What’s cooking ?

Sommaire

Kaggle

Présentation du projet

Problématique

Résolution

Outils

Démonstration

Bilan

Conclusion

Piste d’amélioration

What’s cooking ?

Kaggle

Kaggle

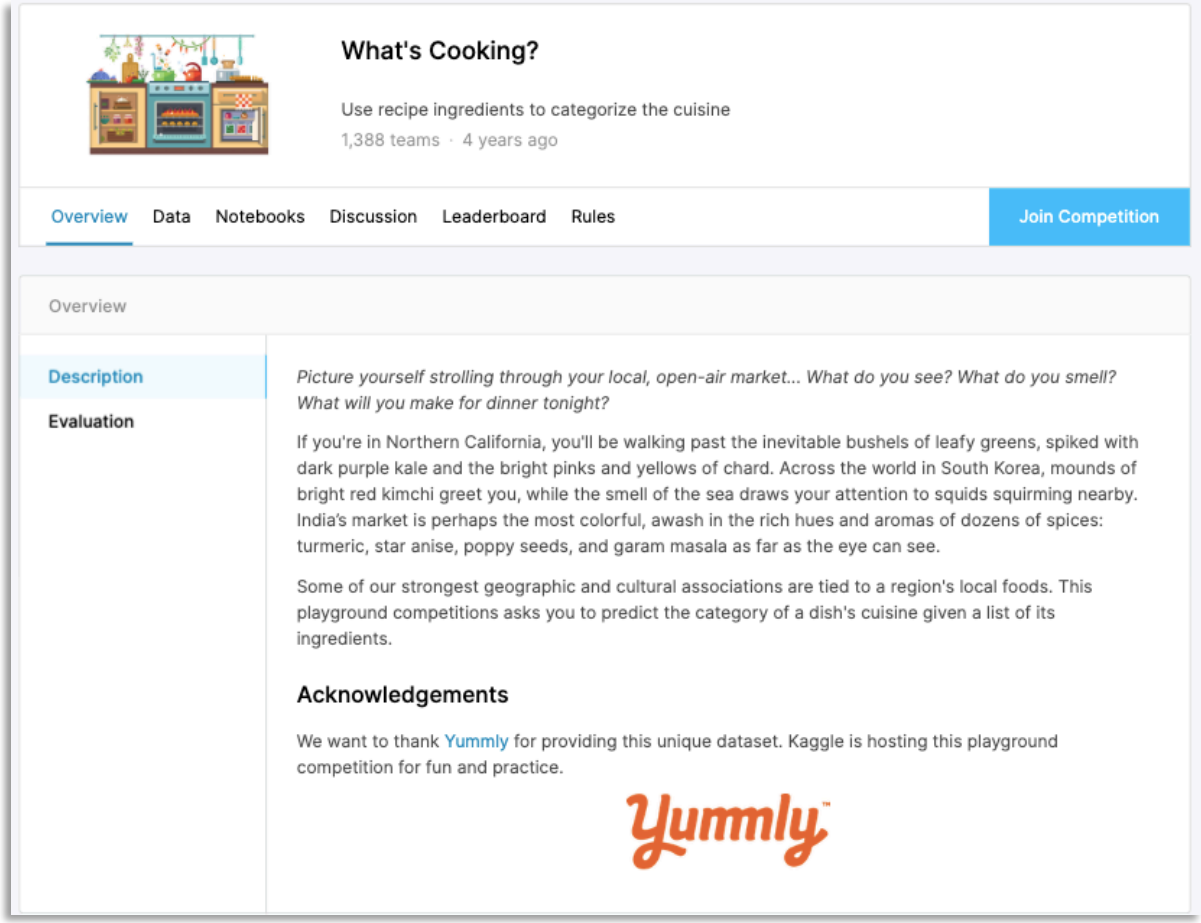
- Plateforme de compétition entre data-scientists
- Problèmes à résoudre avec du machine learning
- Mise à disposition:
  - Grand jeux de donnée
  - Espace de travaille (notebook)

What’s cooking ?

