

2017-2018 - full scope - clustering

September 2, 2018

0.0.1 I) Loading data and preparing dataset - scope 2017-2018

```
In [1]: from pathlib import Path
import pandas as pd
import numpy as np
from datetime import datetime
import time
import matplotlib.pyplot as plt
%matplotlib inline
#%pylab inline
import itertools
import pickle
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import cross_val_score, GridSearchCV
from sklearn.decomposition import PCA
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import confusion_matrix, f1_score, precision_score, recall_score
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import LabelBinarizer
from sklearn.preprocessing import OneHotEncoder
from sklearn.svm import SVC
from sklearn.model_selection import StratifiedKFold
from sklearn.feature_selection import RFECV

In [2]: path_project = Path.home() / Path('Google Drive/Felix')
path_data = path_project / Path("data")
path_dump = path_project / Path("dump")

In [3]: # loading cdv data
file = path_data / Path("felix.csv")
with Path.open(file, 'rb') as fp:
    cdv = pd.read_csv(fp, encoding='cp1252', low_memory=False, index_col = 0)
# loading cdv data without format
file = path_data / Path("felix_ssfmt.csv")
with Path.open(file, 'rb') as fp:
    cdv_ssfmt = pd.read_csv(fp, encoding='cp1252', low_memory=False, index_col = 0)
```

```

        # loading MergeCommunesEnvi data
file = path_data / Path("MergeCommunesEnvi.csv")
with Path.open(file, 'rb') as fp:
    MergeCommunesEnvi = pd.read_csv(fp, encoding='cp1252', low_memory=False, sep=';', in

In [4]: # load various variable set
filename = path_dump / Path("dict_var_groups.sav")
with open(filename, 'rb') as fp:
    dict_var_groups = pickle.load(fp)

scope_2017_2018_ext_var = dict_var_groups['scope_2017_2018_ext_var']
scope_2017_2018_var = dict_var_groups['scope_2017_2018_var']

pred_var = dict_var_groups['pred_var']
com_var = dict_var_groups['com_var']
tech_var = dict_var_groups['tech_var']
text_var = dict_var_groups['text_var']
bizz_var = dict_var_groups['bizz_var']
cat_var = dict_var_groups['cat_var']
cat_max9_var = dict_var_groups['cat_max9_var']
cat_min10_var = dict_var_groups['cat_min10_var']
quant_var = dict_var_groups['quant_var']

In [5]: exclusion = com_var | tech_var | bizz_var | text_var
scope_2017_2018_var_kept = scope_2017_2018_ext_var - exclusion

cat_var_kept = cat_max9_var & scope_2017_2018_var_kept
scope_quant_var = (quant_var & scope_2017_2018_var_kept)
quant_null = np.sum(MergeCommunesEnvi.loc[:, scope_quant_var].isnull())
quant_var_kept = set(quant_null[quant_null < 200].index)

print(f"Out of {MergeCommunesEnvi.shape[1]} variable {len(scope_2017_2018_ext_var)} \
are used in 2017 and 2018 ")
print(f"{len(scope_2017_2018_ext_var & exclusion)} of 'technical' variable \
such as 'inseenum' are excluded ")
print(f"{len(scope_2017_2018_var_kept)} are remaining :")
print(f"\t{len(cat_var & scope_2017_2018_var_kept)} \
categorical variables : ")
print(f"\t\t{len(cat_max9_var & scope_2017_2018_var_kept)} \
with maximum 9 modalities ")
print(f"\t\t{len(cat_min10_var & scope_2017_2018_var_kept)} \
with more modalities ... excluded")
print(f"\t{len(quant_var & scope_2017_2018_var_kept)} \
variables are quantitative ")
print(f"\t\t{len(quant_var_kept)} have less than 200 missing values")
print(f"\t\t{len(scope_quant_var)-len(quant_var_kept)} \
have more ... excluded")

scope = cat_var_kept | quant_var_kept

```

Out of 571 variable 514 are used in 2017 and 2018
30 of 'technical' variable such as 'inseenum' are excluded
484 are remaining :

- 183 categorical variables :
 - 165 with maximum 9 modalities
 - 18 with more modalities ... excluded
- 301 variables are quantitative
 - 263 have less than 200 missing values
 - 38 have more ... excluded

```
In [6]: df = MergeCommunesEnvi.loc[MergeCommunesEnvi['ANNEEFUZ'].isin({2017,2018}),scope]
df.loc[:,scope_2017_2018_var & scope] = cdv_ssfmt.loc[:,scope_2017_2018_var & scope]
df.loc[:,cat_var_kept - {"HEUREUX"}] = cdv.loc[:,cat_var_kept - {"HEUREUX"}]
print(f"\nFinal number of variable kept : {df.shape[1]}")
```

