

Contexte du projet

Préparation textes

- tous les avis de 1 étoile (1M)
- sélection échantillon (50 000)

Preprocessing

- lettres minuscules
- contractions de langage
- tokenisation
- adjectifs / adverbes
- stop words
- lemmatisation

Preprocessing

- création n-grams
- filtrer mots peu ou trop présents

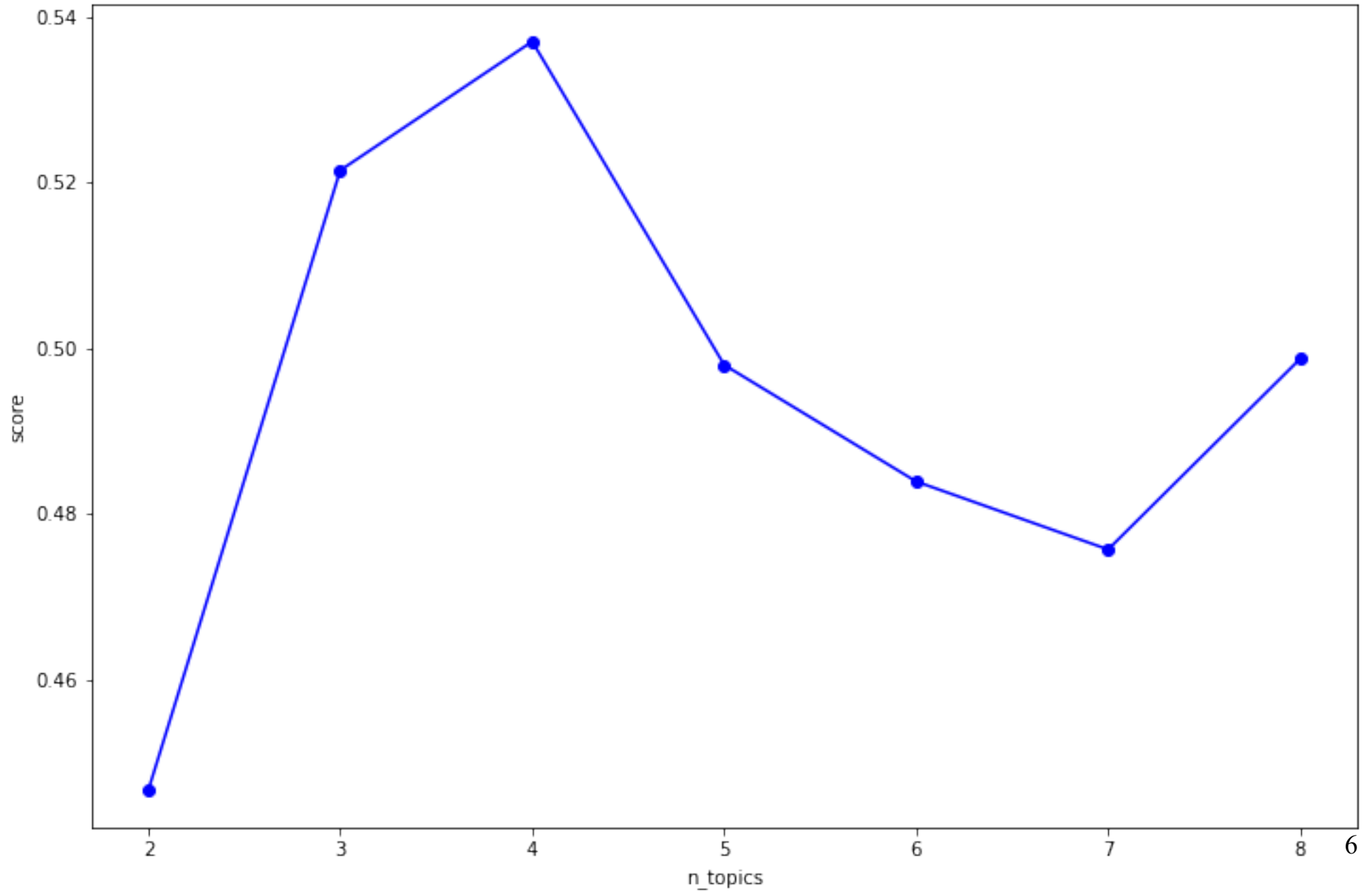
Modélisation

Modèles :

- LDA
- NMF

Score : Cohérence CV

Modélisation (LDA)



Modélisation (LDA)