Présentation du sujet

Présentation du projet

Problématique




```
[
 {
  "id": 10259,
  "cuisine": "greek",
  "ingredients": [
   "romaine lettuce",
   "black olives",
   "grape tomatoes",
   "garlic",
   "pepper",
   "purple onion",
   "seasoning",
   "garbanzo beans",
   "feta cheese crumbles"
  ]
 },
 {
  "id": 25693,
  "cuisine": "southern_us",
  "ingredients": [
   "plain flour",
   "ground pepper",
   "salt",
   "tomatoes",
   "ground black pepper",
   "thyme",
   "eggs",
   "green tomatoes",
   "yellow corn meal",
   "milk",
   "vegetable oil"
  ]
 },
```

Présentation du projet

Résolution - Outils

1) Matrice

Librairie Pandas

	id	cuisine	ingredient	val
0	10259	greek	romaine lettuce	1
1	10259	greek	black olives	1
2	10259	greek	grape tomatoes	1
3	10259	greek	garlic	1
4	10259	greek	pepper	1

```
#### Création des DataFrames d'apprentissage 1/2 ####

util_dfs = pd.pivot_table(data = dfs, values = 'val', index = 'id', columns = 'ingredient')
util_dfs = util_dfs.fillna(0)
```

ingredient	(oz.) tomato sauce	(oz.) tomato paste	(10 oz.) frozen chopped spinach	(10 oz.) frozen chopped spinach, thawed and squeezed dry	(14 oz.) sweetened condensed milk	(14.5 oz.) diced tomatoes	(15 oz.) refried beans	1% low-fat buttermilk	1% low-fat chocolate milk	1% low-fat cottage cheese	...	yukon gold potatoes	yuzu	yuzu juice	za'atar	zest
id																
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
1000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0

5 rows × 6714 columns

Présentation du projet

Résolution - Outils

2) Factorisation

Algorithme NMF, librairie sklearn.decomposition

```
#### Création et apprentissage du modèle de reconnaissance des types de cuisine ####

X = util_dfs

# Création du modèle
model = NMF(n_components = 6, init='random', random_state=0)

# Apprentissage du modèle (décomposition en deux matrices à 6 dimensions)
W = model.fit_transform(X)
```

```
: W
array([[0.        , 0.        , 0.12876519, 0.00161134, 0.        ,
        0.        ],
       [0.11214663, 0.        , 0.01343381, 0.        , 0.07464309,
        0.02471402],
       [0.01455789, 0.        , 0.        , 0.07010848, 0.00127487,
        0.07695549],
       ...,
       [0.        , 0.03656691, 0.00041074, 0.        , 0.        ,
        0.        ],
       [0.00018086, 0.        , 0.06515019, 0.0090832 , 0.07218028,
        0.00614986],
       [0.00298885, 0.00145084, 0.01462306, 0.        , 0.00203041,
        0.0029329 ]])
```

3) Les plus proches voisins

Librairie sklearn.neighbors

```
#### Création d'un modèle des voisins ####

neigh = NearestNeighbors(n_neighbors=10)
neigh.fit(W)
```