Final number of variable kept : 428

```
In [7]: p = df.shape[1]
print(f"{p} columns out of which {len(cat_var_kept)-1} \
are corresponding to categorical features")
```

428 columns out of which 164 are corresponding to categorical features

```
In [8]: df = pd.get_dummies(df,
                           columns=cat_var_kept - {"HEUREUX"},
                           dummy_na = True,
                           drop_first=1)
```

```
In [9]: q = df.shape[1]
print(f"{q} columns after encoding of {len(cat_var_kept)-1} categorical \
variables in {len(cat_var_kept)-1+q-p} binary variables \
(K-1 one hot encoding)")
```

848 columns after encoding of 164 categorical variables in 584 binary variables (K-1 one hot encoding)

```
In [10]: df.shape
```

```
Out[10]: (6036, 848)
```

```
In [11]: # encoding of "HEUREUX" '[nsp]'
df.loc[df["HEUREUX"]==5,"HEUREUX"] = None
df = df.loc[np.isfinite(df['HEUREUX']).index,:]
```

```

# treating remaining missing values
features = df.columns.drop(['HEUREUX'])
df_tmp = df.loc[:,set(features) | {"HEUREUX"}].dropna()

X = df_tmp.loc[:,features]
y = df_tmp["HEUREUX"]

X_train, X_test, y_train, y_test = train_test_split(X,
                                                    y,
                                                    test_size=0.2,
                                                    random_state=42
                                                    )

scaler = StandardScaler().fit(X_train)
X_train = scaler.transform(X_train)
X_test = scaler.transform(X_test)

print(f"Number exemple: {y.shape[0]}\n\
- training set: {y_train.shape[0]}\n\
- test set: {y_test.shape[0]}")
print(f"Number of features: p={X_train.shape[1]}")
print(f"Number of class: {len(np.unique(y))}")
for c in np.unique(y):
    print(f"class {c:0.0f} : {100*np.sum(y==c)/len(y):0.1f}%")

```

```

Number exemple: 5624
- training set: 4499
- test set: 1125
Number of features: p=847
Number of class: 4
class 1 : 2.0%
class 2 : 34.8%
class 3 : 47.9%
class 4 : 15.3%

```

0.0.2 II) Feature selection

```

In [12]: startTime = time.time()

scoring='f1_macro'
step = 0.05

clf = LogisticRegression(C=1,
                        penalty='l1',
                        class_weight='balanced',
                        random_state=42)

```

```

rfecv = RFECV(estimator=clf, step=step, cv=StratifiedKFold(2),
              scoring=scoring)

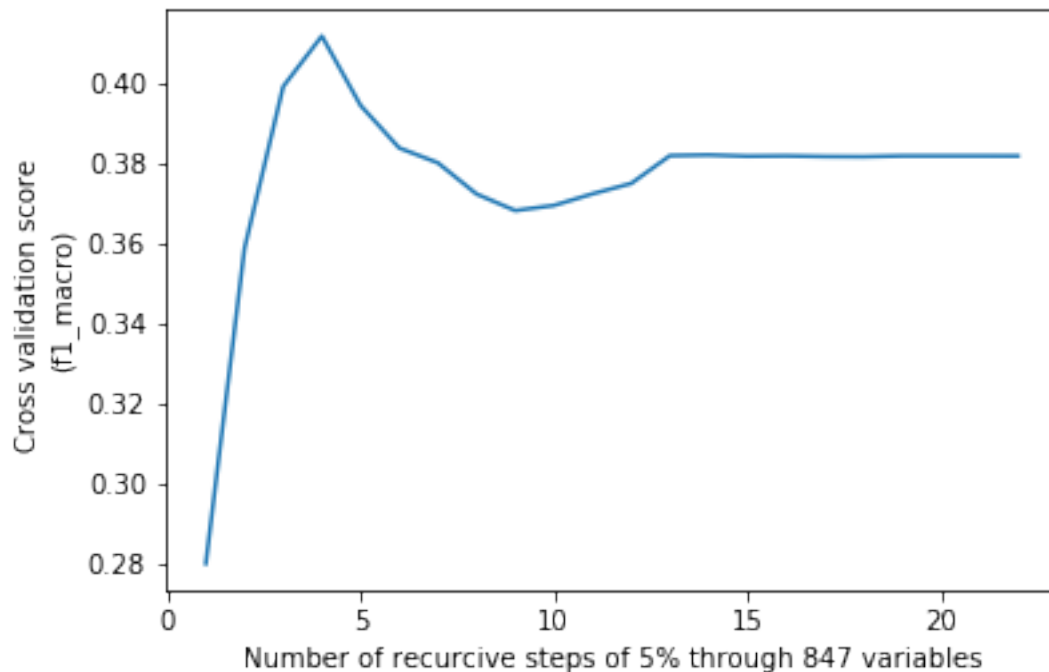
rfecv.fit(X_train, y_train)
print(f"done in {time.time() - startTime:0.1f} s")
print("Optimal number of features : %d" % rfecv.n_features_)

# Plot number of features VS. cross-validation scores
plt.figure()
plt.xlabel(f"Number of recursive steps of {100*step:0.0f}% through {X_train.shape[1]} v
plt.ylabel(f"Cross validation score \n({scoring})")
plt.plot(range(1, len(rfecv.grid_scores_) + 1), rfecv.grid_scores_)
plt.show()

//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)

