# Avis Restau

Étude préliminaire des fonctionnalités pour "Détecter les sujets d'insatisfaction" et "Labelliser automatiquement les photos postées"

# Projet

Détecter les sujets d'insatisfaction présents dans les commentaires postés sur la plateforme.

Labelliser automatiquement les photos postées sur la plateforme. Par exemple, identifier les photos relatives à la nourriture, au décor dans le restaurant ou à l'extérieur du restaurant.

Étude préliminaire des fonctionnalités "Détecter les sujets d'insatisfaction"

S'assurer de la possibilité de collecter de nouvelles données

# Étapes

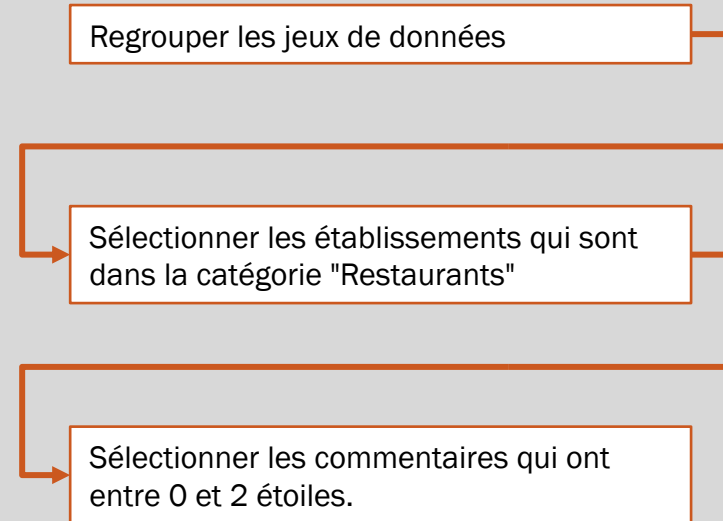| Analyser les commentaires pour détecter les différents sujets d'insatisfaction | Analyser les photos pour déterminer les catégories des photos | Collecter un échantillon (environ 200 restaurants) de données via l'API Yelp |
|---|---|---|
| ↓ | ↓ | ↓ |
| Prétraiter des données textuelles | Prétraiter des images | Récupérer uniquement les champs nécessaires |
| ↓ | ↓ | ↓ |
| Utiliser les techniques de réduction de dimension | Utiliser les techniques de réduction de dimension | Stocker les résultats dans un fichier exploitable (par exemple csv) |
| ↓ | ↓ | |
| Visualiser des données de grandes dimensions | Visualiser des données de grandes dimensions | |

# Description des données

### yelp_academic_dataset_business.json

```json
{
    "business_id":"Xb0VJtN8ldxBSrgmEinHIA",
    "name":"Taco Bell",
    "address":"26 E Schrock Rd",
    "city":"Westerville",
    "state":"OH",
    "postal_code":"43081",
    "latitude":40.1100433197,
    "longitude":-82.924741194,
    "stars":2.5,
    "review_count":24,
    "is_open":1,
    "attributes":{
            ...
    },
    "categories":"Breakfast & Brunch, Restaurants,
Tacos, Mexican, Fast Food",
    "hours":{
      ...
    }
}
```

### yelp_academic_dataset_review.json

```json
{
    "review_id":"m68FT2nDHXAyYZtBtE0n2Q",
    "user_id":"kTnqh6o9zPeiks3JVv-DyQ",
    "business_id":"Xb0VJtN8ldxBSrgmEinHIA",
    "stars":2.0,
    "useful":2,
    "funny":0,
    "cool":0,
    "text":"I will say Westerville needed Taco Bell back in town.  The
food is really good but I must say this is the slowest location I have
ever been to.  Takes approximately 10 minutes to get your food every
time and this is supposed to be fast food.  If it was my choice, I
would not come here but my boys like it.  They need to step up their
game.",
    "date":"2016-08-14 21:48:02"
}
```

# Prétraitement des données

Regrouper les jeux de données

Sélectionner les établissements qui sont dans la catégorie "Restaurants"

Sélectionner les commentaires qui ont entre 0 et 2 étoiles.

