



## Projet DALAS

A quoi ressemblerez-vous au printemps-été 2025 ?

**Zahra SIDDIQUE, Kajitha VIGNESWARALINGAM**

**Encadrants :** Laure Soulier, Jean-Noël Vittaut

Mai 2025

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Les données</b>	<b>4</b>
2.1	Articles . . . . .	4
2.1.1	Sources . . . . .	4
2.1.2	Nettoyage . . . . .	4
2.1.3	Structuration . . . . .	4
2.2	Produits . . . . .	5
2.2.1	Sources . . . . .	5
2.2.2	Nettoyage . . . . .	5
2.2.3	Structuration . . . . .	6
<b>3</b>	<b>Analyse exploratoire des données (EDA)</b>	<b>7</b>
3.1	Articles . . . . .	7
3.2	Analyse temporelle des tendances . . . . .	7
3.3	Produits . . . . .	7
3.4	Visualisations descriptives . . . . .	8
<b>4</b>	<b>Articles et prédictions</b>	<b>13</b>
4.1	Nommage des tendances : <i>n-grams</i> . . . . .	13
4.1.1	Avril 2025 . . . . .	14
4.1.2	Analyse globale . . . . .	16
4.2	Apprentissage des tendances : <i>embeddings</i> . . . . .	18
4.3	Regroupement des tendances : <i>clustering</i> . . . . .	18
<b>5</b>	<b>Recommandation de produits issue des prédictions</b>	<b>21</b>
5.1	Déduction des tendances . . . . .	21
5.2	Association trends-produits . . . . .	21
5.3	Recommandations finales . . . . .	21
5.4	Évaluation et visualisations . . . . .	22
<b>6</b>	<b>Limites et biais</b>	<b>25</b>
6.1	Biais de représentativité des données . . . . .	25
6.2	Biais liés au traitement du texte . . . . .	25
6.3	Biais dans le clustering . . . . .	25
6.4	Biais dans la recommandation . . . . .	25
6.5	Biais temporels . . . . .	25
<b>7</b>	<b>Conclusion</b>	<b>26</b>

# 1 Introduction

« *À quoi ressemblerez-vous au printemps-été 2025 ?* » Dans un contexte où les tendances vestimentaires évoluent rapidement sous l'influence des médias, des réseaux sociaux et des mutations sociétales, il devient pertinent d'interroger les codes esthétiques qui façonnent notre apparence à chaque saison. Ce projet vise à analyser les orientations stylistiques envisagées pour le printemps-été 2025, à travers une étude de contenus issus de magazines de mode et "lifestyle".

En mobilisant une approche qualitative fondée sur la lecture critique d'articles récents, nous nous attachons à identifier les pièces phares, les palettes chromatiques, les matières, imprimés et les coupes qui composent les tendances émergentes. L'objectif est d'en proposer une synthèse structurée, enrichie d'une sélection de tenues illustratives portées par nos modèles, traduisant les esthétiques dominantes, "trends" de la saison.

Ce travail a pour objectif d'observer et de présenter les grandes tendances de mode prévues pour le printemps-été 2025, afin de mieux comprendre les choix stylistiques qui s'offriront aux consommateurs. Il vise également à proposer des critères d'évaluation permettant d'éclairer les décisions d'achat pour le public, ainsi que les orientations de production pour les marques.

## 2 Les données

Cette section présente les données mobilisées dans notre étude : les articles de spécialisés en mode et les produits de vêtements en ligne. Nous détaillons ici les sources utilisées, les étapes de nettoyage nécessaires, ainsi que les méthodes de structuration appliquées pour préparer ces données à l'analyse.

### 2.1 Articles

#### 2.1.1 Sources

Les articles utilisés dans ce projet proviennent de trois sites spécialisés dans la mode et le lifestyle, reconnus pour la qualité de leur couverture des tendances saisonnières :

- *Women's Wear Daily* : <https://wwd.com/fashion-news/fashion-trends>
- *Who What Wear* : <https://www.whowhatwear.com/fashion/trends/spring> et <https://www.whowhatwear.com/fashion/trends/summer>
- *Harper's Bazaar* : <https://www.harpersbazaar.com/fashion/> et <https://www.harpersbazaar.com/fashion/trends/>

Ces sources ont été sélectionnées pour leur diversité éditoriale et leur régularité dans la publication de contenus liés aux tendances printemps-été.