```

done in 2295.9 s
Optimal number of features : 91



```

In [13]: lasso_mask = rfecv.support_
         X_train = X_train[:,lasso_mask]

```

```
X_test = X_test[:,lasso_mask]
print(f"Number of features: p={X_train.shape[1]}")
```

Number of features: p=91

```
In [14]: lasso_features = set(df.columns.drop(['HEUREUX'])[lasso_mask])
lasso_features
```

```
Out[14]: {'ACM3_Charge de logement négligeable',
'ACM4',
'ACM5',
'ACM7',
'ACM8',
'ACM9',
'AGE5',
'AGGL05_20 000 à 100 000 habitants',
'AGGL05_Plus de 100 000 habitants',
'AGGL09_50 000 à 100 000 habitants',
'ASSOENVI_Oui',
'ASSOPARE_[Nsp]',
'CADVIE',
'CADVIE3_Satisfait de son cadre de vie',
'CDV5',
'CDV5_4',
'CHOIXNUC_Plutôt des inconvénients',
"CHOVOLON_Plutôt pas d'accord",
'CHOVOLON_[Nsp]',
'CLASSES0_La classe moyenne supérieure',
'CLASSES0_Les défavorisés',
'CLASSES0_Les gens aisés',
'CLASSES0_Les privilégiés',
'DEPL0G',
'DIPL4_Diplômé du supérieur',
'DIPLOME_BEPC, brevet des collèges',
'ECHPOL_Plutôt à gauche',
'ETATSAN',
'FREQSPOR',
'INQAGRES',
'INQNUCLE',
'JUSTICE',
'LIEN_3_Frère, soeur',
'LIEN_3_Parent, beau-parent',
'LIEN_5_Enfant, beau-fils, belle-fille',
'LIEN_7_Frère, soeur',
'NB20_4_Un enfant de moins de 20 ans',
'NB99_4_Trois et plus enfants de 20 ans et plus',
'NB_A106',
```

'NB_A108',
 'NB_A119',
 'NB_A124',
 'NB_A401',
 'NB_D104',
 'NB_D108',
 'NB_D601',
 'NB_F102_NB_ECL',
 'NB_F107_NB_COU',
 'NB_F113_NB_COU',
 'NB_F120_NB_ECL',
 'NB_F121_NB_COU',
 'NB_F303_NB_SALLES',
 'NIVFRAN',
 'NIVPERSO',
 "OPIIMMIG_Le départ d'un grand nombre d'immigrés",
 'P15_RSECOCC',
 'PCSENQ8_Employé',
 "PREFPALI_Une amélioration de votre pouvoir d'achat",
 'PREOIMMI_Oui',
 'PREOMALA_Oui',
 'RAISPAUV_[Nsp]',
 'REV_TR7',
 'RE_WEB_Oui',
 'RURAUABA_PU',
 'SALCOMPC_nan',
 'SALCOMPI_Salarié',
 'SALCOMPI_nan',
 'SECUR3_Nsp, NR',
 'SENSIENV',
 'SEXE_4_Homme',
 'SITUEMP6_CDD, intérim',
 'SITUEMP6_Public CDI',
 'SITUEMP_Exerce un emploi',
 'SOUFFDEP_Oui',
 'SOUFFDOS_Oui',
 'STATLOGB_Locataire ou sous-locataire parc privé',
 'STATMAT_Séparé(e), divorcé(e)',
 'TEMPSTRA_Temps plein',
 'TYPCHAUF_Une installation individuelle propre à votre logement',
 "TYPEEMPL_D'une entreprise privée",
 'TYPOSQT_Classes moyennes inférieures',
 'TYPOSQT_Classes moyennes supérieures',
 'UDA10_Ouest',
 'UDA10_Sud-Ouest',
 'VISITFAM_Oui',
 'statut99_R',
 'zau1999_Commune espace dominante rurale',

```
'zau1999_Commune pôle urbain',
'zau2010_Commune appartenant à couronne d'un grand pôle",
'zau2010_Commune appartenant à grand pôle (10 000 emplois ou +)',
'zau2010_Commune isolée hors influence des pôles'}
```