# Traitement des données texte

Very nice décor, with band playing jazz, but don't go with a group larger than 4 after 6:30PM

| Mise en minuscule du texte | Suppression des accents | Expansion des contractions |
|---|---|---|
| very nice décor, with band playing jazz, but don't go with a group larger than 4 after 6:30PM | very nice decor, with band playing jazz, but don't go with a group larger than 4 after 6:30PM | very nice decor, with band playing jazz, but do not go with a group larger than 4 after 6:30PM |

| Suppression de la ponctuation | Suppression des mots avec des chiffres | Stemmisation |
|---|---|---|
| very nice decor, with band playing jazz but do not go with a group larger than 4 after 6:30PM | very nice decor with band playing jazz but do not go with a group larger than after | not open at very nice decor with band play jazz but do not go with a group large than after |

| Stop words | very nice decor band play jazz not group large |
|---|---|
| very nice decor band play jazz not group large | |

# Répartition des mots

# Latent Dirichlet Allocation (LDA)

11 catégories réparties en 5 groupes

### Service

- Call, manag,custom,would,told
- Veri, restaur, dish, servic, lobster

### Nourriture

- Taco, steak, food, chip, good
- Food, chicken, tast, like, ric
- Sushi, coffe, roll, ice, tuna

### Commande

- pizza, salad order, good like
- Burger, fri , order, chees, sandwich

### Emplacement

- Order, go, food, place, time
- Food, place, like, good, one

### Attente

- Order, food, us, minut, wait
- Tabl, us, wait, bar, seat

11 catégories avec un score de cohérence de 0,40

# Latent Dirichlet Allocation (LDA)

==Ordered== take out that was to be 30 mins. I hour later we finally received our order. Pirate Pak burger dry, cold and unfit for consumption. Fish and chips also cold but did eat the fish which was at least fresh. Milk shake melted. I suspect that the ==order== was ready at the 30 ==minute== mark but somehow they either had a mix up or put the ==order== somewhere where it was missed. Will give them the benefit of the doubt on temperature. No excuses on the Pirate Pak. If you must use this White Spot, check your take out ==order== before you leave the restaurant for quality and temperature. I will not return.

## Attente

- Order, food, us, minut, wait

# K-means

**Service**

- food, place, good, servic, like
- bar, bartend, drink, beer, servic

**Nourriture**

- bbq, rib, brisket, pork, sauc

**Commande**

- burger, fri, order, food, good
- sandwich, order, bread, chees, good
- chicken, rice, order, fri, food
- taco, order, food, bell, mexican

**Emplacement**

- coffe, cup, place, breakfast, shop
- pizza, order, crust, chees, place

**Attente**

- us, tabl, food, wait, minut
- order, wait, call, time, minu

11 catégories avec un score de silhouette de 0,0074

# K-means

Ordered take out that was to be 30 mins. I hour later we finally received our order. Pirate Pak burger dry, cold and unfit for consumption. Fish and chips also cold but did eat the fish which was at least fresh. Milk shake melted. I suspect that the order was ready at the 30 minute mark but somehow they either had a mix up or put the order somewhere where it was missed. Will give them the benefit of the doubt on temperature. No excuses on the Pirate Pak. If you must use this White Spot, check your take out order before you leave the restaurant for quality and temperature. I will not return.

## Attente

- order, wait, call, time, minu

# Non Negative Matrix Factorization (NMF)

Service

- food, servic, good, terribl, horribl

Nourriture

- chicken, tast, veri, sauc, good,
- pizza, crust, chees, slice, top
- burger, fri, bun, onion, chees

8 catégories avec un score de silhouette de 0,0042

Emplacement

- place, go, time, like, get
- us, tabl, ask, drink, server

Attente

- wait, minut, hour, min, long
- order, call, deliveri, time, said

# Non Negative Matrix Factorization (NMF)

Ordered take out that was to be 30 mins. I hour later we finally received our order. Pirate Pak burger dry, cold and unfit for consumption. Fish and chips also cold but did eat the fish which was at least fresh. Milk shake melted. I suspect that the order was ready at the 30 minute mark but somehow they either had a mix up or put the order somewhere where it was missed. Will give them the benefit of the doubt on temperature. No excuses on the Pirate Pak. If you must use this White Spot, check your take out order before you leave the restaurant for quality and temperature. I will not return.