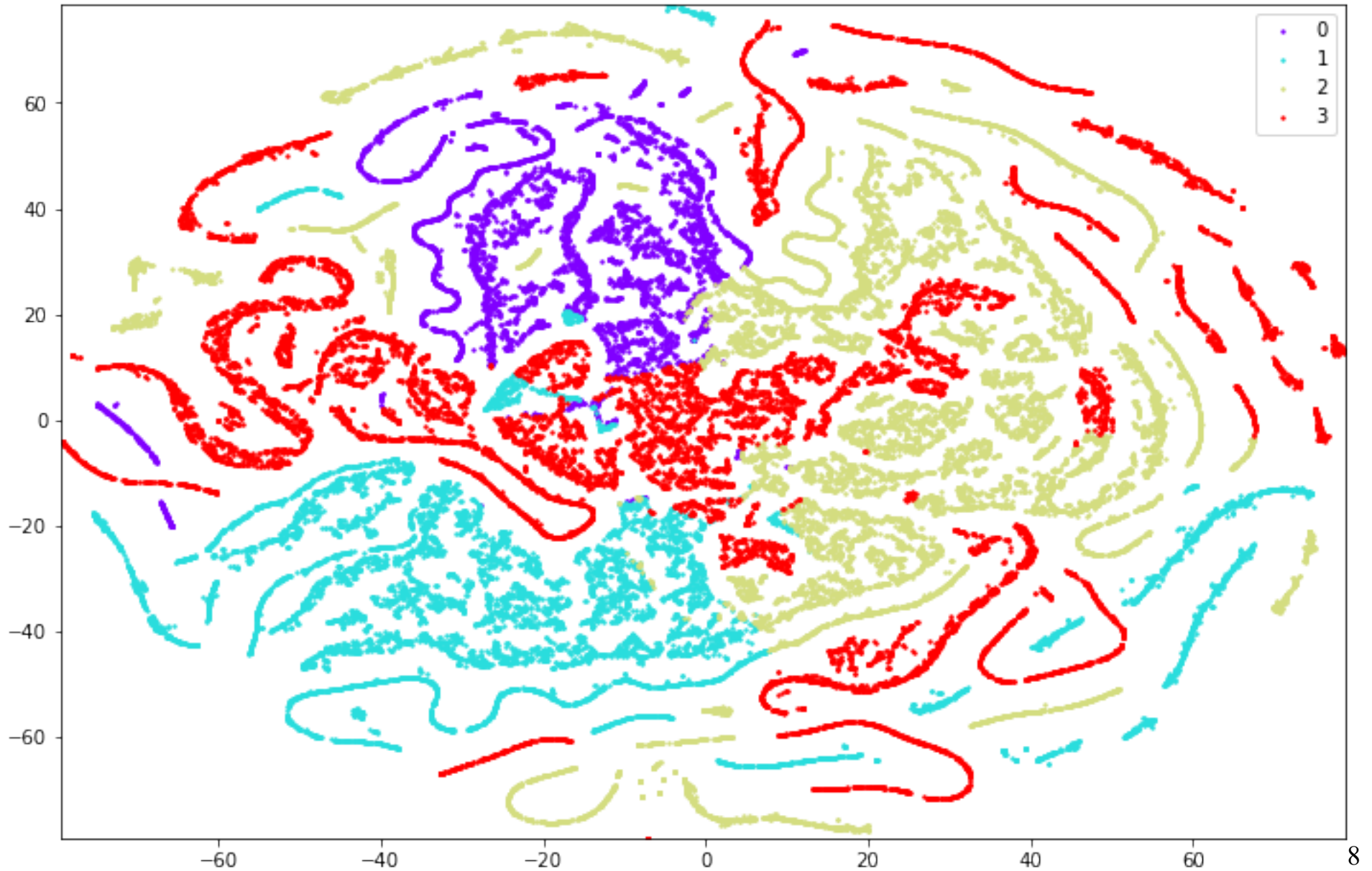
A word cloud representing a topic related to hotels and rooms. The most prominent words are 'hotel' and 'room'. Other visible words include 'stay', 'night', 'people', 'door', 'bathroom', 'car', 'say', 'parking', 'day', 'way', 'told', 'staff', 'desk', 'lot', and 'front'.

A word cloud representing a topic related to food and restaurants. The most prominent words are 'chicken', 'pizza', 'restaurant', and 'ordered'. Other visible words include 'taste', 'cheese', 'meat', 'menu', 'eat', 'meal', 'sauce', 'tasted', 'fry', 'got', 'order', 'flavor', 'came', 'sandwich', 'salad', and 'way'.