0.0.3 III) Model valuation

a) Random Forest

```
In [15]: startTime = time.time()
n_estimators_range = [16,32,64,128,256]
max_depth_range = [2,4,8,16,32,64,128,256]
param_grid = dict(n_estimators=n_estimators_range, max_depth = max_depth_range)

params = {'max_features' : 'sqrt',
          'random_state' : 32,
          'min_samples_split' : 2,
          'class_weight' : 'balanced'
        }

clf = RandomForestClassifier(**params)

grid = GridSearchCV(clf,
                    scoring='f1_macro',
                    param_grid=param_grid)

grid.fit(X_train, y_train)
print(f"Determination of optimal hyperparameters in \
{time.time() - startTime:0.1f} s")
print(f"Optimal values are {grid.best_params_} \n\
F1 weighted Score of cross validation {100*grid.best_score_:0.2f}%")

# Learning on full training set with optimals hyperparameters and score on test set
params = {'max_features' : 'sqrt', 'random_state' : 32,
          'min_samples_split' : 2, 'class_weight' : 'balanced',
          'n_estimators' : grid.best_params_['n_estimators'],
          'max_depth' : grid.best_params_['max_depth']}
clf = RandomForestClassifier(**params).fit(X_train, y_train)

//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
```



```

'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)

```

Determination of optimal hyperparameters in 110.8 s
 Optimal values are {'max_depth': 4, 'n_estimators': 128}
 F1 weighted Score of cross validation 40.55%

```

In [16]: def plot_confusion_matrix(cm, classes,
                                   normalize=False,
                                   title='Confusion matrix',
                                   cmap=plt.cm.Blues):
    """
    This function prints and plots the confusion matrix.
    Normalization can be applied by setting `normalize=True`.
    """
    if normalize:
        cm = cm.astype('float') / cm.sum(axis=1)[:, np.newaxis]
        print("Normalized confusion matrix")
    else:
        print('Confusion matrix, without normalization')

    print(cm)

    plt.imshow(cm, interpolation='nearest', cmap=cmap)
    plt.title(title)
    plt.colorbar()
    tick_marks = np.arange(len(classes))
    plt.xticks(tick_marks, classes, rotation=45)
    plt.yticks(tick_marks, classes)

    fmt = '.2f' if normalize else 'd'
    thresh = cm.max() / 2.
    for i, j in itertools.product(range(cm.shape[0]), range(cm.shape[1])):
        plt.text(j, i, format(cm[i, j], fmt),
                 horizontalalignment="center",
                 color="white" if cm[i, j] > thresh else "black")

