

milestone4

June 19, 2022

```
[1]: import pandas as pd
from sqlalchemy import create_engine

[18]: from sklearn.ensemble import RandomForestRegressor
from sklearn.multioutput import MultiOutputClassifier
from sklearn.pipeline import Pipeline
import pickle
from sklearn.model_selection import train_test_split
from sklearn.model_selection import GridSearchCV
from sklearn.metrics import confusion_matrix
from sklearn import metrics
import numpy as np

[3]: #check sqlalchemy
import sqlalchemy
print (sqlalchemy.__version__ )
```

1.4.22

```
[53]: #load data from database
def load_data():
    engine = create_engine('sqlite:///casas.db')
    df = pd.read_sql("SELECT * FROM casas", engine)
    X = df[['apartments', 'houses', 'areas_sqm', 'quartos', 'Latitude',
    → 'Longitude', 'CodPostal', 'Cod_Condition', 'T0', 'T1', 'T2', 'T3', 'T4',
    → 'T5', 'T6']]
    y = df[['Price']]
    return X, y

[54]: #call function
X,y = load_data()

[55]: X.head()
```

```
[55]:
```

	apartments	houses	areas_sqm	quartos	Latitude	Longitude	CodPostal	\
0	1	0	96	2	40.6175	-8.647778	3810	
1	1	0	119	2	40.6175	-8.647778	3810	
2	1	0	72	2	40.6175	-8.647778	3810	
3	1	0	90	2	40.6175	-8.647778	3810	

4	1	0	126	2	40.6175	-8.647778	3810
---	---	---	-----	---	---------	-----------	------

	Cod_Condition	T0	T1	T2	T3	T4	T5	T6
0	4	0	0	1	0	0	0	0
1	4	0	0	1	0	0	0	0
2	4	0	0	1	0	0	0	0
3	4	0	0	1	0	0	0	0
4	6	0	0	1	0	0	0	0

```
[56]: y.head()
```

```
[56]: Price
0    145000
1    155000
2    118000
3    185000
4    295000
```

```
[7]: X.shape, y.shape
```

```
[7]: ((4180, 15), (4180, 1))
```

```
[8]: #split data
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.3,
→random_state=42)
```

0.0.1 Substituir pelo modelo correto

```
[9]: pipeline=Pipeline([
    ('clf',RandomForestRegressor())
])
```

```
[31]: pipeline.get_params().keys()
```

```
[31]: dict_keys(['memory', 'steps', 'verbose', 'clf', 'clf__bootstrap',
'clf__ccp_alpha', 'clf__criterion', 'clf__max_depth', 'clf__max_features',
'clf__max_leaf_nodes', 'clf__max_samples', 'clf__min_impurity_decrease',
'clf__min_impurity_split', 'clf__min_samples_leaf', 'clf__min_samples_split',
'clf__min_weight_fraction_leaf', 'clf__n_estimators', 'clf__n_jobs',
'clf__oob_score', 'clf__random_state', 'clf__verbose', 'clf__warm_start'])
```

```
[10]: #choose parameters
parameters = {
    "clf__n_estimators": [10, 50, 100]
}
```

```
[11]: # use GridSearchCV
cv = GridSearchCV(pipeline, param_grid=parameters, verbose=1)
```

```
[12]: #fit model
model=cv.fit(X_train, y_train)
```