#### 2.1.2 Nettoyage

Une fois extraits, les articles ont fait l'objet d'un prétraitement visant à éliminer les éléments non textuels (balises HTML résiduelles, scripts, mentions de cookies, formulaires de connexion, etc.). Ce nettoyage a permis d'obtenir un contenu textuel clair, exploitable pour l'analyse de tendances.

#### 2.1.3 Structuration

Les articles nettoyés ont été stockés dans des fichiers au format `.csv`, selon les champs suivants : `["Title", "Author", "Publication Date", "Content", "Source"]`.

Un classement temporel a été réalisé en répartissant les articles dans des dossiers mensuels selon leur date de publication. Par ailleurs, un fichier global, `all_data.txt`, regroupe l'ensemble des articles, toutes dates confondues, afin de faciliter les traitements textuels à grande échelle.

Répartition des fichiers par mois :

- avril-25 : 114 fichiers

- décembre-24 : 3 fichiers
- février-25 : 21 fichiers
- janvier-25 : 9 fichiers
- mai-25 : 66 fichiers
- mars-25 : 53 fichiers
- novembre-24 : 8 fichiers
- octobre-24 : 6 fichiers
- septembre-24 : 5 fichiers

Comme on le constate, la répartition du nombre de fichiers est hétérogène. Cela peut introduire un biais de représentation des données.

## 2.2 Produits

### 2.2.1 Sources

Afin de garantir la qualité des analyses futures portant sur les tendances de mode, une phase cruciale de **pré-traitement des données** a été appliquée au corpus initial issu du **scraping du site de la marque Miu Miu**.

Nous nous sommes notamment concentrées sur les **nouvelles collections du mois de mars 2025**, ainsi que sur une sélection de **produits permanents**.

Le fichier brut contenait des informations produit **hétérogènes** et **imparfaitement structurées**, nécessitant un **nettoyage rigoureux**.

### 2.2.2 Nettoyage

Les champs **titre** et **description**, riches en informations sémantiques, ont été standardisés à l'aide de techniques de **traitement automatique du langage naturel (NLP)**.

- Une **lemmatisation** a été appliquée à l'aide de **spaCy** avec le modèle **fr\_core\_news\_sm**, pour ramener chaque mot à sa forme canonique.
- Les **stop words** et ponctuations ont été supprimés afin de préparer le corpus à une future **vectorisation**.

Les champs **matières**, **couleurs** et **type** ont été extraits automatiquement à partir des descriptions textuelles par détection de **motifs fréquents** et de **mots-clés spécifiques**.

Le champ **image**, contenant l'URL de l'image du produit, a permis l'extraction des **couleurs dominantes** à l'aide d'un **clustering KMeans** appliqué sur les pixels.

- Les valeurs **RGB principales** ont été stockées dans les colonnes **dominant\_colors** et **flattened\_colors**.

- Le **nom des couleurs** associées a été obtenu par **mappage** à un dictionnaire CSS, stocké dans la colonne `color_names`.

### 2.2.3 Structuration

Les champs numériques comme `prix_numeric` ont été :

- Convertis au format `float64`.
- Nettoyés pour éviter toute **erreur de parsing**.

Les entrées **incomplètes** (produits sans image, sans description ou sans prix) ont été supprimées afin de garantir la **cohérence globale du dataset**.

Le jeu de données final, comportant 251 produits, est donc structuré selon les colonnes suivantes :

```
[url, titre, description, prix_numeric, couleurs, matières, type,  
  image, dominant_colors_array, color_names, dominant_colors,  
  flattened_colors]
```

### 3 Analyse exploratoire des données (EDA)

#### 3.1 Articles

#### 3.2 Analyse temporelle des tendances

Afin de mieux comprendre l'évolution des sujets abordés dans les articles au fil du temps, nous avons réalisé une analyse temporelle basée sur les dates de publication et d'apparition des produits.

Les volumes d'articles publiés par mois montrent une dynamique fluctuante, avec des pics notables correspondant à certaines périodes clés. Par exemple, on observe une augmentation significative des articles liés aux sacs *tote bag* au printemps 2025, ce qui traduit un intérêt croissant autour de ce produit.