```

```

plt.tight_layout()
plt.ylabel('True label')
plt.xlabel('Predicted label')

In [20]: # Model evaluation
class_names = ["Jamais",
               "Occasionnellement",
               "Assez souvent",
               "Très souvent" ]
y_test_pred = clf.predict(X_test)
f1_scores = f1_score(y_test, y_test_pred, labels = [1,2,3,4], average=None)
for i,c in enumerate(class_names):
    print(f"f1 score class '{c}' : {100*f1_scores[i]:0.1f}%")
accuracy = clf.score(X_test, y_test)
f1_macro = f1_score(y_test, y_test_pred, average='macro')
f1_weighted = f1_score(y_test, y_test_pred, average='weighted')
print(f"Average scores :\nf1 macro : {f1_macro*100:0.4f} %\n\
f1 weighted : {f1_weighted*100:0.4f} %\naccuracy : {accuracy*100:0.4f} %")

# Compute confusion matrix
cnf_matrix = confusion_matrix(y_test, y_test_pred)
np.set_printoptions(precision=2)

# Plot non-normalized confusion matrix
plt.figure()
plot_confusion_matrix(cnf_matrix, classes=class_names,
                      title='Confusion matrix, without normalization')

# Plot normalized confusion matrix
plt.figure()
plot_confusion_matrix(cnf_matrix, classes=class_names, normalize=True,
                      title='Normalized confusion matrix')

plt.show()

# save performance
score_rf = {
    'f1_scores' : f1_scores,
    'accuracy' : accuracy,
    'f1_macro' : f1_macro,
    'f1_weighted' : f1_weighted
}

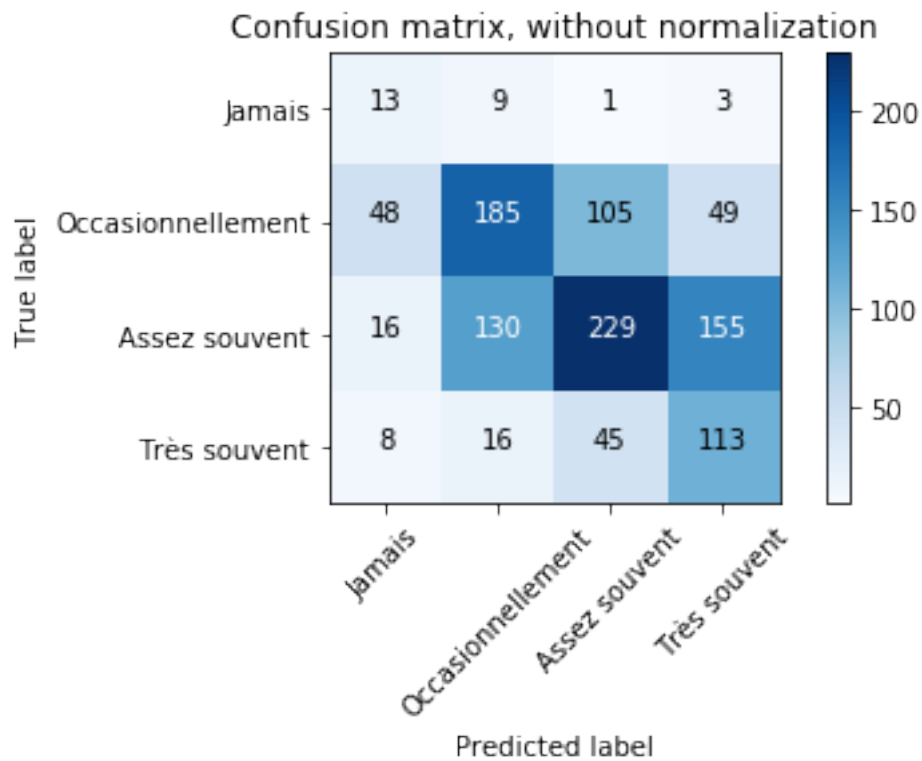
f1 score class 'Jamais' : 23.4%
f1 score class 'Occasionnellement' : 50.9%
f1 score class 'Assez souvent' : 50.3%
f1 score class 'Très souvent' : 45.0%
Average scores :

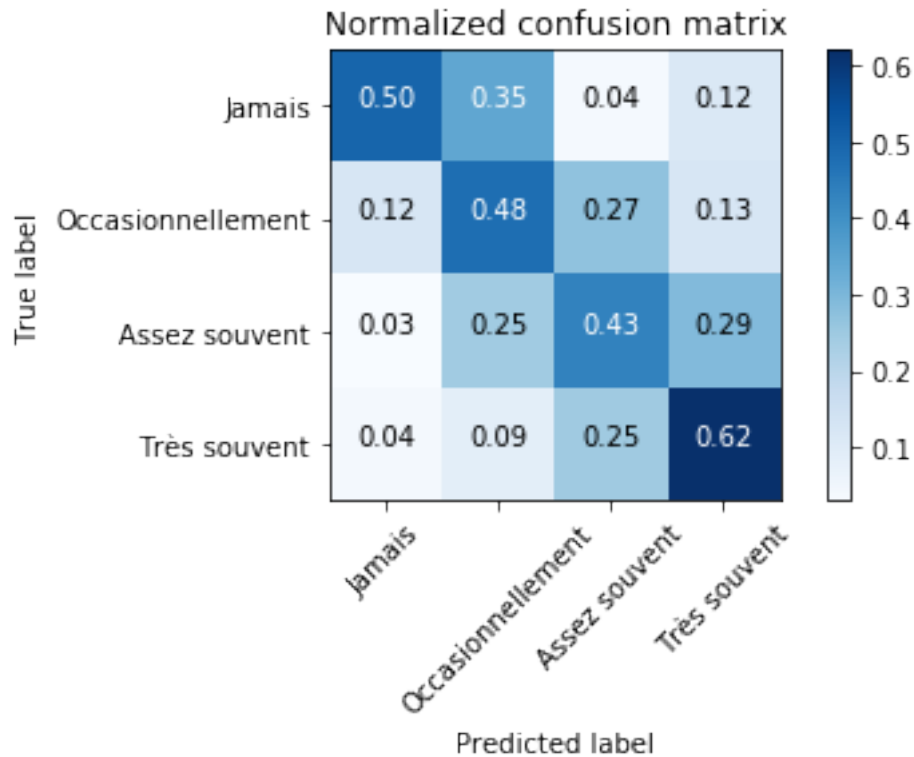
```

```

f1 macro : 42.4168 %
f1 weighted : 49.0430 %
accuracy : 48.0000 %
Confusion matrix, without normalization
[[ 13   9   1   3]
 [ 48 185 105  49]
 [ 16 130 229 155]
 [  8  16  45 113]]
Normalized confusion matrix
[[ 0.5  0.35 0.04 0.12]
 [ 0.12 0.48 0.27 0.13]
 [ 0.03 0.25 0.43 0.29]
 [ 0.04 0.09 0.25 0.62]]