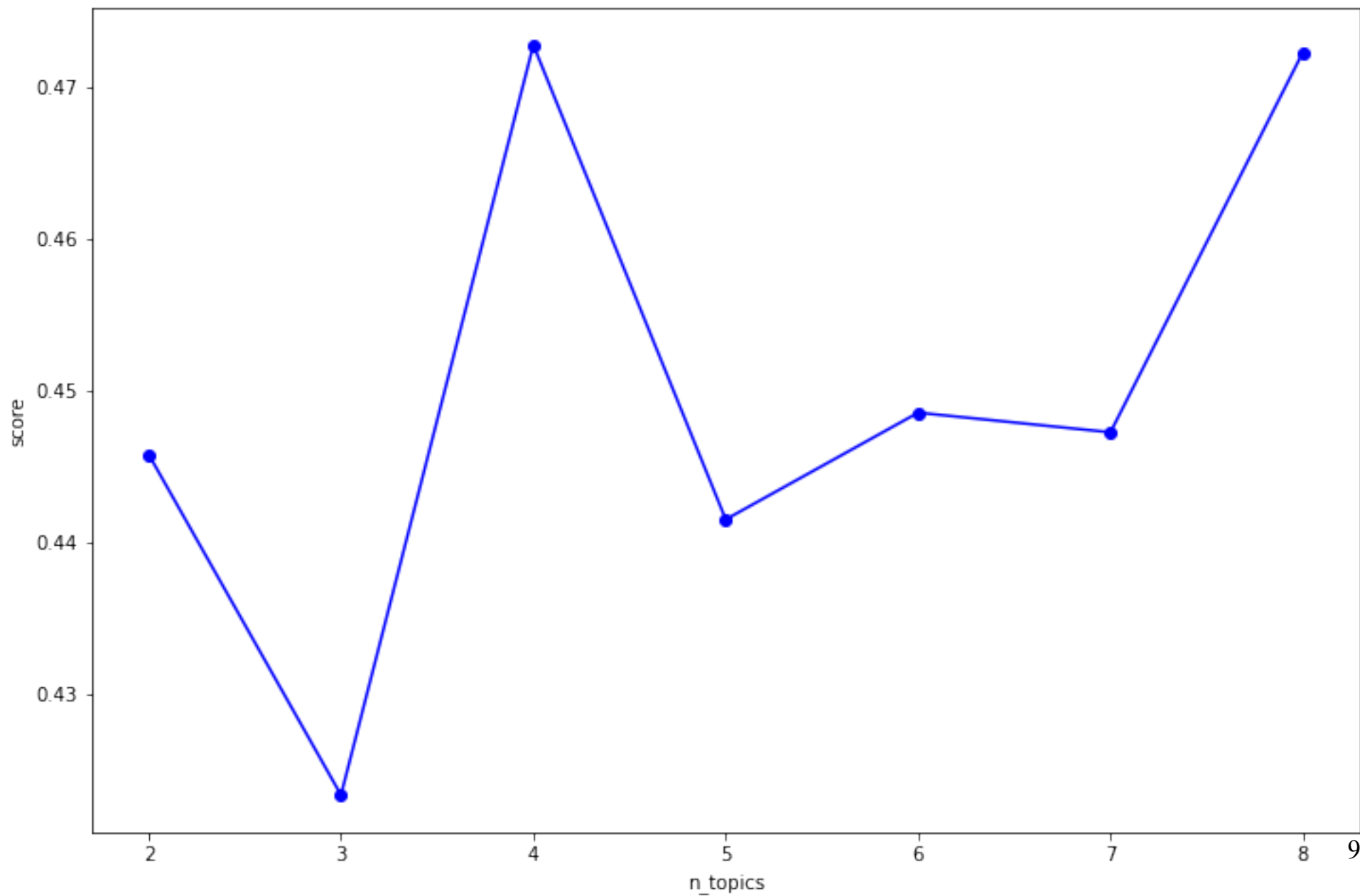
A word cloud representing a topic related to ordering and service. The most prominent words are 'minute', 'order', and 'said'. Other visible words include 'people', 'drink', 'wait', 'bar', 'manager', 'told', 'restaurant', 'server', 'went', 'table', 'came', 'customer', 'asked', and 'hour'.

A word cloud representing a topic related to communication and business. The most prominent words are 'told', 'day', 'said', and 'call'. Other visible words include 'called', 'went', 'take', 'company', 'work', 'asked', 'got', 'customer', 'store', 'business', 'car', 'month', 'appointment', 'money', 'week', and 'phone'.

Modélisation (LDA)



Modélisation (NMF)



Modélisation (NMF)

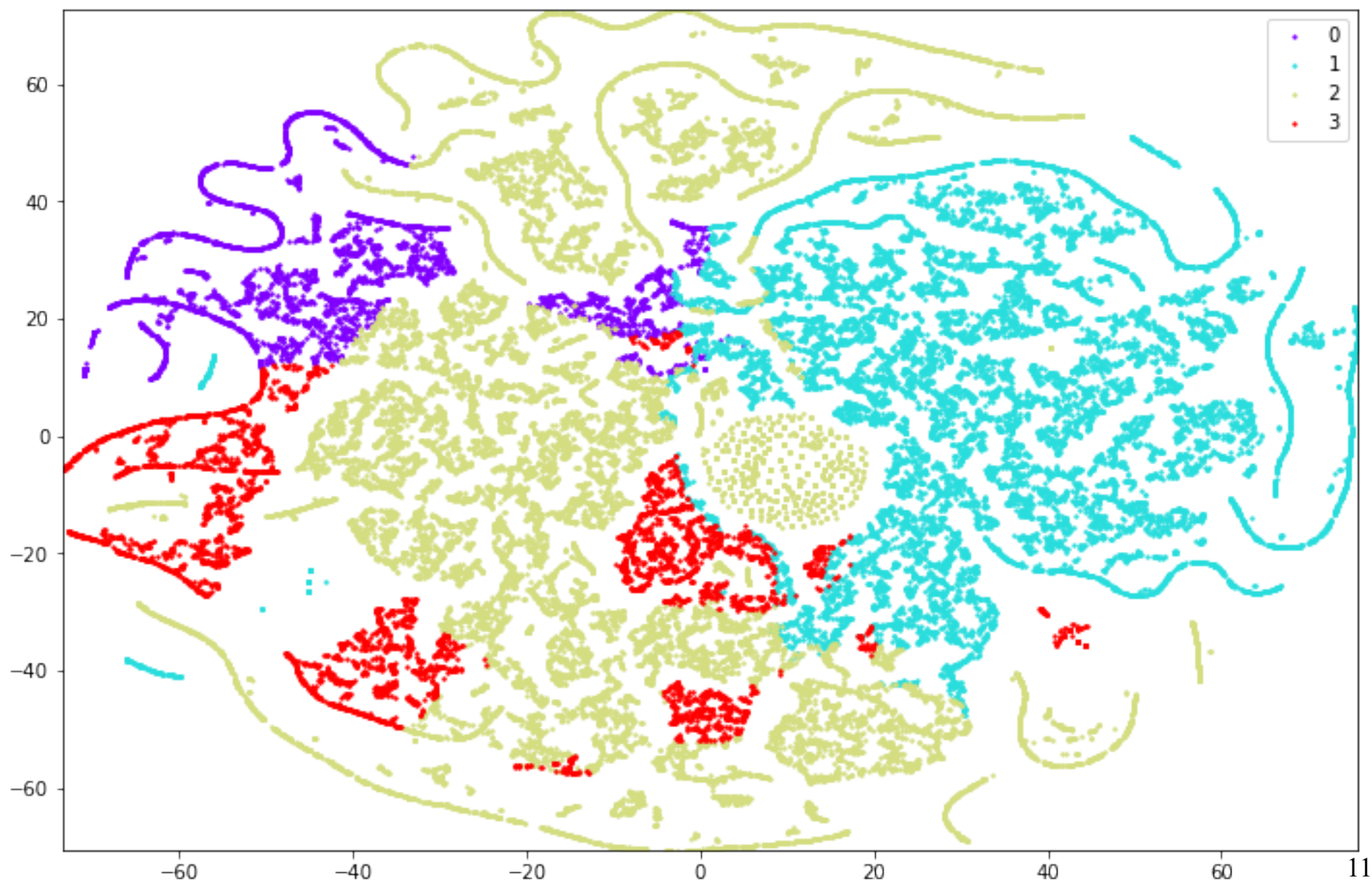
minute said told table asked came got waitress left server manager went wished friend drink hour another people

car said told called business store went week got company manager take work day know asked going customer phone

order ordered minute restaurant bar table people drink experience customer another came got two manager meal server chicken took

room hotel stay desk night front people day floor dog check staff door work year bed know thing going hour manager

Modélisation (NMF)



Préparation Images

Dataset :

- 5 catégories
- 200 images par catégorie

Modélisation

Méthode :

- preprocessing
- features (modèle)
- réduction (PCA)
- prédiction (K_means)
- visualisation (tsne)

Modélisation



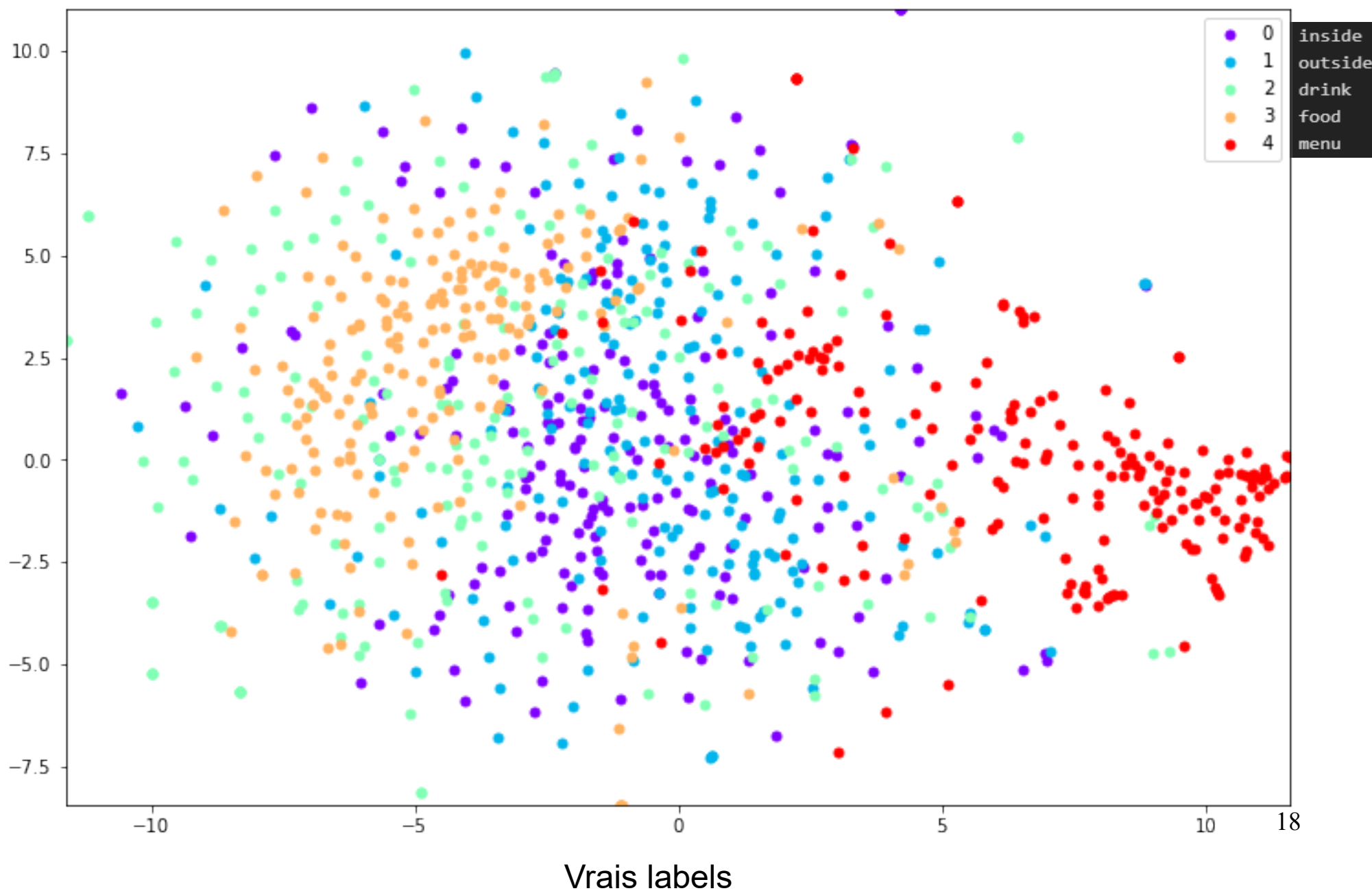
Modélisation (SIFT)



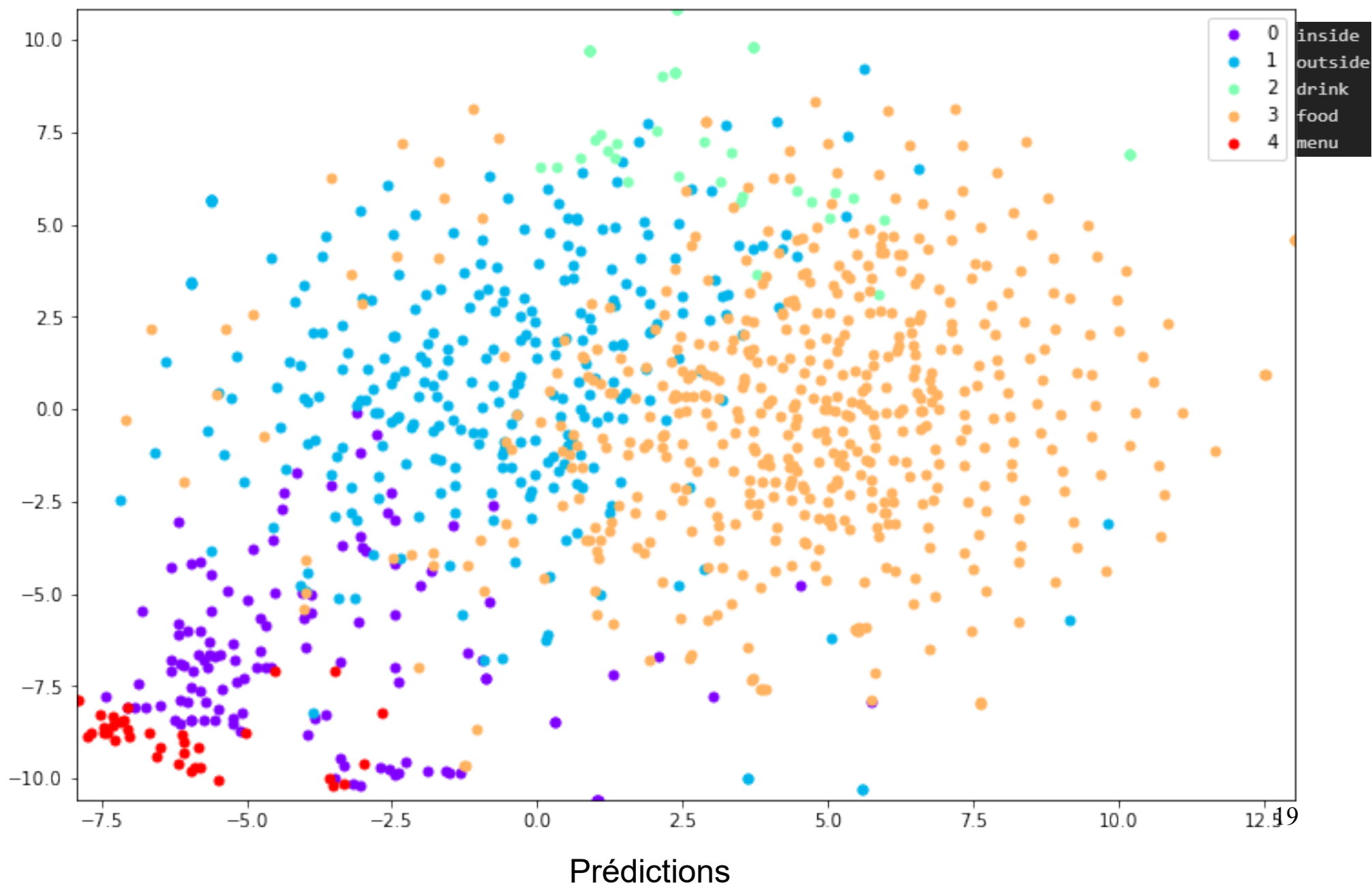
Modélisation (SIFT)

- 1184 features par image
- 723 features après PCA
- prédiction K-means : ARI = 0,09

Visualisation (SIFT)



Visualisation (SIFT)

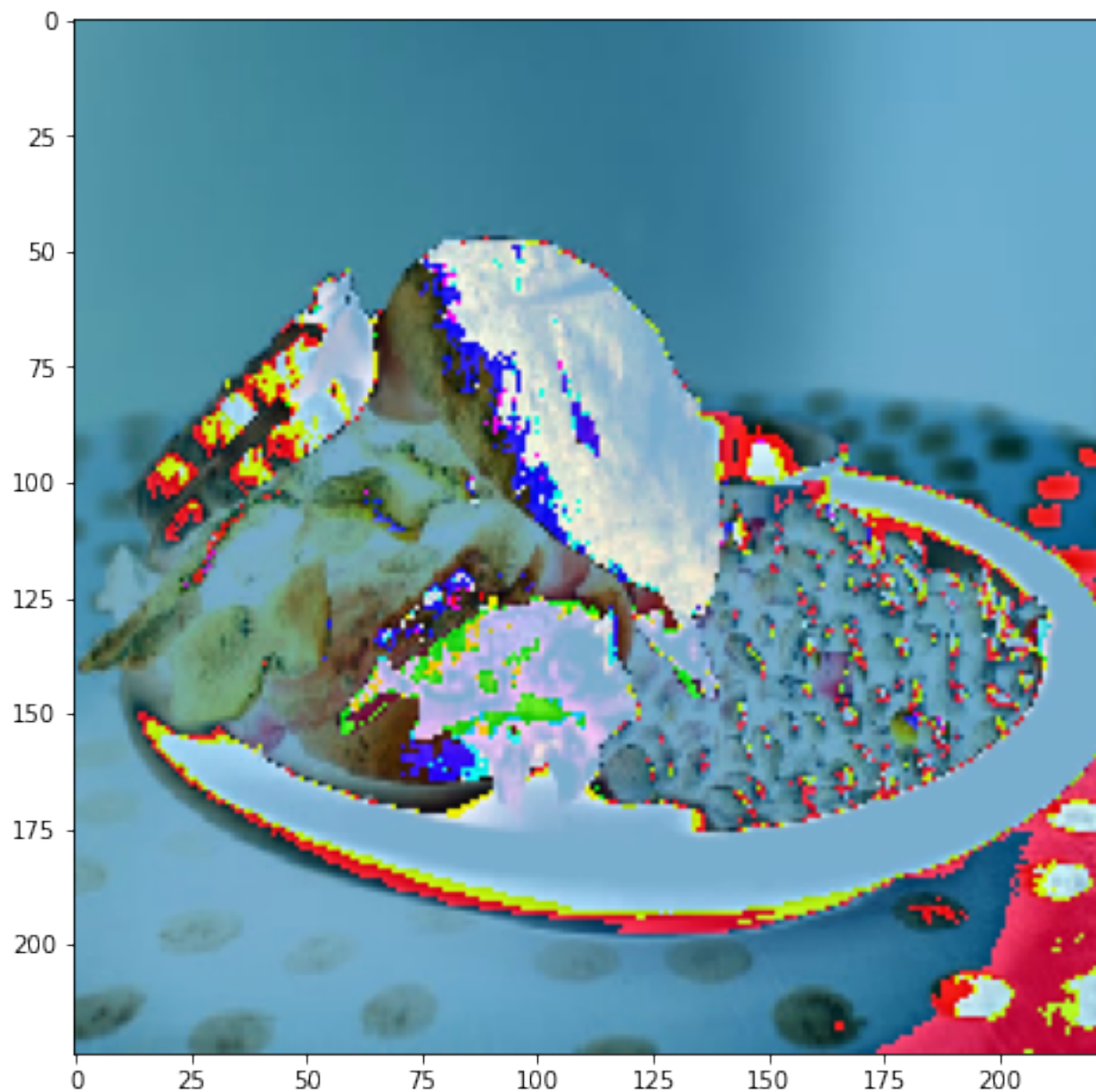


Modélisation (VGG16)

Model: "vgg16"

Layer (type)	Output Shape	Param #
input_2 (InputLayer)	[(None, None, None, 3)]	0
block1_conv1 (Conv2D)	(None, None, None, 64)	1792
block1_conv2 (Conv2D)	(None, None, None, 64)	36928
block1_pool (MaxPooling2D)	(None, None, None, 64)	0
block2_conv1 (Conv2D)	(None, None, None, 128)	73856
block2_conv2 (Conv2D)	(None, None, None, 128)	147584
block2_pool (MaxPooling2D)	(None, None, None, 128)	0
block3_conv1 (Conv2D)	(None, None, None, 256)	295168
block3_conv2 (Conv2D)	(None, None, None, 256)	590080
block3_conv3 (Conv2D)	(None, None, None, 256)	590080
block3_pool (MaxPooling2D)	(None, None, None, 256)	0
block4_conv1 (Conv2D)	(None, None, None, 512)	1180160
block4_conv2 (Conv2D)	(None, None, None, 512)	2359808
block4_conv3 (Conv2D)	(None, None, None, 512)	2359808
block4_pool (MaxPooling2D)	(None, None, None, 512)	0
block5_conv1 (Conv2D)	(None, None, None, 512)	2359808
block5_conv2 (Conv2D)	(None, None, None, 512)	2359808
block5_conv3 (Conv2D)	(None, None, None, 512)	2359808
block5_pool (MaxPooling2D)	(None, None, None, 512)	0
global_average_pooling2d_1 (GlobalAveragePooling2D)	(None, 512)	0

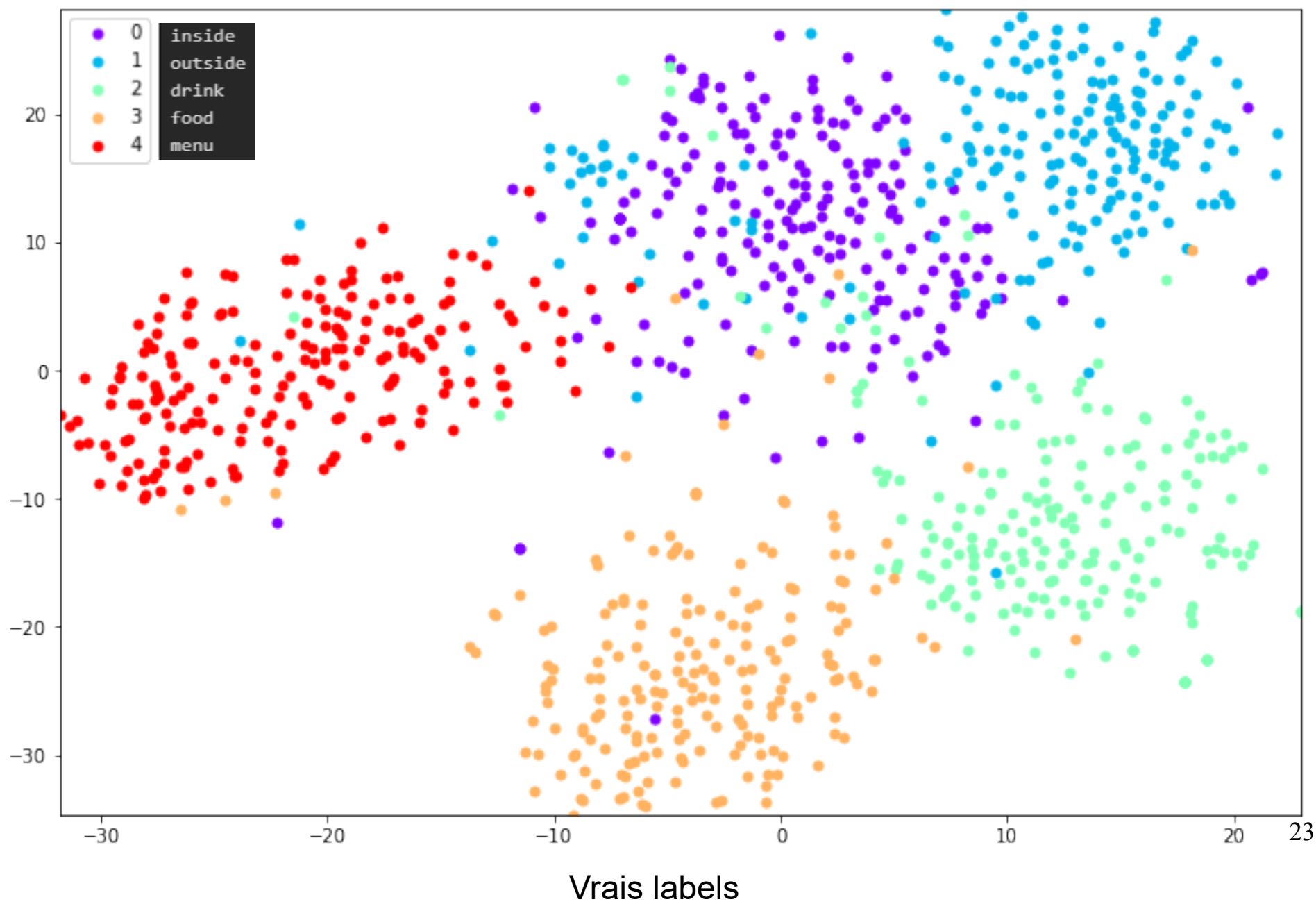
Modélisation (VGG16)



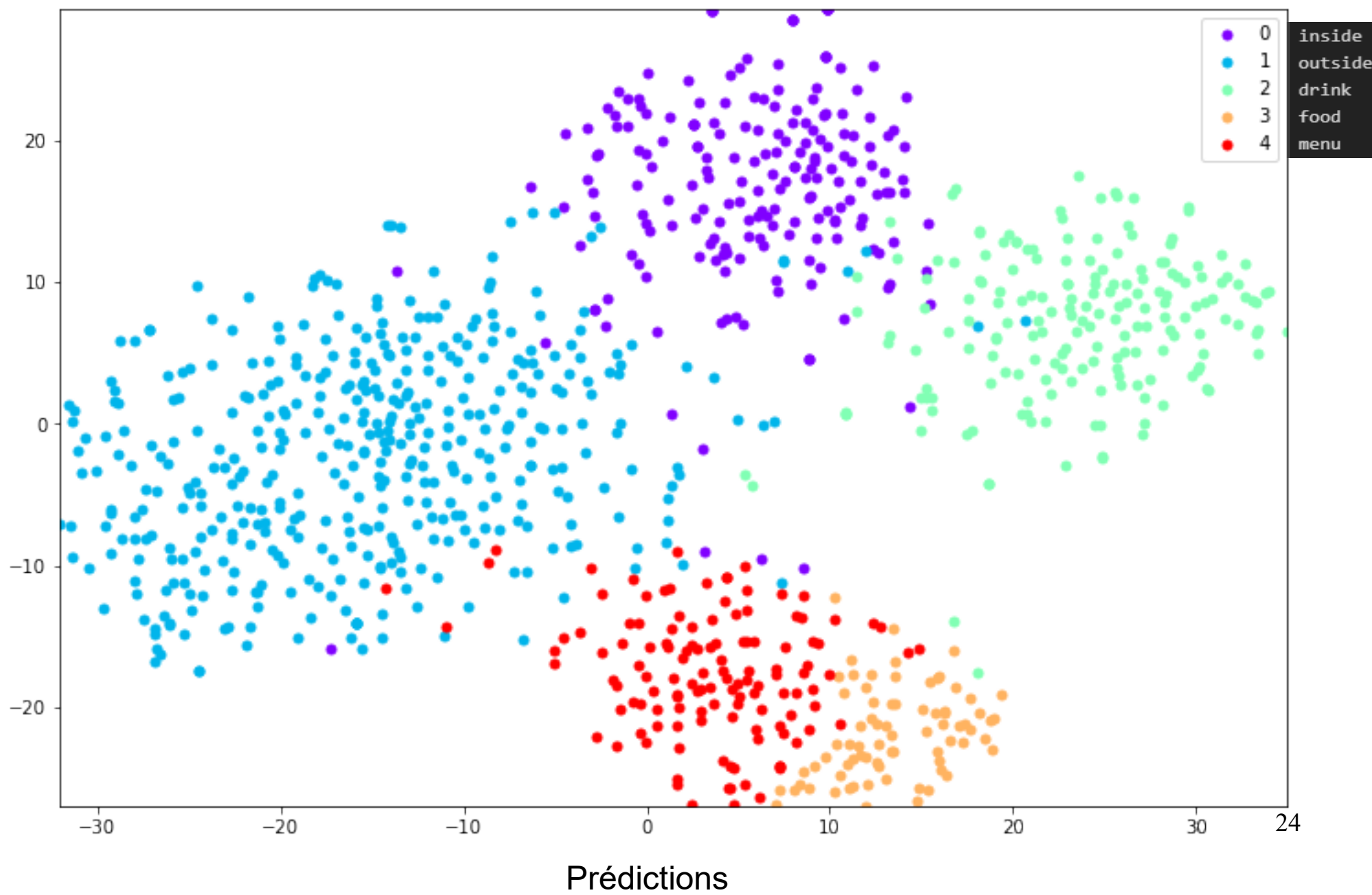
Modélisation (VGG16)

- 512 features par image
- 362 features après PCA
- prédiction K-means : $ARI = 0,57$

Visualisation (VGG16)



Visualisation (VGG16)



Conclusion