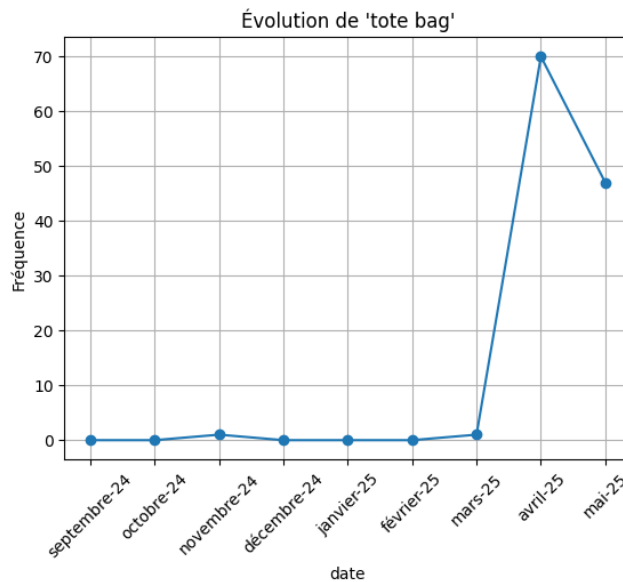


FIGURE 1 – évolution de "tote bag"

Cette analyse temporelle est essentielle pour contextualiser les regroupements thématiques obtenus par la suite via le clustering et pour ajuster les recommandations de produits en fonction des variations d'intérêt observées.

#### 3.3 Produits

**Objectifs :** L'analyse exploratoire des données (EDA) vise à comprendre la **structure statistique du corpus nettoyé**, à détecter d'éventuels **biais ou incohérences**, et à orienter les **choix méthodologiques** pour les modèles à venir.

Dans cette section, nous explorons les **distributions des variables** et analysons les **relations entre elles**, afin d'identifier les **trends émergents**.

**Dimensions du corpus :** Le jeu de données final contient **251 produits de mode**, structurés en **12 colonnes**, incluant des champs textuels (titres, descriptions), numériques (prix), catégoriels (matières, couleurs, types), et visuels (couleurs extraites des images) .

### 3.4 Visualisations descriptives

Afin de mieux comprendre la distribution des variables du corpus, nous avons généré plusieurs visualisations, présentées ci-dessous.

**Dimensions du corpus :** Le jeu de données final contient **251 produits de mode**, structurés en **12 colonnes**, incluant des champs textuels (titres, descriptions), numériques (prix), catégoriels (matières, couleurs, types), et visuels (couleurs extraites des images).

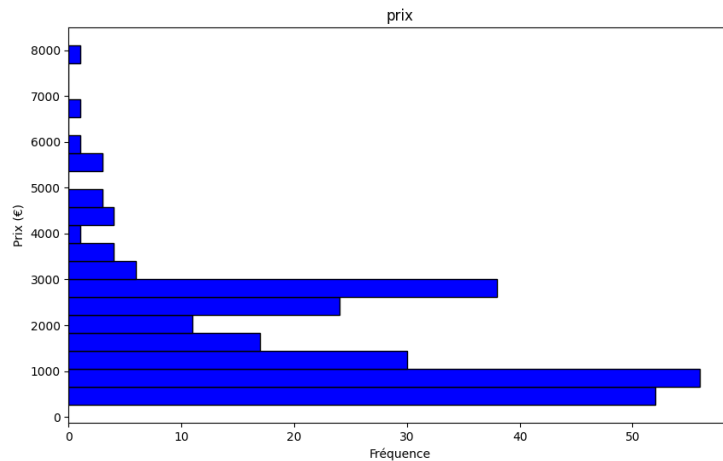


FIGURE 2 – Distribution des prix des produits Miu Miu

La majorité des produits se situent dans une gamme de **prix élevés**, typique du segment luxe. La distribution est asymétrique, avec un pic autour de **1500–2000€**.