**Attente**

- order, call, deliveri, time, said

# API Yelp

API GraphQL

## Requête

```
{
  search(
    term: "restaurants",
    location: "$location",
    categories:"restaurants",
    limit:$limit,
    offset:$offset
  ){
    total,
    business {
      name
      id
      reviews(){
        text
        rating
      }
    }
  }
}
```

Limité à un extrait de 3 commentaires par établissement

## Format JSON

```
{
  "name":"Fuckoffee",
  "id":"-22PaX8rQyh0ZQrdCiDz4A",
  "reviews":[
    {
      "text":"They make perfect flat white, I love
      the fruity taste. Nice and comfortable place,
      friendly people.\nI miss the crowd in the shop
      before this COVID-19...",
      "rating":5
    },
    {
      "text":"Lol more like fuck racist
      establishments. This business has an a-frame
      sign that they're proudly boasting, which
      contains a very explicit racist message...",
      "rating":1
    },
    {
      "text":"Do not support racist businesses. I
      don't care how good your coffee is, your
      ignorance has no place here.",
      "rating":1
    }
  ]
},
```

## Format CSV

"They make perfect flat white, I love the fruity taste. Nice and comfortable place, friendly people.\nI miss the crowd in the shop before this COVID-19...", 5

"Lol more like fuck racist establishments. This business has an a-frame sign that they're proudly boasting, which contains a very explicit racist message...",1

"Do not support racist businesses. I don't care how good your coffee is, your ignorance has no place here.",1

Les données sont prêtes à être traitées avec le traitement standard

# Traitement des images

Objectif : **Labelliser automatiquement les photos postées sur la plateforme.**

**200 000 images**

**5 catégories**
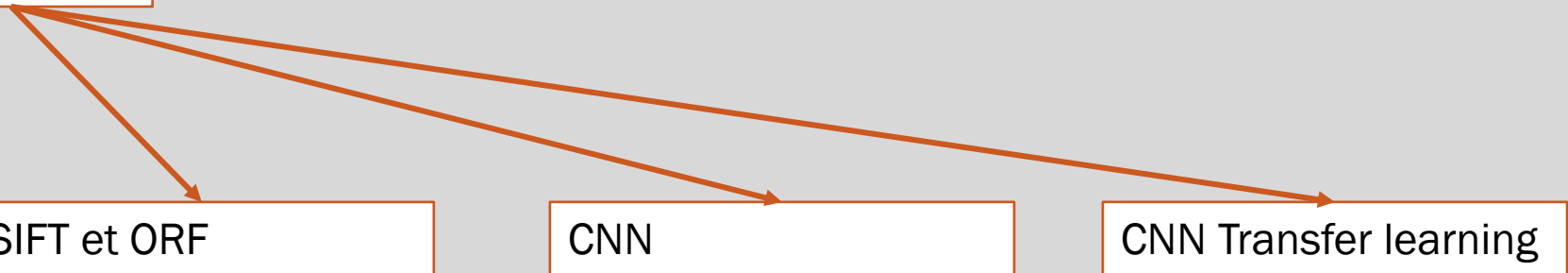- Interior
- Food
- Outside
- Menu
- drink

```
{
    "photo_id":"Un_Og6jfhazVn7CxszkKEw",
    "business_id":"R1sIqGfKpF5f3HV3vyNsbg",
    "caption":"Drinks! Front is beer obviously. The pink one is the blackberry lemonade and the yellow one is
the spicy margarita they made",
    "label":"drink"
}
```

