

OpenClassrooms

Projet P6:

Classifiez automatiquement des biens de consommation

N. Alves

Classifiez automatiquement des biens de consommation

Sommaire

- Contexte
- Présentation des informations disponibles
- Étude du texte
- Étude des images
- Conclusion et perspectives

Classifiez automatiquement des biens de consommation

- Contexte
- Présentation des informations disponibles
- Étude du texte
- Étude des images
- Conclusion et perspectives

Classifiez automatiquement des biens de consommation

Contexte

Sur la place de marché, des vendeurs proposent des articles à des acheteurs en postant une photo et une description. Lors de la mise en ligne des produits, il est attribué manuellement une catégorie au produit.

⇒ **On souhaite automatiser la classification de produits, pour cela une étude de faisabilité est réalisée.**



place de marché

Classifiez automatiquement des biens de consommation

- Contexte
- **Présentation des informations disponibles**
- Étude du texte
- Étude des images
- Conclusion et perspectives

Classifiez automatiquement des biens de consommation

Présentation des informations disponibles

Les informations disponibles sont :

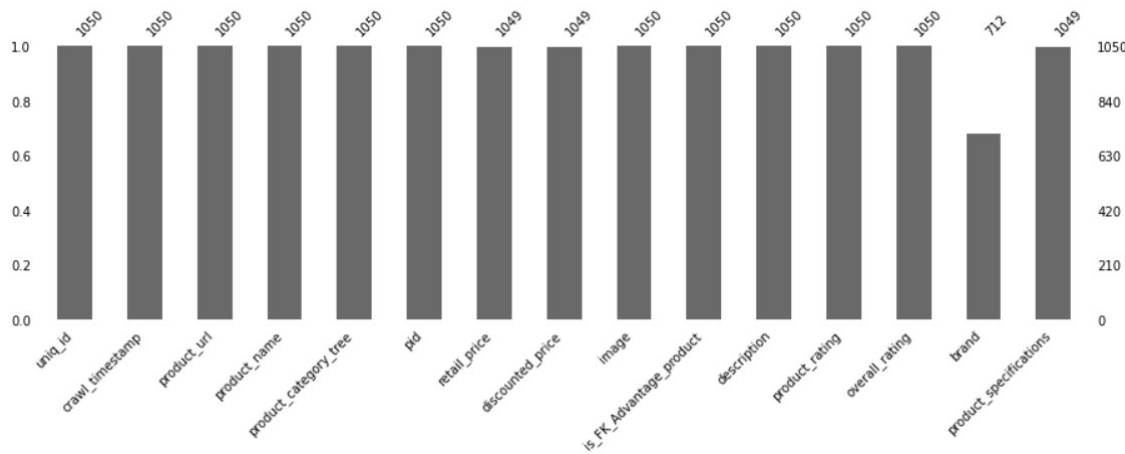
- un fichier csv,
- un répertoire contenant les photos des produits.

Ligne du fichier → un article

Colonne → caractéristiques de l'article

- descriptif,
- catégorie,
- prix,
- adresse internet etc.

⇒ **1050** photos et descriptions de produits.



Les informations disponibles, valeurs manquantes

Classifiez automatiquement des biens de consommation

- Contexte
- Présentation des informations disponibles
- Étude du texte
- Étude des images
- Conclusion et perspectives

Classifiez automatiquement des biens de consommation

Étude du texte

Le jeu de donnée :

Il y a quatre niveaux de précision dans la classification:

⇒ **Le premier niveau de précision comporte 7 catégories.**

Niveau 0: 7 catégories

Niveau 1: 62 catégories

Niveau 2: 239 catégories

Niveau 3: 348 catégories

➤ Les catégories principales du premier niveau:

- 'Home furnishing',
- 'Baby care',
- 'Watches',
- 'Home decor & festive needs',
- 'Kitchen & dining',
- 'Beauty and personal care',
- 'Computers'

Classifiez automatiquement des biens de consommation

Étude du texte

Modélisation non supervisée

Partie exploratoire : on utilise la modélisation Latent Dirichlet Allocation (LDA)

On cherche à définir 20 groupes (topics).

⇒ Chaque catégorie modélisée peut être reliée à une catégorie déjà définie.

⇒ **On retrouve l'ensemble des catégories déjà définies au préalable.**

Exemple : Topic 5 → 'Kitchen & dining'

Topic 5:

mug ceramic coffee gift perfect make design love one give

Classifiez automatiquement des biens de consommation

- Contexte
- Présentation des informations disponibles
- Étude du texte
- Étude des images
- Conclusion et perspectives

Classifiez automatiquement des biens de consommation

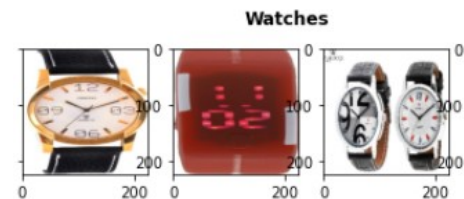
Étude des images

Le jeu de donnée :

- 7 catégories principales contenant chacune 150 photos.
- 62 sous- catégories

Le prétraitement pour diversifier les données d'entrées:

- Rescale = $1./255$ → normalisation des valeurs
- shear_range = 0.2 → transformation géométrique
- zoom_range = 0.2 → zoom aléatoire
- horizontal_flip = True → image symétrique aléatoire



Classifiez automatiquement des biens de consommation

Étude des images

La Modélisation :

Partie exploratoire : on utilise la modélisation supervisée avec transfer learning

(CNN VGG 19)

⇒ Premier calcul : 7 catégories principales contenant chacune 150 photos.

⇒ Second calcul : 62 sous- catégories

Layer (type)	Output Shape	Param #
input_2 (InputLayer)	[(None, 224, 224, 3)]	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv4 (Conv2D)	(None, 56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0

Couches du réseau de neurones

Classifiez automatiquement des biens de consommation

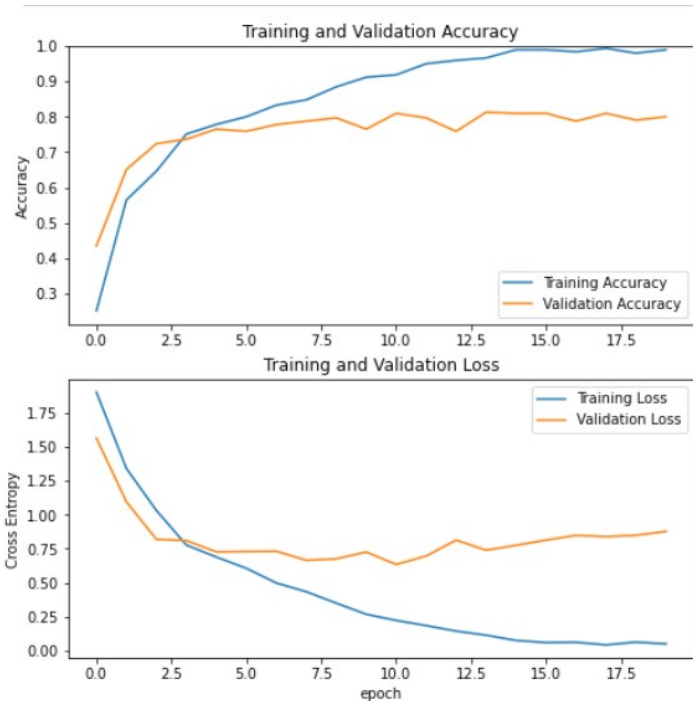
Étude des images

La Modélisation :

Partie exploratoire → on utilise la modélisation supervisée
avec transfer learning (CNN VGG 19)

⇒ Premier calcul : 7 catégories principales contenant chacune 150 photos.

⇒ **Accuracy = 0,8**



Classifiez automatiquement des biens de consommation

Étude des images

La Modélisation :

Premier calcul : Les photos mal classées.

Le jeu de validation comporte 7 catégories contenant chacune 45 photos
(315 photos au global pour la validation)

Catégories	Photos mal classées	
Home furnishing	12	27 %
Baby care	8	18 %
Watches	1	2 %
Homedecor & festive needs	17	38 %
Kitchen & dining	24	53 %
Beauty and personal care	9	20 %
Computers	18	40 %

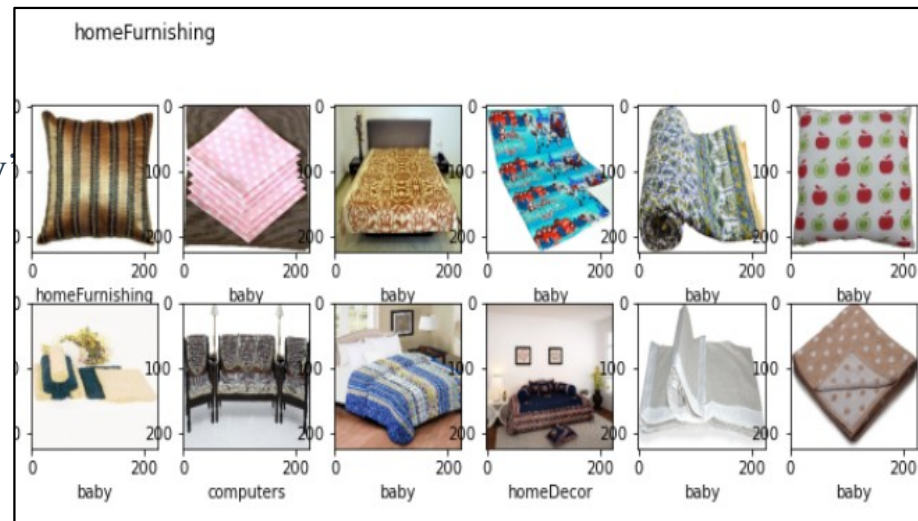
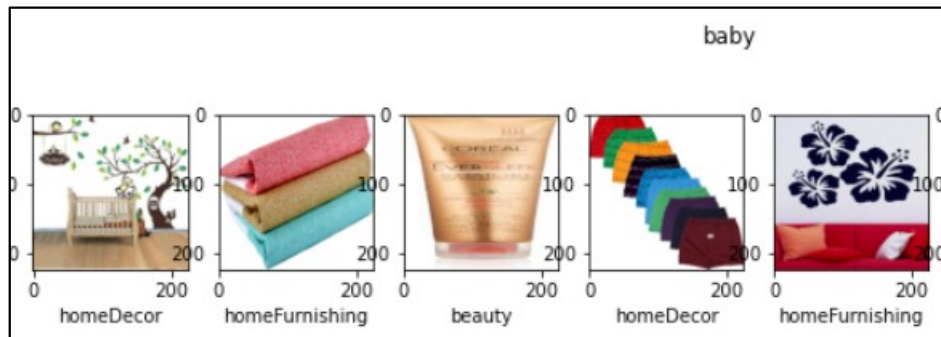
Classifiez automatiquement des biens de consommation

Étude des images

La Modélisation : Premier calcul, les photos mal classées.

Difficultés : il y a des similarités entre catégories

- bouteille de shampoing : catégorie 'baby' et la catégorie 'beauty'
- linge de maison : catégorie 'baby' et 'home Furnishing'



Classifiez automatiquement des biens de consommation

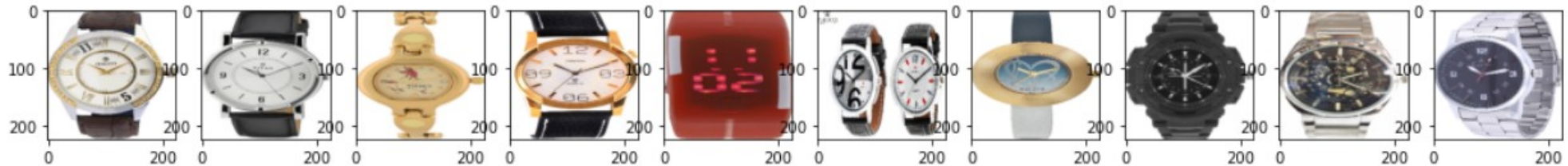
Étude des images

La Modélisation : Premier calcul, les photos mal classées.

→ Les photos de montres sont très similaires, il y a peu d'erreur de prédictions sur cette catégorie.

Watches

Watches



Classifiez automatiquement des biens de consommation

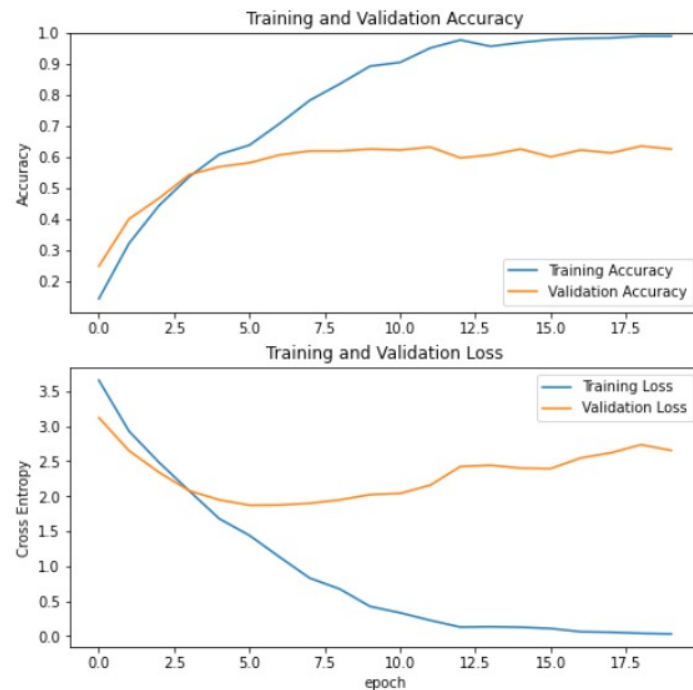
Étude des images

La Modélisation :

Partie exploratoire → on utilise la modélisation supervisée
avec transfer learning (CNN VGG 19)

⇒ Second calcul : 62 catégories.

⇒ **Accuracy = 0,6**



Classifiez automatiquement des biens de consommation

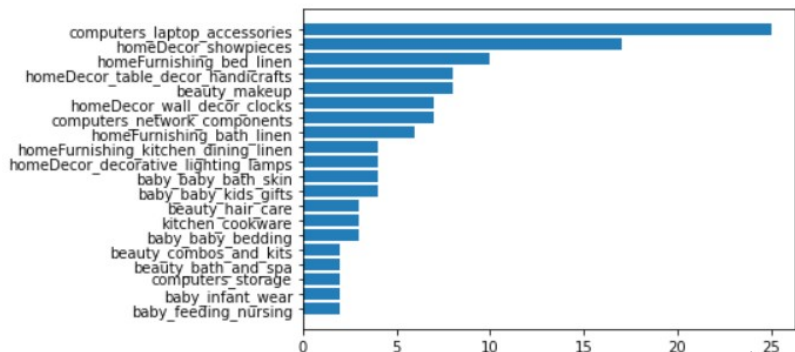
Étude des images

La Modélisation : Les photos mal classées.

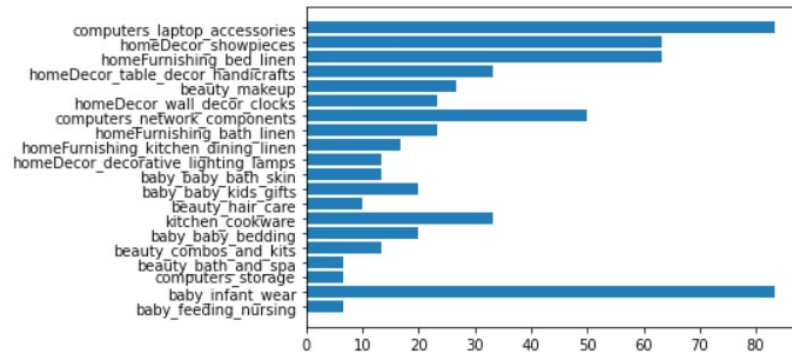
Le jeu de validation comporte 62 catégories contenant une répartition des photos déséquilibrées

(315 photos au global pour la validation)

→ Pour améliorer l'apprentissage : rééquilibrage des catégories et/ou augmenter le volume d'informations.



Nombre de photos mal classées



Nombre de photos par catégorie

Segmentez des clients d'un site e-commerce

- Contexte
- Présentation des informations disponibles
- Étude du texte
- Étude des images
- Conclusion et perspectives

Merci de votre attention !