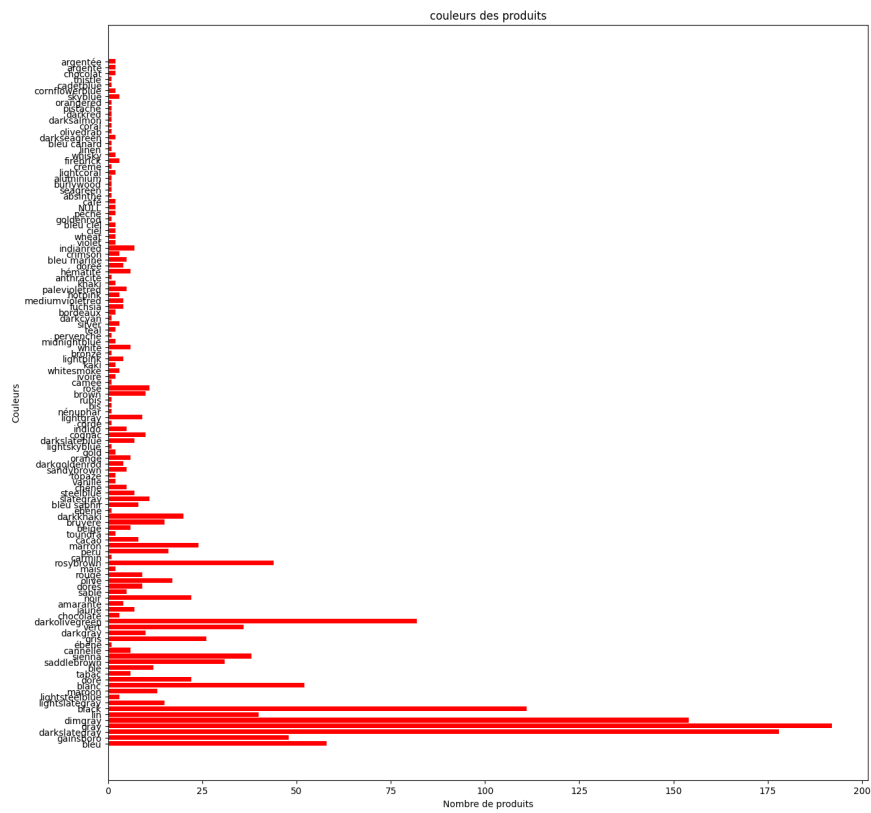


FIGURE 3 – Répartition des couleurs dominantes

La palette de la collection met en avant des tons **neutres et sobres** tels que le *noir*, *blanc*, *gris* ainsi que des teintes pastel comme le *rose clair* ou le *bleu ciel*, révélant une esthétique minimaliste et féminine.



FIGURE 4 – Matières les plus utilisées dans les produits

Les matières dominantes sont le **coton**, la **soie**, le **cuir** et la **laine**. Cette répartition reflète une volonté de conjuguer **qualité textile** et **confort**, caractéristiques clés du prêt-à-porter de luxe.

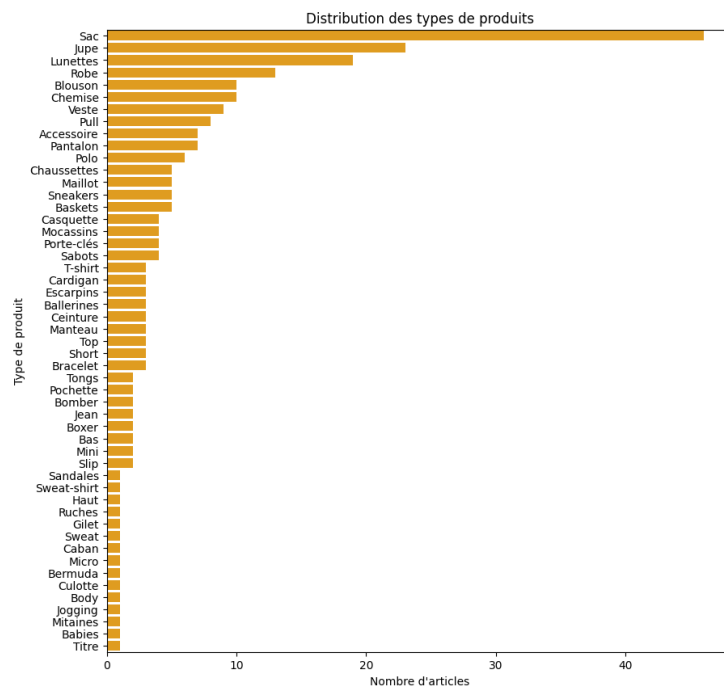


FIGURE 5 – Répartition des types de produits

Les *vestes*, *robes* et *pantalons* sont les types de produits les plus présents, ce qui témoigne de l'importance de ces pièces dans les collections de saison. Les accessoires et chaussures sont plus rares dans ce corpus.