**Algorithmes**

SIFT et ORF

CNN

CNN Transfer learning

# SIFT et ORB

Traitement de l'image
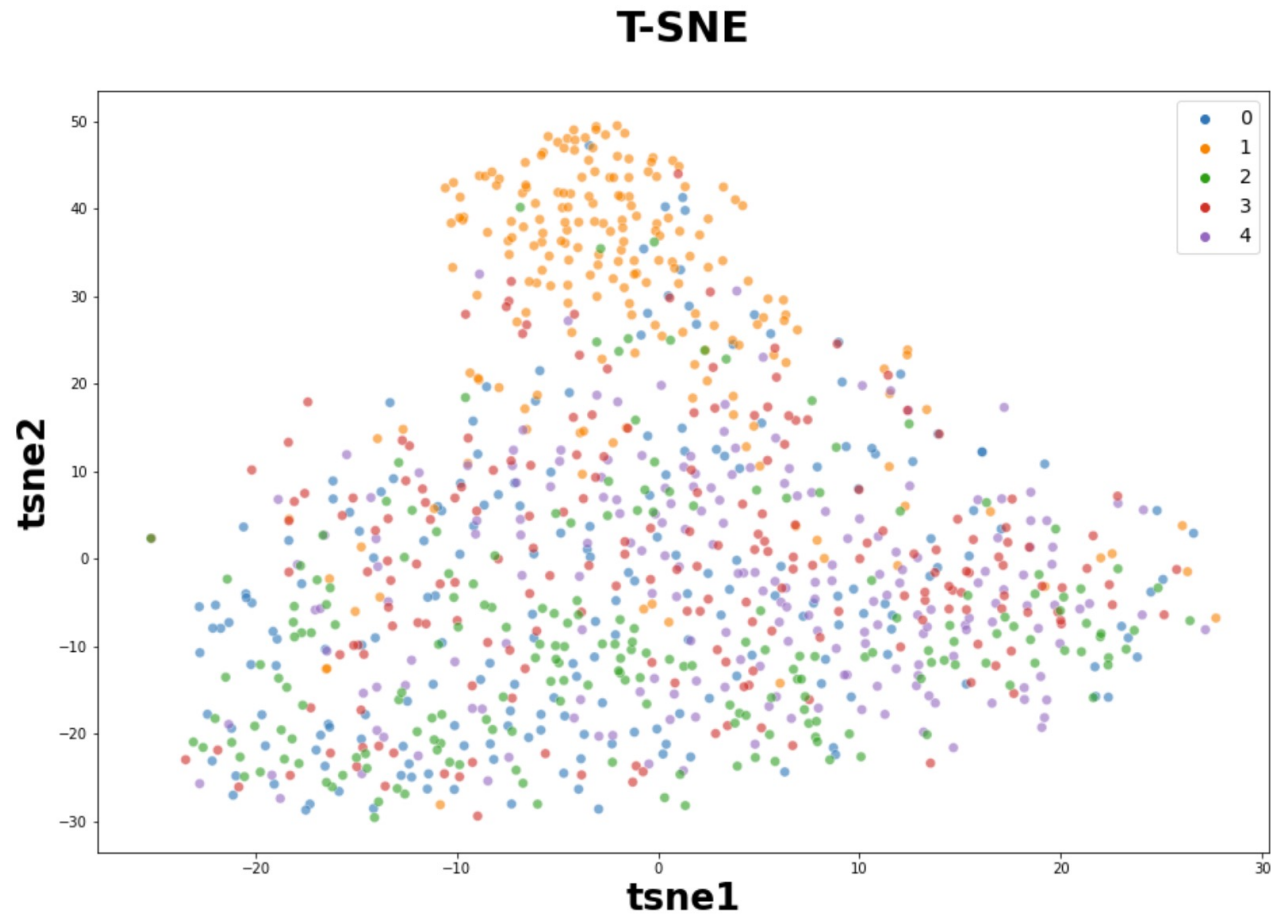
Original

Nuances de gris