Présentation du projet

Résolution - Outils

4) Réduction de dimensions

Algortihme TSNE - Librairie sklearn.manifold

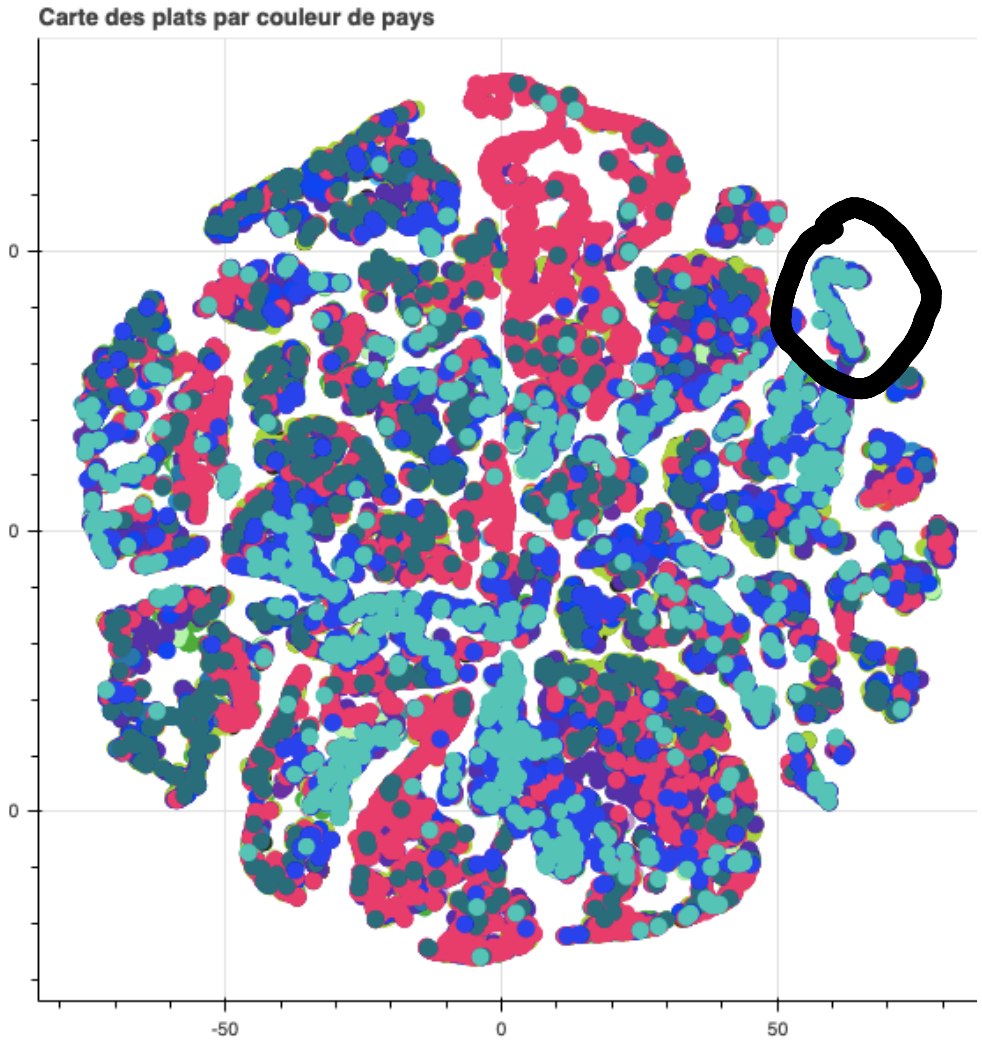
Librairie Bokeh

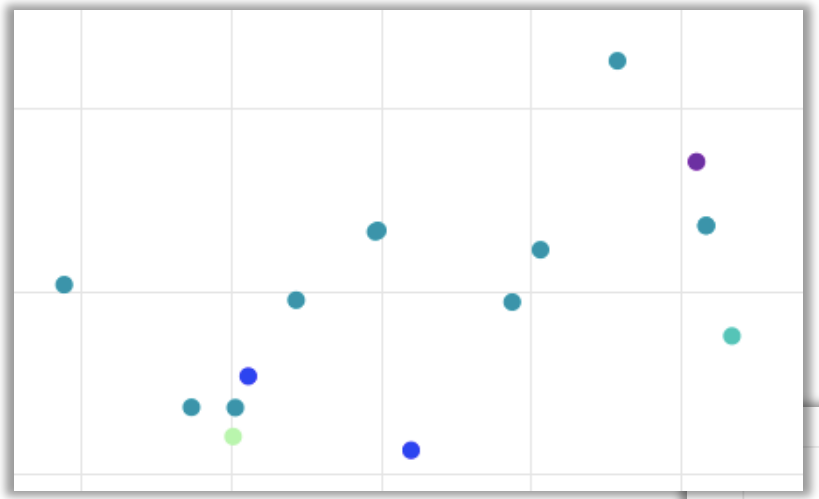
```
#### Calcul des points en 2D correspondants aux plats dans l'ordre de W ####

points_plats = tsne(n_components = 2).fit_transform(W)
```