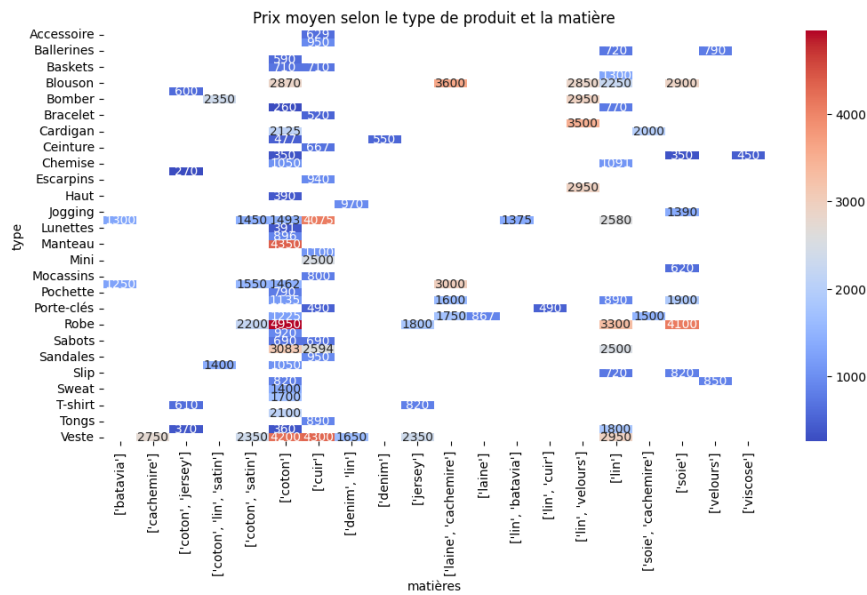


FIGURE 8 – Fréquence croisée entre types et matières

Cette heatmap révèle les associations dominantes entre **types de produits** et **matières**. Par exemple, les *cardigans* sont très souvent en *cachemire*, tandis que les *jupes* utilisent davantage le *coton*.

**Analyse sémantique des descriptions produits :** Afin de mieux comprendre la structure thématique des descriptions textuelles, nous avons appliqué une méthode de vectorisation TF-IDF, suivie d'un clustering non supervisé (KMeans) sur les vecteurs obtenus. Ce processus a permis d'identifier **6 clusters sémantiques distincts** regroupant les produits ayant une description assez similaires.

Les étapes clés de cette analyse sont les suivantes :

- **Nettoyage linguistique** : suppression des mots vides (*stopwords*) en français.
- **Vectorisation** : représentation des descriptions par des vecteurs TF-IDF avec un maximum de 100 termes discriminants.
- **Clustering KMeans** : segmentation des produits en 6 groupes selon leurs caractéristiques lexicales.
- **Extraction des mots-clés** : pour chaque cluster, les termes les plus représentatifs ont été extraits à partir des centres de gravité des clusters.

**Analyse des couleurs dominantes :** Pour caractériser visuellement les produits, un clustering a été réalisé sur les couleurs dominantes extraites des images. Ce traitement a permis d'identifier **12 clusters de couleurs** représentatifs des

principales teintes présentes dans le corpus. Chaque cluster correspond à une couleur moyenne (RGB) obtenue via K-Means sur les pixels dominants.

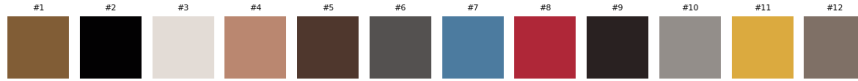


FIGURE 9 – Palette des 12 clusters de couleurs dominantes extraits

Les résultats mettent en évidence une diversité de tons allant : des **neutres chauds** (beige, brun, tabac), aux **gris et noirs** classiques, en passant par des **teintes vives ou pastel** (fuchsia, bleu, vert doux).

## 4 Articles et prédictions

À partir du corpus d’articles collectés, nous identifions les **tendances dominantes** à l’aide de **méthodes textuelles**. Cette section détaille les techniques utilisées pour :

- **extraire les tendances**,
- analyser leur **fréquence et structure**,
- regrouper celles qui relèvent de **dynamiques similaires**.

### 4.1 Nommage des tendances : *n-grams*

Afin d’identifier les **motifs récurrents** dans les articles de presse de mode, nous extrayons des **n-grams** (séquences de  $n$  mots consécutifs) sur deux niveaux :

- sur l’ensemble du **corpus agrégé**,
- et sur des **sous-corpus mensuels**.

Nous considérons des valeurs de  $n$  allant de 1 (unigrammes) à 5 (penta-grammes), permettant d’identifier aussi bien des mots-clés isolés (“*shirt*”, “*butter yellow*”) que des expressions complexes (“*mary jane flat*”, “*high rise wide leg*”).