Fitting 5 folds for each of 3 candidates, totalling 15 fits

```
/Users/anateresaneto/opt/miniconda3/envs/my_env/lib/python3.9/site-  
packages/sklearn/pipeline.py:346: DataConversionWarning: A column-vector y was  
passed when a 1d array was expected. Please change the shape of y to  
(n_samples,), for example using ravel().  
    self._final_estimator.fit(Xt, y, **fit_params_last_step)  
/Users/anateresaneto/opt/miniconda3/envs/my_env/lib/python3.9/site-  
packages/sklearn/pipeline.py:346: DataConversionWarning: A column-vector y was  
passed when a 1d array was expected. Please change the shape of y to  
(n_samples,), for example using ravel().  
    self._final_estimator.fit(Xt, y, **fit_params_last_step)  
/Users/anateresaneto/opt/miniconda3/envs/my_env/lib/python3.9/site-  
packages/sklearn/pipeline.py:346: DataConversionWarning: A column-vector y was  
passed when a 1d array was expected. Please change the shape of y to  
(n_samples,), for example using ravel().  
    self._final_estimator.fit(Xt, y, **fit_params_last_step)  
/Users/anateresaneto/opt/miniconda3/envs/my_env/lib/python3.9/site-  
packages/sklearn/pipeline.py:346: DataConversionWarning: A column-vector y was  
passed when a 1d array was expected. Please change the shape of y to  
(n_samples,), for example using ravel().  
    self._final_estimator.fit(Xt, y, **fit_params_last_step)  
/Users/anateresaneto/opt/miniconda3/envs/my_env/lib/python3.9/site-  
packages/sklearn/pipeline.py:346: DataConversionWarning: A column-vector y was  
passed when a 1d array was expected. Please change the shape of y to  
(n_samples,), for example using ravel().  
    self._final_estimator.fit(Xt, y, **fit_params_last_step)  
/Users/anateresaneto/opt/miniconda3/envs/my_env/lib/python3.9/site-  
packages/sklearn/pipeline.py:346: DataConversionWarning: A column-vector y was  
passed when a 1d array was expected. Please change the shape of y to  
(n_samples,), for example using ravel().  
    self._final_estimator.fit(Xt, y, **fit_params_last_step)  
/Users/anateresaneto/opt/miniconda3/envs/my_env/lib/python3.9/site-  
packages/sklearn/pipeline.py:346: DataConversionWarning: A column-vector y was  
passed when a 1d array was expected. Please change the shape of y to  
(n_samples,), for example using ravel().  
    self._final_estimator.fit(Xt, y, **fit_params_last_step)  
/Users/anateresaneto/opt/miniconda3/envs/my_env/lib/python3.9/site-  
packages/sklearn/pipeline.py:346: DataConversionWarning: A column-vector y was  
passed when a 1d array was expected. Please change the shape of y to  
(n_samples,), for example using ravel().  
    self._final_estimator.fit(Xt, y, **fit_params_last_step)  
/Users/anateresaneto/opt/miniconda3/envs/my_env/lib/python3.9/site-
```

```

packages/sklearn/pipeline.py:346: DataConversionWarning: A column-vector y was
passed when a 1d array was expected. Please change the shape of y to
(n_samples,), for example using ravel().
    self._final_estimator.fit(Xt, y, **fit_params_last_step)
/Users/anateresaneto/opt/miniconda3/envs/my_env/lib/python3.9/site-
packages/sklearn/pipeline.py:346: DataConversionWarning: A column-vector y was
passed when a 1d array was expected. Please change the shape of y to
(n_samples,), for example using ravel().
    self._final_estimator.fit(Xt, y, **fit_params_last_step)
/Users/anateresaneto/opt/miniconda3/envs/my_env/lib/python3.9/site-
packages/sklearn/pipeline.py:346: DataConversionWarning: A column-vector y was
passed when a 1d array was expected. Please change the shape of y to
(n_samples,), for example using ravel().
    self._final_estimator.fit(Xt, y, **fit_params_last_step)
/Users/anateresaneto/opt/miniconda3/envs/my_env/lib/python3.9/site-
packages/sklearn/pipeline.py:346: DataConversionWarning: A column-vector y was
passed when a 1d array was expected. Please change the shape of y to
(n_samples,), for example using ravel().
    self._final_estimator.fit(Xt, y, **fit_params_last_step)
/Users/anateresaneto/opt/miniconda3/envs/my_env/lib/python3.9/site-
packages/sklearn/pipeline.py:346: DataConversionWarning: A column-vector y was
passed when a 1d array was expected. Please change the shape of y to
(n_samples,), for example using ravel().
    self._final_estimator.fit(Xt, y, **fit_params_last_step)
/Users/anateresaneto/opt/miniconda3/envs/my_env/lib/python3.9/site-
packages/sklearn/pipeline.py:346: DataConversionWarning: A column-vector y was
passed when a 1d array was expected. Please change the shape of y to
(n_samples,), for example using ravel().
    self._final_estimator.fit(Xt, y, **fit_params_last_step)
/Users/anateresaneto/opt/miniconda3/envs/my_env/lib/python3.9/site-
packages/sklearn/pipeline.py:346: DataConversionWarning: A column-vector y was
passed when a 1d array was expected. Please change the shape of y to
(n_samples,), for example using ravel().
    self._final_estimator.fit(Xt, y, **fit_params_last_step)

```

```

[13]: #predict
y_pred=model.predict(X_test)