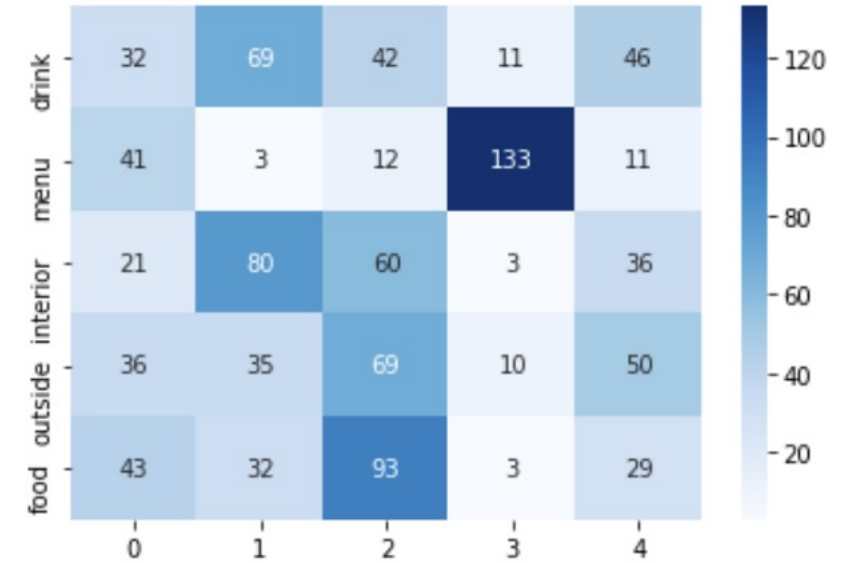
Mise à l'échelle
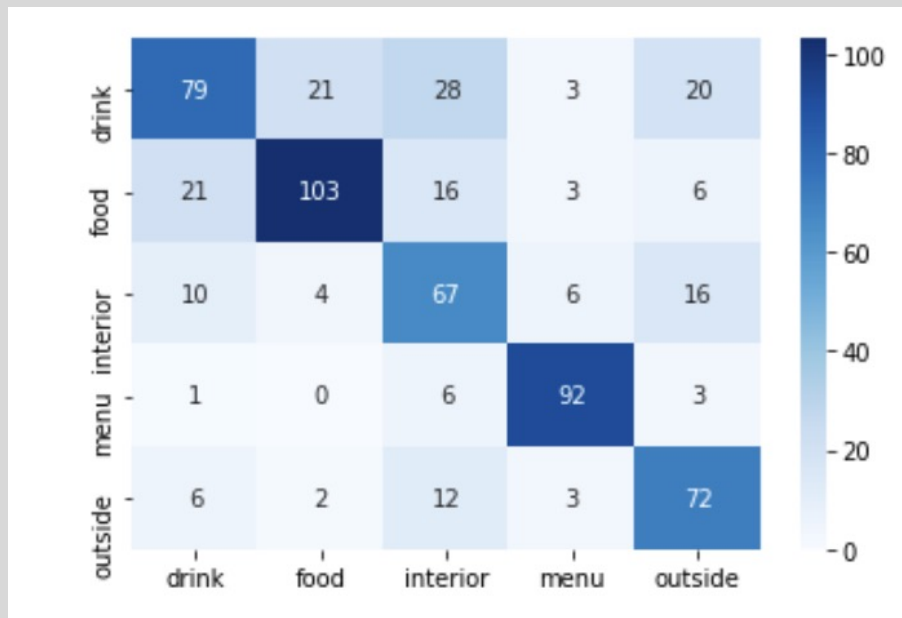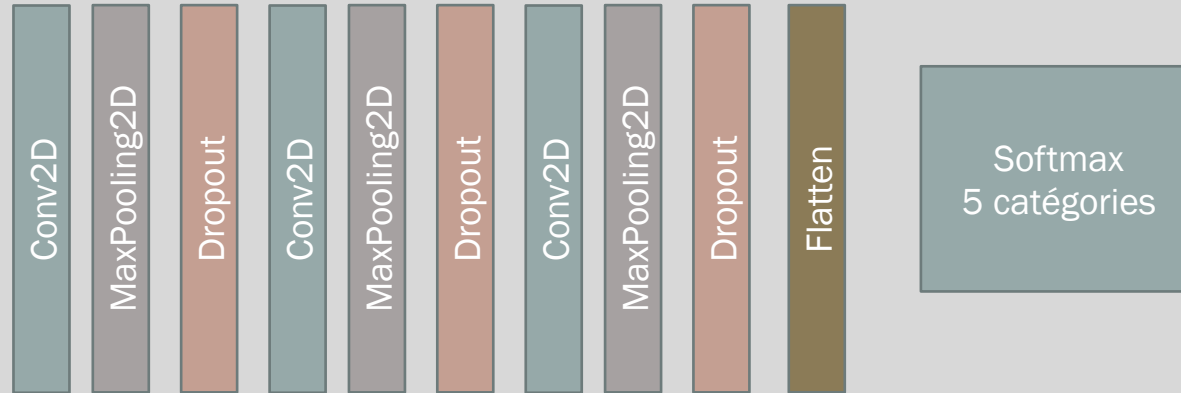
Histogramme