Des **statistiques de fréquence** sont générées pour visualiser les n-grams les plus représentés. Trois types de visualisations sont proposées :

- **Nuage de mots** global pour les trigrams les plus fréquents.
- **Histogramme** des bigrams les plus présents entre février et avril.

Ces représentations offrent une première lecture lexicale des tendances émergentes.

Observons ces visualisations pour le mois d’avril 2025 et l’analyse tous mois confondus, avec les top 100 n-grams.

#### 4.1.1 Avril 2025

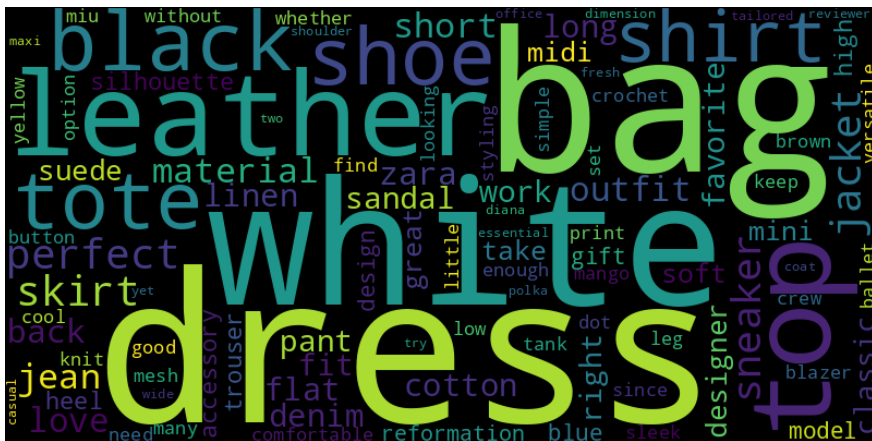


FIGURE 10 – Wordcloud unigrams - avril 2025



FIGURE 11 – Wordcloud bigrams - avril 2025

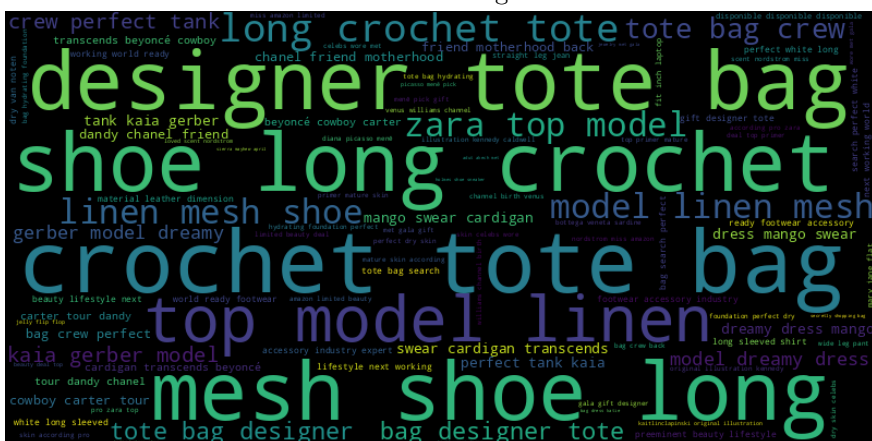


FIGURE 12 – Wordcloud trigrams - avril 2025



### 4.1.2 Analyse globale



FIGURE 16 – Wordcloud unigrams - global

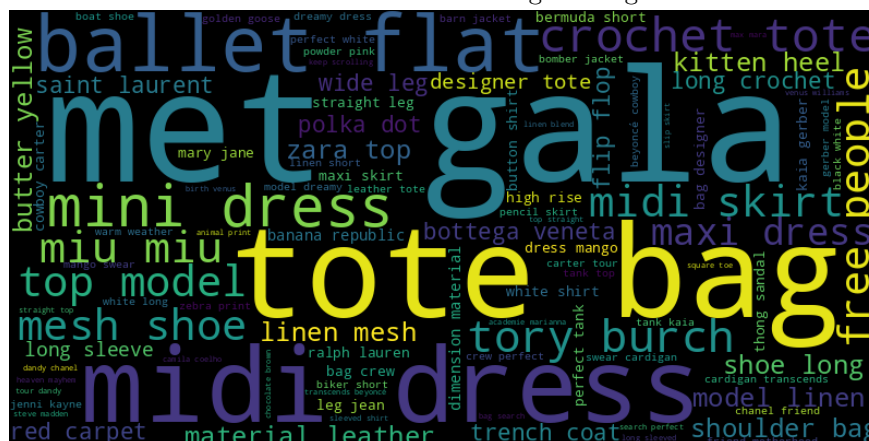


