	algoritmo	acur	sens	esp	efic	VPP	VPN	TExec
0	1	Tree	88.02	87.68	88.37	88.03	88.46	87.59	0.00
1	2	KNN	93.19	90.46	96.31	93.38	96.55	89.83	0.38
2	3	SVM	50.03	0.00	50.03	25.02	0.00	100.00	34.92
3	4	MLP	89.88	93.14	87.08	90.11	86.08	93.67	0.00
4	5	Naive	74.72	83.47	69.61	76.54	61.61	87.82	0.00
5	6	Dummy	49.97	49.97	0.00	24.98	100.00	0.00	0.00

In [43]:

```
dfAlg
```

Out[43]:

	id_alg	algoritmo	acur	sens	esp	efic	VPP	VPN	mat_conf	
0	1	Tree	97.20	97.20	0.00	48.60	100.00	0.00	[[186954, 0], [5388, 0]]	DecisionTreeClassifie
1	2	KNN	96.81	97.97	40.37	69.17	98.76	29.03	[[184644, 2310], [3824, 1564]]	KNeighborsClas
2	3	SVM	97.20	97.20	0.00	48.60	100.00	0.00	[[186954, 0], [5388, 0]]	SVC(C:
3	4	MLP	97.20	97.20	0.00	48.60	100.00	0.00	[[186954, 0], [5388, 0]]	MLPCla:
4	5	Naive	65.73	99.40	6.66	53.03	65.14	86.30	[[121780, 65174], [738, 4650]]	Gau
5	6	Dummy	97.20	97.20	0.00	48.60	100.00	0.00	[[186954, 0], [5388, 0]]	DummyCla

In [44]:

```
dfAlgUnder
```

Out[44]:

	id_alg	algoritmo	acur	sens	esp	efic	VPP	VPN	mat_conf	
0	1	Tree	88.53	88.65	88.42	88.53	88.37	88.70	[[4686, 617], [600, 4710]]	DecisionTreeCla
1	2	KNN	93.69	94.17	93.22	93.69	93.14	94.24	[[4939, 364], [306, 5004]]	KNeighbors
2	3	SVM	50.03	0.00	50.03	25.02	0.00	100.00	[[0, 5303], [0, 5310]]	SV
3	4	MLP	65.84	99.53	59.45	79.49	31.79	99.85	[[1686, 3617], [8, 5302]]	MLF
4	5	Naive	75.94	87.88	69.74	78.81	60.15	91.71	[[3190, 2113], [440, 4870]]	
5	6	Dummy	49.97	49.97	0.00	24.98	100.00	0.00	[[5303, 0], [5310, 0]]	Dummy

In [45]:

```
dfAlgOver
```

Out[45]:

	id_alg	algoritmo	acur	sens	esp	efic	VPP	VPN	mat_conf	
0	1	Tree	89.51	89.47	89.54	89.51	89.57	89.44	[[167669, 19521], [19726, 167155]]	DecisionTreeClassif
1	2	KNN	96.78	96.56	97.00	96.78	97.02	96.54	[[181607, 5583], [6463, 180418]]	KNeighborsCla
2	3	SVM	49.96	0.00	49.96	24.98	0.00	100.00	[[0, 187190], [0, 186881]]	SVC((
3	4	MLP	90.63	94.52	87.35	90.94	86.27	94.99	[[161493, 25697], [9363, 177518]]	MLPCla
4	5	Naive	76.60	88.63	70.26	79.45	61.06	92.15	[[114303, 72887], [14661, 172220]]	Ga
5	6	Dummy	50.04	50.04	0.00	25.02	100.00	0.00	[[187190, 0], [186881, 0]]	DummyCl

In [46]:

```
print('Original')
dfAlg[colsExibirMin]
```

```
Original
```

Out[46]:

	id_alg	algoritmo	acur	sens	esp	efic	TExec
0	1	Tree	97.20	97.20	0.00	48.60	0.04
1	2	KNN	96.81	97.97	40.37	69.17	7.60
2	3	SVM	97.20	97.20	0.00	48.60	27.78
3	4	MLP	97.20	97.20	0.00	48.60	9.97
4	5	Naive	65.73	99.40	6.66	53.03	0.04
5	6	Dummy	97.20	97.20	0.00	48.60	0.00

In [47]:

```
print('Oversample')
dfAlgOver[colsExibirMin]
```

Oversample

Out[47]:

	id_alg	algoritmo	acur	sens	esp	efic	TExec
0	1	Tree	89.51	89.47	89.54	89.51	0.18
1	2	KNN	96.78	96.56	97.00	96.78	47.88
2	3	SVM	49.96	0.00	49.96	24.98	3348.35
3	4	MLP	90.63	94.52	87.35	90.94	46.72
4	5	Naive	76.60	88.63	70.26	79.45	0.08
5	6	Dummy	50.04	50.04	0.00	25.02	0.01

In [48]:

```
print('Oversample - Teste com dados originais')
dfAlgOver2[colsExibirMin]
```

Oversample - Teste com dados originais

Out[48]:

	id_alg	algoritmo	acur	sens	esp	efic	TExec
0	1	Tree	88.02	87.68	88.37	88.03	0.00
1	2	KNN	93.19	90.46	96.31	93.38	0.38
2	3	SVM	50.03	0.00	50.03	25.02	34.92
3	4	MLP	89.88	93.14	87.08	90.11	0.00
4	5	Naive	74.72	83.47	69.61	76.54	0.00
5	6	Dummy	49.97	49.97	0.00	24.98	0.00

In [49]:

```
print('Undersample')
dfAlgUnder[colsExibirMin]
```

Undersample

Out[49]:

	id_alg	algoritmo	acur	sens	esp	efic	TExec
0	1	Tree	88.53	88.65	88.42	88.53	0.00
1	2	KNN	93.69	94.17	93.22	93.69	0.12
2	3	SVM	50.03	0.00	50.03	25.02	2.56
3	4	MLP	65.84	99.53	59.45	79.49	0.07
4	5	Naive	75.94	87.88	69.74	78.81	0.00
5	6	Dummy	49.97	49.97	0.00	24.98	0.00

In [50]:

```
print('Undersample - Teste com dados originais')
dfAlgUnder2[colsExibirMin]
```

Undersample - Teste com dados originais

Out[50]:

	id_alg	algoritmo	acur	sens	esp	efic	TExec
0	1	Tree	88.53	88.65	88.42	88.53	0.00
1	2	KNN	93.69	94.17	93.22	93.69	0.11
2	3	SVM	50.03	0.00	50.03	25.02	0.96
3	4	MLP	65.84	99.53	59.45	79.49	0.00
4	5	Naive	75.94	87.88	69.74	78.81	0.00
5	6	Dummy	49.97	49.97	0.00	24.98	0.00

In [51]:

```
dfAlgUnder2.T.to_dict()[0]['algoritmo']
```

Out[51]:

'Tree'

In [52]:

```
dill.dump_session('notebook_env.db')
```

In [ ]:

```

```