```





0.04 IV) Load, learn and valuate model on clusters

```
In [21]: # loading cdv data
file = path_data / Path("clustTest1.csv")
with Path.open(file, 'rb') as fp:
    clustTest1 = pd.read_csv(fp, encoding='utf-8', low_memory=False, sep=";", index_col=0)
```

```
In [22]: clustTest1.head(2)
```

```
Out[22]:
```

	clust1	clust2	clust3	clust4	clust5
INTER6					
390001	1	4	5	1.0	2.0
390002	2	6	4	7.0	5.0

```
In [23]: #score = dict()

n_estimators_range = [16,32,64,128]
max_depth_range = [2,4,8,16,32,64]
param_grid = dict(n_estimators=n_estimators_range, max_depth = max_depth_range)
params = {'max_features' : 'sqrt',
          'random_state' : 32,
          'min_samples_split' : 2,
          'class_weight' : 'balanced'}
```

```

    }

score_clustering_methods = []
clustering_methods = clustTest1.columns[0:3]

for method in clustering_methods:
    print(f"\nAnalysis cluster method {method}")
    cluster_list = clustTest1[method].unique()
    print(f"liste of clusters : {cluster_list}")
    score_cluster = []
    for cluster in cluster_list:
        index_scope = clustTest1.loc[clustTest1[method]==cluster,:].index
        print(f"cluster {cluster} : {len(index_scope)} elements")

        # treating remaining missing values
        features = df.columns.drop(['HEUREUX'])[lasso_mask]
        df_tmp = df.loc[index_scope, set(features) | {"HEUREUX"}].dropna()

        X = df_tmp.loc[:, features]
        y = df_tmp["HEUREUX"]

        X_train, X_test, y_train, y_test = train_test_split(X,
                                                            y,
                                                            test_size=0.2,
                                                            random_state=42
                                                            )

        scaler = StandardScaler().fit(X_train)
        X_train = scaler.transform(X_train)
        X_test = scaler.transform(X_test)

        print(f"Number exemple: {y.shape[0]}\n\
- training set: {y_train.shape[0]}\n\
- test set: {y_test.shape[0]}")
        print(f"Number of features: p={X_train.shape[1]}")
        print(f"Number of class: {len(np.unique(y))}")
        for c in np.unique(y):
            print(f"class {c:0.0f} : {100*np.sum(y==c)/len(y):0.1f}%")

        startTime = time.time()
        clf = RandomForestClassifier(**params)
        grid = GridSearchCV(clf,
                            scoring='f1_macro',
                            param_grid=param_grid)

        grid.fit(X_train, y_train)
        print(f"Optimal values are {grid.best_params_} \n\

```

```

F1 weighted Score of cross validation {100*grid.best_score_:0.2f}%")

# Learning on full training set with optimals hyperparameters and score on test
params_opt = {'max_features' : 'sqrt', 'random_state' : 32,
              'min_samples_split' : 2, 'class_weight' : 'balanced',
              'n_estimators' : grid.best_params_['n_estimators'],
              'max_depth' : grid.best_params_['max_depth']}
clf = RandomForestClassifier(**params_opt).fit(X_train, y_train)

y_test_pred = clf.predict(X_test)
f1_scores = f1_score(y_test, y_test_pred, labels = [1,2,3,4], average=None)
accuracy = clf.score(X_test, y_test)
f1_macro = f1_score(y_test, y_test_pred, average='macro')
f1_weighted = f1_score(y_test, y_test_pred, average='weighted')

res = {'f1_scores' : f1_scores,
       'accuracy' : accuracy,
       'f1_macro' : f1_macro,
       'f1_weighted' : f1_weighted}

cl = {'cluster' : cluster,
      'model' : 'RandomForestClassifier',
      'params' : params_opt,
      'metrics' : res
      }

score_cluster.append(cl)

d = {'clustering_method' : method,
     'cluster_scores' : score_cluster
     }
score_clustering_methods.append(d)

```

```

Analysis cluster method clust1
liste of clusters : [1 2 3 4 5 6]
cluster 1 : 295 elements
Number exemple: 286
    - training set: 228
    - test set: 58
Number of features: p=91
Number of class: 4
class 1 : 5.2%
class 2 : 40.9%
class 3 : 38.8%
class 4 : 15.0%

```

[illegible]

[illegible]

```
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: UndefinedVariableWarning:
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: UndefinedVariableWarning:
'precision', 'predicted', average, warn_for)
```

Optimal values are {'max_depth': 4, 'n_estimators': 128}

F1 weighted Score of cross validation 42.24%

cluster 3 : 3633 elements

Number exemple: 3529

- training set: 2823

- test set: 706

Number of features: p=91

Number of class: 4

class 1 : 1.5%

class 2 : 33.1%

class 3 : 49.4%

class 4 : 16.0%

```
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: UndefinedVariableWarning:
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: UndefinedVariableWarning:
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: UndefinedVariableWarning:
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: UndefinedVariableWarning:
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: UndefinedVariableWarning:
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: UndefinedVariableWarning:
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: UndefinedVariableWarning:
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: UndefinedVariableWarning:
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: UndefinedVariableWarning:
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: UndefinedVariableWarning:
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: UndefinedVariableWarning:
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: UndefinedVariableWarning:
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: UndefinedVariableWarning:
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: UndefinedVariableWarning:
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: UndefinedVariableWarning:
'precision', 'predicted', average, warn_for)
```

[illegible]

Optimal values are {'max_depth': 4, 'n_estimators': 128}

F1 weighted Score of cross validation 40.73%

cluster 4 : 218 elements

Number exemple: 214

- training set: 171

- test set: 43

Number of features: p=91

Number of class: 4

class 1 : 0.9%

class 2 : 27.6%

class 3 : 50.0%

class 4 : 21.5%

```
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/model_selection/_split.py:605: Warning:
% (min_groups, self.n_splits)), Warning)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
```

[illegible]

```
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
```

Optimal values are {'max_depth': 2, 'n_estimators': 128}

F1 weighted Score of cross validation 44.93%

cluster 5 : 137 elements

Number exemple: 134

- training set: 107

- test set: 27

Number of features: p=91

Number of class: 4

```
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1137: Undefined
'recall', 'true', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
```

[illegible]

[illegible]

```

//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)

class 1 : 2.2%
class 2 : 38.1%
class 3 : 44.8%
class 4 : 14.9%
Optimal values are {'max_depth': 8, 'n_estimators': 64}
    F1 weighted Score of cross validation 32.56%
cluster 6 : 24 elements
Number exemple: 22
    - training set: 17
    - test set: 5
Number of features: p=91
Number of class: 4
class 1 : 4.5%
class 2 : 40.9%
class 3 : 31.8%
class 4 : 22.7%

//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1137: Undefined
'recall', 'true', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/model_selection/_split.py:605: Warning:
% (min_groups, self.n_splits)), Warning)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)

```

[illegible]

[illegible]

```
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
```

Optimal values are {'max_depth': 2, 'n_estimators': 128}
F1 weighted Score of cross validation 39.28%

Analysis cluster method clust2

```
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1137: Undefined
'recall', 'true', average, warn_for)
```

liste of clusters : [4 6 5 1 3 2 7]
cluster 4 : 212 elements
Number exemple: 205
- training set: 164
- test set: 41
Number of features: p=91
Number of class: 4
class 1 : 7.8%
class 2 : 42.4%
class 3 : 32.2%
class 4 : 17.6%

[illegible]

```
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
```

Optimal values are {'max_depth': 2, 'n_estimators': 16}

F1 weighted Score of cross validation 39.99%

cluster 6 : 1137 elements

Number exemple: 1108

- training set: 886

- test set: 222

Number of features: p=91

Number of class: 4

class 1 : 1.8%

class 2 : 33.1%

class 3 : 48.5%

class 4 : 16.6%

```
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
```

[illegible]


```
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
```

Optimal values are {'max_depth': 4, 'n_estimators': 128}

F1 weighted Score of cross validation 43.79%

cluster 5 : 750 elements

Number exemple: 741

- training set: 592

- test set: 149

Number of features: p=91

Number of class: 4

class 1 : 1.5%

class 2 : 32.7%

class 3 : 48.9%

class 4 : 17.0%

```
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
```

```
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
```

Optimal values are {'max_depth': 4, 'n_estimators': 128}
F1 weighted Score of cross validation 48.43%

```
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
```

```
cluster 1 : 1257 elements
Number exemple: 1191
    - training set: 952
    - test set: 239
Number of features: p=91
Number of class: 4
class 1 : 1.8%
class 2 : 37.7%
class 3 : 47.0%
class 4 : 13.4%
```

```
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
```

[illegible]

```
Number of class: 4
class 1 : 1.2%
class 2 : 31.5%
class 3 : 52.6%
class 4 : 14.6%
```

[illegible]

[illegible]

```

    'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
    'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
    'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
    'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
    'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
    'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
    'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
    'precision', 'predicted', average, warn_for)