FIGURE 17 – Wordcloud bigrams - global

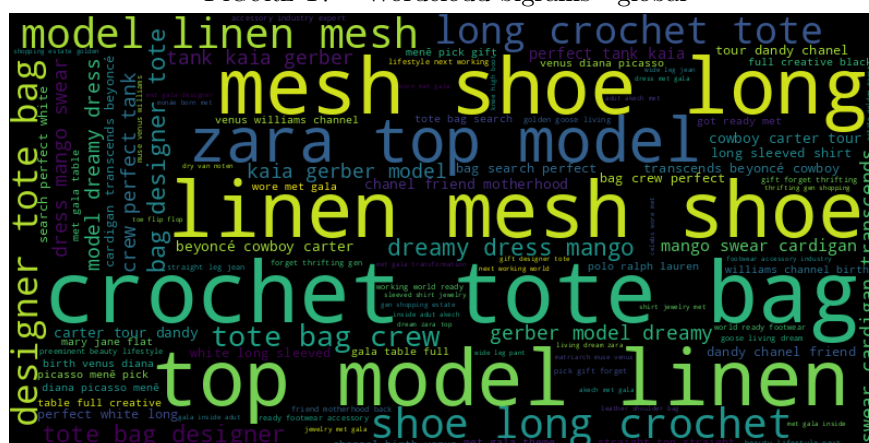


FIGURE 18 – Wordcloud trigrams - global



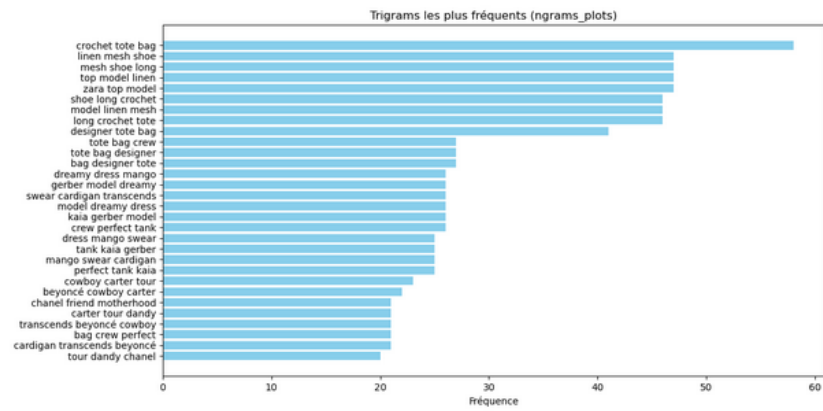


FIGURE 19 – Barplot trigrams - global

## 4.2 Apprentissage des tendances : *embeddings*

Pour enrichir l'analyse sémantique, nous projetons les n-grams dans un **espace vectoriel** grâce à des **modèles d'embeddings**.

Nous utilisons le modèle all-MiniLM-L6-v2 de *sentence-transformers*, entraîné pour capturer les **relations sémantiques** dans les textes en anglais.

L'apprentissage est focalisé sur les **bigrams** et **trigrams**, particulièrement efficaces pour représenter :

- des **couleurs** (“*powder pink*”).
- des **coupes** (“*oversized blazer*”).
- des **motifs** (“*polka dots*”),
- des **matières** (“*linen shirt*”),
- des **marques** (“*miu miu*”).
- des **pièces** (“*flip flop*”).