```

```

[27]: def build_model():
    pipeline = Pipeline([
        ('clf',MultiOutputClassifier(RandomForestRegressor()))
    ])
    parameters = {
        'clf__estimator__n_estimators':[50, 100, 150]
    }
    #create grid_search
    cv = GridSearchCV(pipeline, param_grid = parameters, verbose = 1)

```

```
return cv
```

```
[28]: model = build_model()
```

```
[29]: print('MAE:', metrics.mean_absolute_error(y_test, y_pred))
      print('MSE:', metrics.mean_squared_error(y_test, y_pred))
      print('RMSE:', np.sqrt(metrics.mean_squared_error(y_test, y_pred)))
      print('VarScore:', metrics.explained_variance_score(y_test, y_pred))
```

MAE: 45821.41872795983

MSE: 7166046811.822071

RMSE: 84652.50623473631

VarScore: 0.5834095873470151

```
[51]: file = 'casas.pkl'
      pickle.dump(model, open(file, 'wb'))
```

```
[ ]: # desenvolver o mapa
```

```
[34]: import numpy as np # library to handle data in a vectorized manner

import pandas as pd # library for data analysis
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import json # library to handle JSON files

# transforming json file into a pandas dataframe library
from pandas.io.json import json_normalize

#!conda install -c conda-forge geopy --yes # uncomment this line if you haven't
→completed the Foursquare API lab
#from geopy.geocoders import Nominatim # convert an address into latitude and
→longitude values

import requests # library to handle requests
from pandas.io.json import json_normalize # transform JSON file into a pandas
→dataframe

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors

# import k-means from clustering stage
from sklearn.cluster import KMeans

#!conda install -c conda-forge folium=0.5.0 --yes
import folium # map rendering library
```

```
print('Libraries imported.')
```

Libraries imported.

```
[35]: !pip install shapely
import shapely.geometry

!pip install pyproj
import pyproj

import math

def lonlat_to_xy(lon, lat):
    proj_latlon = pyproj.Proj(proj='latlong',datum='WGS84')
    proj_xy = pyproj.Proj(proj="utm", zone=33, datum='WGS84')
    xy = pyproj.transform(proj_latlon, proj_xy, lon, lat)
    return xy[0], xy[1]

def xy_to_lonlat(x, y):
    proj_latlon = pyproj.Proj(proj='latlong',datum='WGS84')
    proj_xy = pyproj.Proj(proj="utm", zone=33, datum='WGS84')
    lonlat = pyproj.transform(proj_xy, proj_latlon, x, y)
    return lonlat[0], lonlat[1]

def calc_xy_distance(x1, y1, x2, y2):
    dx = x2 - x1
    dy = y2 - y1
    return math.sqrt(dx*dx + dy*dy)
```

Collecting shapely

Downloading Shapely-1.8.2-cp39-cp39-macosx_10_9_x86_64.whl (1.2 MB)

|| 1.2 MB 4.1 MB/s eta 0:00:01

Installing collected packages: shapely

Successfully installed shapely-1.8.2

Collecting pyproj

Downloading pyproj-3.3.1-cp39-cp39-macosx_10_9_x86_64.whl (8.2 MB)

|| 8.2 MB 4.4 MB/s eta 0:00:01

Requirement already satisfied: certifi in

./opt/miniconda3/envs/my_env/lib/python3.9/site-packages (from pyproj)
(2022.6.15)

Installing collected packages: pyproj

Successfully installed pyproj-3.3.1

```
[36]: aveiro_center = 40.6410163 , -8.653466
address = 'Forum Aveiro'
print('Coordinate of {}: {}'.format(address, aveiro_center))
```

Coordinate of Forum Aveiro: (40.6410163, -8.653466)

```
[40]: aveiro_center_x, aveiro_center_y = lonlat_to_xy(aveiro_center[1],  
→aveiro_center[0]) # City center in Cartesian coordinates  
  
k = math.sqrt(3) / 2 # Vertical offset for hexagonal grid cells  
x_min = aveiro_center_x - 3000  
x_step = 300  
y_min = aveiro_center_y - 3000 - (int(21/k)*k*300 - 6000)/2  
y_step = 300 * k  
  
latitudes = []  
longitudes = []  
distances_from_center = []  
xs = []  
ys = []  
for i in range(0, int(21/k)):  
    y = y_min + i * y_step  
    x_offset = 150 if i%2==0 else 0  
    for j in range(0, 21):  
        x = x_min + j * x_step + x_offset  
        distance_from_center = calc_xy_distance(aveiro_center_x,  
→aveiro_center_y, x, y)  
        if (distance_from_center <= 3001):  
            lon, lat = xy_to_lonlat(x, y)  
            latitudes.append(lat)  
            longitudes.append(lon)  
            distances_from_center.append(distance_from_center)  
            xs.append(x)  
            ys.append(y)  
  
print(len(latitudes), 'candidate neighborhood centers generated.')
```