```

Optimal values are {'max_depth': 4, 'n_estimators': 64}

F1 weighted Score of cross validation 38.63%

cluster 2 : 857 elements

Number exemple: 839

- training set: 671

- test set: 168

Number of features: p=91

Number of class: 4

class 1 : 1.7%

class 2 : 31.8%

class 3 : 49.7%

class 4 : 16.8%

```

//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
    'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
    'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
    'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
    'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
    'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
    'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
    'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
    'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
    'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
    'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
    'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
    'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
    'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
    'precision', 'predicted', average, warn_for)

```

[illegible]

[illegible]

Optimal values are {'max_depth': 4, 'n_estimators': 128}

F1 weighted Score of cross valdation 38.15%

```
cluster 7 : 569 elements
```

Number example: 558

- training set: 446

- test set: 112

Number of features: $p=91$

Number of class: 4

```
class 1 : 2.9%
```

```
class 2 : 43.0%
```

```
class 3 : 40.9%
```

```
class 4 : 13.3%
```

[illegible]

[illegible]

[illegible]

Optimal values are {'max_depth': 4, 'n_estimators': 64}
F1 weighted Score of cross validation 36.72%

```
Analysis cluster method clust3
liste of clusters : [5 4 1 2 3]
cluster 5 : 373 elements
Number exemple: 370
    - training set: 296
    - test set: 74
Number of features: p=91
Number of class: 4
class 1 : 4.3%
class 2 : 41.4%
class 3 : 37.6%
class 4 : 16.8%
```

```
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
```

[illegible]

[illegible]

[illegible]

Optimal values are {'max_depth': 4, 'n_estimators': 64}

F1 weighted Score of cross valdation 40.84%

```
cluster 1 : 1593 elements
```

Number exemple: 1526

- training set: 1220

- test set: 306

Number of features: $p=91$

Number of class: 4

```
class 1 : 2.0%
```

```
class 2 : 35.1%
```

```
class 3 : 48.4%
```

```
class 4 : 14.5%
```

```
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: UndefinedVariableWarning:
'precision', 'predicted', average, warn_for)
```

[illegible]

Optimal values are {'max_depth': 4, 'n_estimators': 128}

F1 weighted Score of cross validation 41.76%

cluster 2 : 1246 elements

Number exemple: 1216

- training set: 972

- test set: 244

Number of features: p=91

Number of class: 4

class 1 : 1.8%

class 2 : 33.0%

class 3 : 48.8%

class 4 : 16.4%

```
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
```


[illegible]

```
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
```

Optimal values are {'max_depth': 4, 'n_estimators': 128}

F1 weighted Score of cross validation 40.98%

cluster 3 : 142 elements

Number exemple: 138

- training set: 110

- test set: 28

Number of features: p=91

Number of class: 4

class 1 : 5.1%

class 2 : 39.9%

class 3 : 42.0%

class 4 : 13.0%

```
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
```

[illegible]

[illegible]

```
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
```

Optimal values are {'max_depth': 4, 'n_estimators': 32}
 F1 weighted Score of cross validation 27.28%

```
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
//anaconda/envs/py36/lib/python3.6/site-packages/sklearn/metrics/classification.py:1135: Undefined
'precision', 'predicted', average, warn_for)
```

```
In [26]: print(f"F1 macro on full dataset : {100*score_rf['f1_macro']:0.1f}%")
        for score_method in score_clustering_methods:
            print(f"method {score_method['clustering_method']}:")
            average_score = 0
            for i, score_cluster in enumerate(score_method['cluster_scores']):
                print(f"cluster {score_cluster['cluster']}, f1 macro {100*score_cluster['metric']}:")
                average_score = average_score + score_cluster['metrics']['f1_macro']
            average_score = average_score / (i+1)
            print(f"average f1 macro on clusters {100*average_score:0.1f}%")
```

F1 macro on full dataset : 42.4%
 method clust1:
 cluster 1, f1 macro 34.6%
 cluster 2, f1 macro 40.4%
 cluster 3, f1 macro 42.6%
 cluster 4, f1 macro 44.2%
 cluster 5, f1 macro 26.4%
 cluster 6, f1 macro 61.1%
 average f1 macro on clusters 41.6%
 method clust2:
 cluster 4, f1 macro 36.8%
 cluster 6, f1 macro 37.1%
 cluster 5, f1 macro 41.5%
 cluster 1, f1 macro 32.9%
 cluster 3, f1 macro 33.2%

cluster 2, f1 macro 30.8%
cluster 7, f1 macro 48.3%
average f1 macro on clusters 37.2%
method clust3:
cluster 5, f1 macro 43.2%
cluster 4, f1 macro 38.7%
cluster 1, f1 macro 30.5%
cluster 2, f1 macro 48.0%
cluster 3, f1 macro 32.5%
average f1 macro on clusters 38.6%