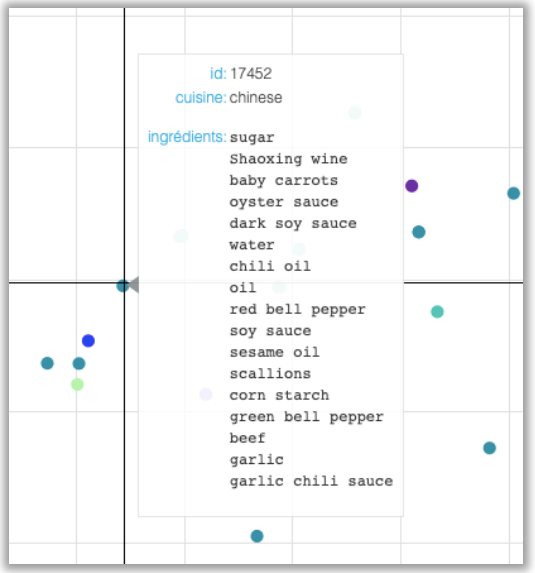
```
[[[-56.082382, 26.717314],
  [-16.43293, 54.312527],
  [-67.24952, -23.151327],
  [-56.076916, 26.720102],
  [-42.324726, -37.560867],
  [12.171578, 35.736145],
```

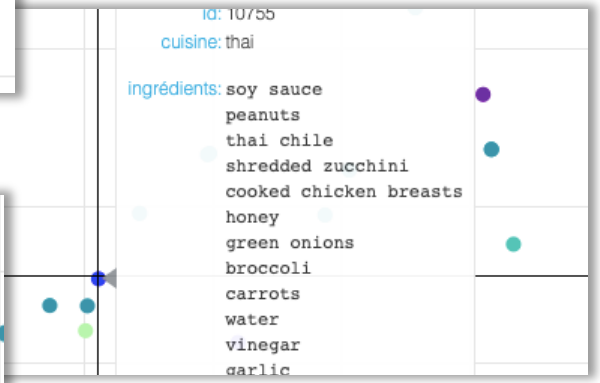
Présentation du projet

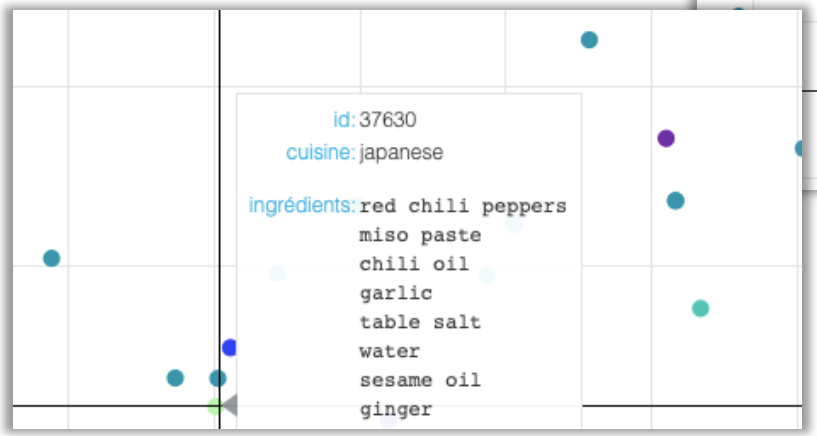
Résolution - Outils











What’s cooking ?

Démonsatration

What’s cooking ?

Bilan

Bilan

Conclusion

Familiarisation

Découverte

Bilan

Piste d’amélioration

Le nombre des plus proches voisins

Le nombre de dimension des vecteurs

Normalisation