<ipython-input-35-88702e7652dd>:12: DeprecationWarning: This function is deprecated. See: <https://pyproj4.github.io/pyproj/stable/gotchas.html#upgrading-to-pyproj-2-from-pyproj-1>

```
xy = pyproj.transform(proj_latlon, proj_xy, lon, lat)
```

<ipython-input-35-88702e7652dd>:18: DeprecationWarning: This function is deprecated. See: <https://pyproj4.github.io/pyproj/stable/gotchas.html#upgrading-to-pyproj-2-from-pyproj-1>

```
lonlat = pyproj.transform(proj_xy, proj_latlon, x, y)
```

<ipython-input-35-88702e7652dd>:18: DeprecationWarning: This function is deprecated. See: <https://pyproj4.github.io/pyproj/stable/gotchas.html#upgrading-to-pyproj-2-from-pyproj-1>

```
lonlat = pyproj.transform(proj_xy, proj_latlon, x, y)
```

<ipython-input-35-88702e7652dd>:18: DeprecationWarning: This function is deprecated. See: <https://pyproj4.github.io/pyproj/stable/gotchas.html#upgrading-to-pyproj-2-from-pyproj-1>


```
<ipython-input-35-88702e7652dd>:18: DeprecationWarning: This function is
deprecated. See: https://pyproj4.github.io/pyproj/stable/gotchas.html#upgrading-
to-pyproj-2-from-pyproj-1
```

```
lonlat = pyproj.transform(proj_xy, proj_latlon, x, y)
```

```
<ipython-input-35-88702e7652dd>:18: DeprecationWarning: This function is
deprecated. See: https://pyproj4.github.io/pyproj/stable/gotchas.html#upgrading-
to-pyproj-2-from-pyproj-1
```

```
lonlat = pyproj.transform(proj_xy, proj_latlon, x, y)
```

```
[42]: map_aveiro = folium.Map(location=aveiro_center, zoom_start=13)
folium.Marker(aveiro_center, popup='Forum Aveiro').add_to(map_aveiro)
for lat, lon in zip(latitudes, longitudes):
    #folium.CircleMarker([lat, lon], radius=2, color='blue', fill=True,
    →fill_color='blue', fill_opacity=1).add_to(map_lisbon)
    folium.Circle([lat, lon], radius=150, color='blue', fill=False).
    →add_to(map_aveiro)
    #folium.Marker([lat, lon]).add_to(map_lisbon)
map_aveiro
```

```
[42]: <folium.folium.Map at 0x7fa6c21c6550>
```

```
[45]: # instantiate a feature group for the X, Y in the dataframe Theatres
casas = folium.map.FeatureGroup()

# loop through the theatres and add each to the theatres feature group
for lat, lng, in zip(X.Latitude, X.Longitude):
    casas.add_child(
        folium.features.CircleMarker(
            [lat, lng],
            radius=10, # circle markers
            color='red',
            fill=True,
            fill_color='yellow',
            fill_opacity=0.6
        )
    )

# add incidents to map
map_aveiro.add_child(casas)
```

```
[45]: <folium.folium.Map at 0x7fa6c21c6550>
```

```
[59]: # instantiate a feature group for the theatres in the dataframe
casas = folium.map.FeatureGroup()

# loop through the casas and add each to the casas feature group
for lat, lng, in zip(X.Latitude, X.Longitude):
    casas.add_child(
        folium.features.CircleMarker(
```

```

        [lat, lng],
        radius=10, # circle markers
        color='red',
        fill=True,
        fill_color='yellow',
        fill_opacity=0.6
    )
)

# add pop-up text to each marker on the map
latitudes = list(X.Latitude)
longitudes = list(X.Longitude)
labels = list(X.apartments)

for lat, lng, label in zip(latitudes, longitudes, labels):
    folium.Marker([lat, lng], popup=label).add_to(map_aveiro)

# add incidents to map
map_aveiro.add_child(casas)

```

[59]: <folium.folium.Map at 0x7fa6c21c6550>

[]: