



RNCP35288 CDSD

# The North Face ecommerce

## Unsupervised Machine Learning

Rédacteur : Jean-Yves Vuillequez





## Contexte et enjeu

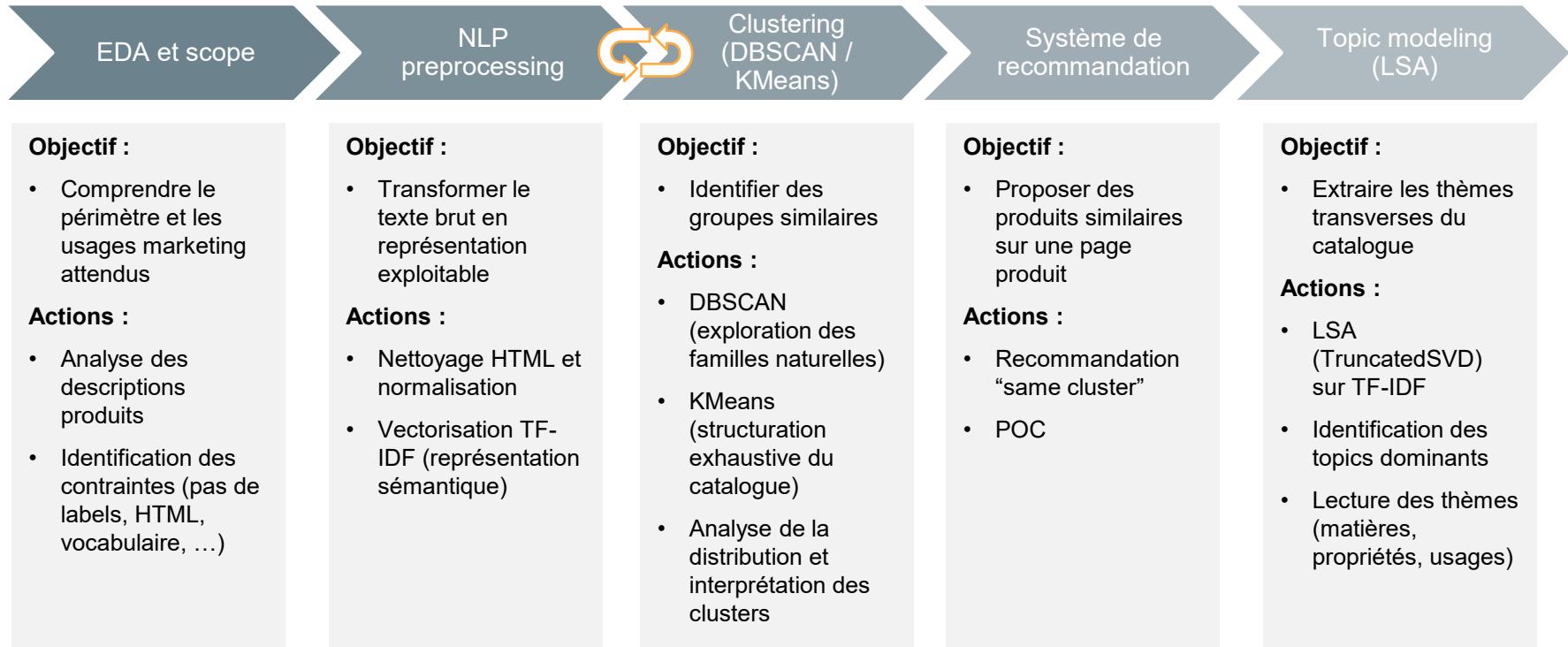


### Objectif :

- Le catalogue e-commerce de The North Face repose sur des descriptions produits textuelles riches mais hétérogènes.
- En l'absence de labels structurés ou de données utilisateurs exploitables, l'enjeu est d'améliorer la structuration du catalogue et les recommandations produits.

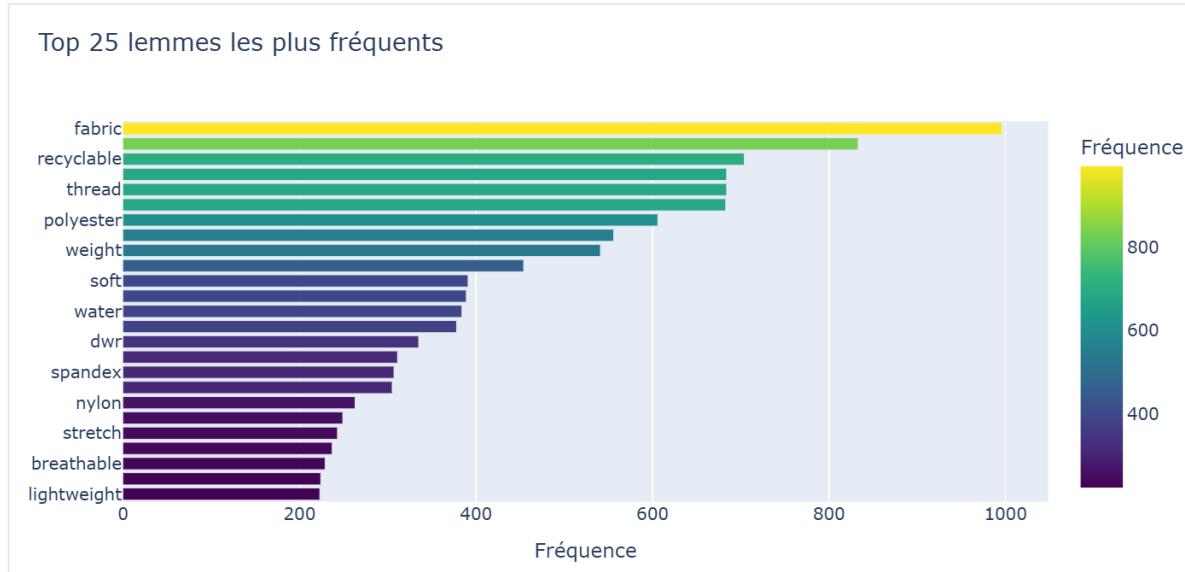


# Méthodologie du projet





# Scope sémantique



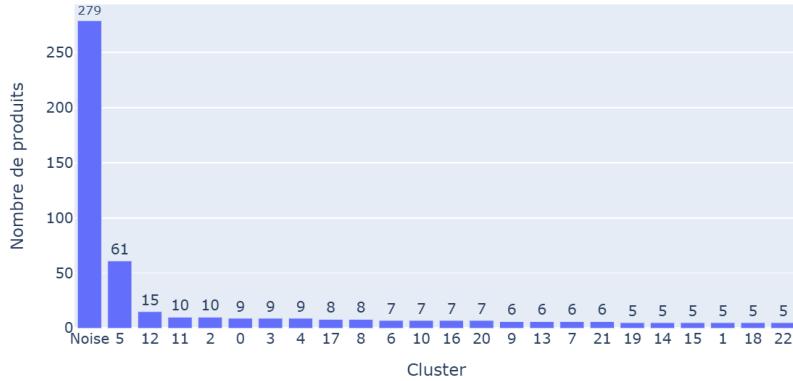
## Analyse sémantique :

- Catalogue e-commerce riche et des descriptions textuelles hétérogènes
- Très fortement orienté textile, avec une dominance du vocabulaire **matière / textile** (fabric, nylon, spandex, ...)
- Forte présence des propriétés techniques et orienté sur les **performances fonctionnelles** des produits (water, breathable, lightweight, ...)
- Signal fort sur la **durabilité / éco-conception** (704 occurrences)



# Clustering : comparaison des approches

Distribution des clusters DBSCAN (noise=55.8%)



Distribution des clusters KMeans (K=15)



## Analyse du Clustering :

- DBSCAN permet d'identifier des familles de **produits très homogènes**, mais laisse **une part importante** du catalogue non assignée.
- Kmeans structure **l'ensemble du catalogue** en un nombre maîtrisé de catégories
- **Choix : KMeans pour la recommandation**



# Résultats du clustering KMeans

## Cluster 0 (n=61)



Cluster 4 (n=15)



## Analyse KMeans :

- **15 clusters globalement équilibrés**, couvrant l'ensemble du catalogue
  - **Clusters sémantiquement cohérents**, structurés autour des matières et des usages
  - **Exemples** : waterproof / breathable, organic cotton, ...



# Système de recommandation

Produit 165. Matching: [2, 299, 258, 3, 301]

Produit 44. Matching: [10, 17, 402, 259, 92]

Produit 142. Matching: [10, 17, 402, 259, 70]



Produit sélectionné :

L/s steersman shirt - This ain't no city shirt, unless you're in Manila. The Steersman is de-twanged ranch hand in a...

Pour pourriez aussi aimer :

- 74: Logo hat - Your unflappable road-tripping companion - this 100% organic cotton hat disguises bed head, cuts glare...
- 356: S/s a/c shirt - Life's heated moments: meeting the girlfriend's parents, the job interview, runouts on the...
- 162: Astrid tank - Conjure up Cardiff's sholdery waves with the Astrid Tank. The unbelievably soft, strappy tank has a...
- 38: Elias fz sweatshirt - Clean as a Wingate splitter. Our Elias is made from a welcoming blend of 95% organic cotton and...
- 199: Margot dress - Dig the Margot out of your board bag, throw it on and head out for live salsa; this super comfy dress...

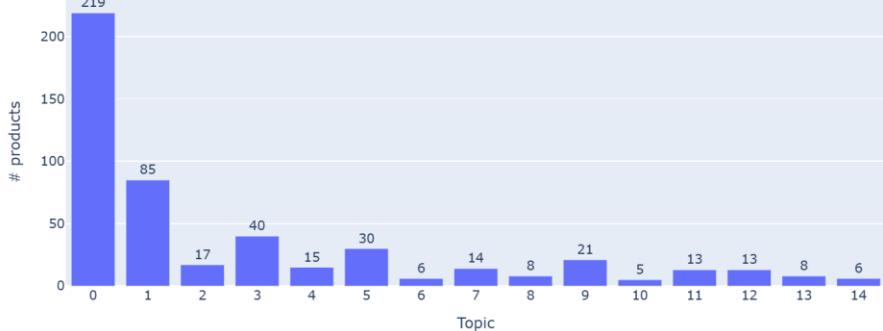
Principe de fonctionnement :

- Pour un produit donné, il est recommandé d'autres produits appartenant au même cluster
- logique simple, répétable et préserve l'intégrité de la recommandation
- Saisie d'un **product id** avec comme retour cinq recommandations associées



# Topic modeling

## Distribution des topics



## Analyse LSA :

- LSA permet à un produit d'être associé à plusieurs thèmes
  - **Topic(s) dominants** : Identifier des dimensions transverses (matières + vocabulaire très générique)
  - **Topics minoritaires** : Identifier les thématiques plus spécifiques (thèmes de "niche" qui pourraient devenir des **filtres / tags**)





Jedha

Merci pour votre attention  
Des questions ?