# SIFT

Réduction de dimension PCA, puis T-SNE

# Réseau de neurones convolutifs (CNN)

Précision de 65%




Intérieur détecté

# Apprentissage par transfert (VGG16)

Précision de 86%

Conv2D | Conv2D | MaxPooling2D | Conv2D | Conv2D | MaxPooling2D | Conv2D | Conv2D | Conv2D | MaxPooling2D | Conv2D | Conv2D | Conv2D | MaxPooling2D | Conv2D | Conv2D | Conv2D | MaxPooling2D | Flatten | Dense

Softmax
5 catégories

VGG16 standard -1 couche

Ajout





Intérieur détecté

# Visualisation des données
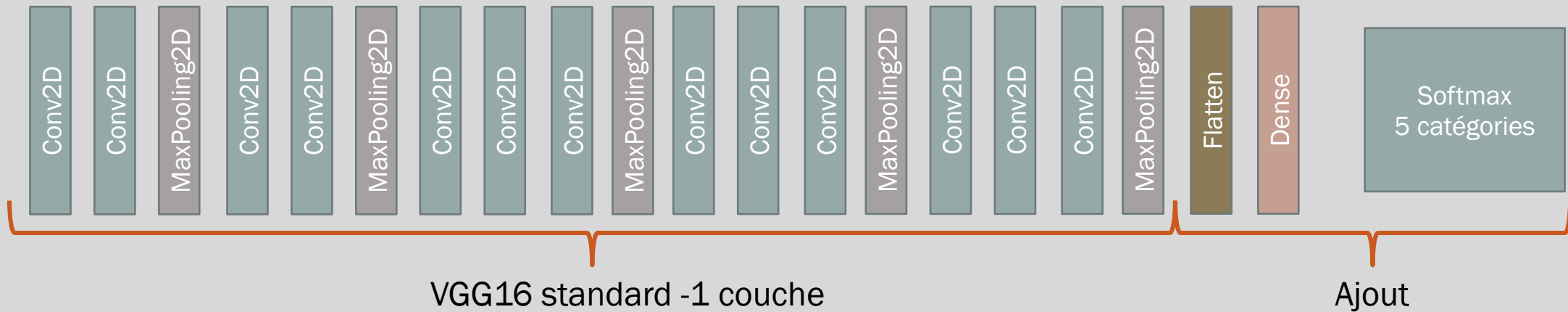
# Conclusion (texte)

**Faisabilité** de détection des sujets d'insatisfaction en utilisant  LDA, mais avec des aménagements
- Détection de la langue

**Avantage**
- Rapidité d'exécution des prédictions

**Inconvénients**
- Difficulté d'estimer les catégories

# Conclusion (images)

Faisabilité de classification d'images avec un modèle de Deep Learning CNN avec Transfer Learning

Avantage
- Bonnes prédictions à 86%
- Rapidité d'exécution des prédictions

Inconvénients
- Nécessite plus de ressources machine
- Utilise des images labellisées pour l'entrainement