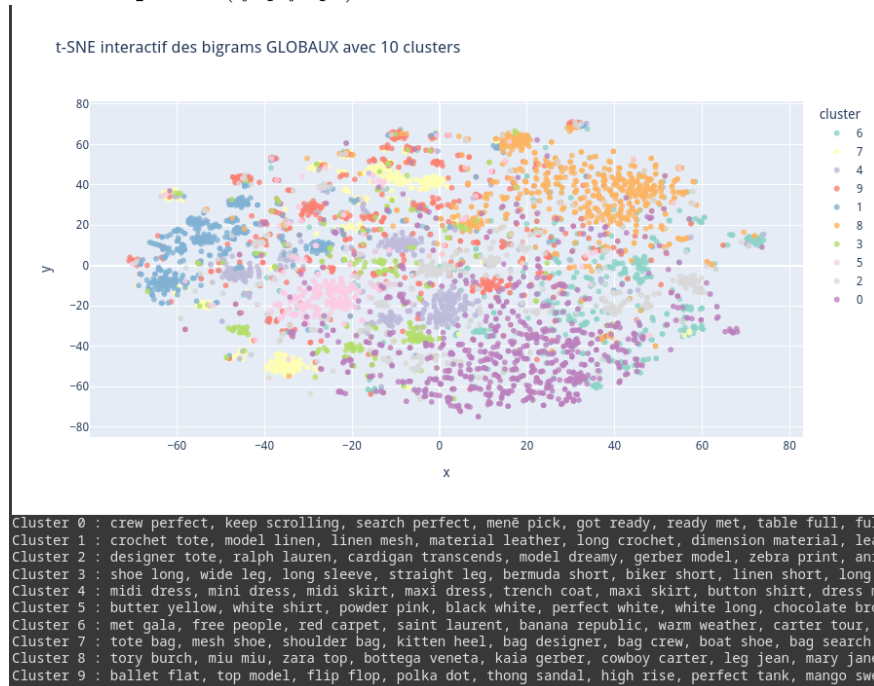


FIGURE 20 – t-SNE sur les bigrams globaux, nbsamples=5000, 10 clusters

## 4.3 Regroupement des tendances : *clustering*

Les vecteurs d'embeddings sont ensuite regroupés par **similarité sémantique** via un **algorithme de clustering non supervisé** (KMeans). L'objectif est de faire émerger automatiquement des **familles stylistiques** cohérentes. On teste les valeurs de clusters 10,30 et 50, appliquées sur nbsamples=5000, sur bigrams (figure 20) et trigrams (figure 21-22) globaux (figures 20-21-22) et par

mois.

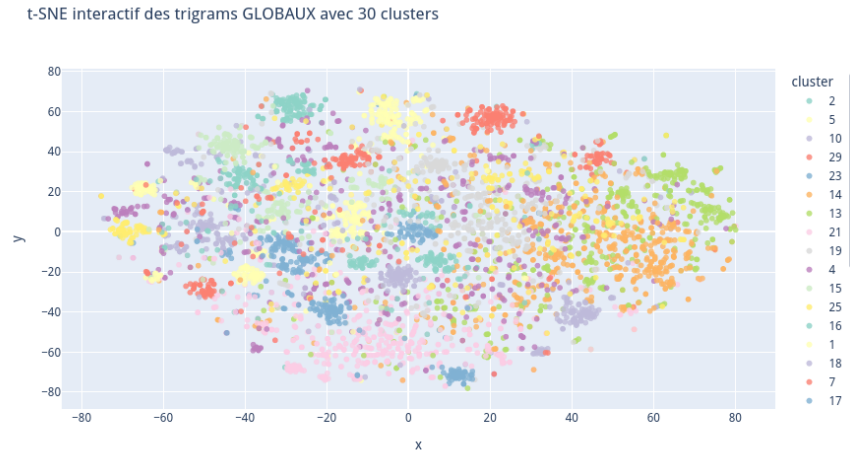


FIGURE 21 – t-SNE sur les trigrammes globaux, nbsamples=5000, 30 clusters  
Chaque cluster agrège des expressions stylistiquement proches. Par exemple :

```
Cluster 0 : leather slingback pump, babouche leather pump, kitten heel pump, mary jane pump, pump tory burd
Cluster 1 : search perfect white, full creative black, pageant rhapsody blue, creative black back, yellow p
Cluster 10 : top model linen, model linen mesh, long crochet tote, dry van Noten, linen crochet tote, linen
Cluster 11 : buzziest pop culture, striped poplin shirt, banana republic poplin, basic poplin shirt, organ
Cluster 12 : gen shopping estate, thrifting gen shopping, site secretly shopping, scent nordstrom miss, dea
Cluster 13 : swear cardigan transcends, got ready met, friend motherhood back, date since updated, earlier
Cluster 14 : gerber model dreamy, cowboy carter tour, carter tour dandy, tour dandy chanel, preeminent beau
Cluster 15 : perfect white long, straight top straight, straight leg jean, wide leg jean, wide leg pant, kn
Cluster 16 : white long sleeved, leather bomber jacket, skin celebs wore, long sleeve top, white long slee
Cluster 17 : met gala theme, met gala table, wore met gala, gala table full, met gala transformation, jewel
Cluster 18 : diana picasso mené, picasso mené pick, venus diana picasso, illustration kennedy caldwell, ori
Cluster 19 : crew perfect tank, perfect tank kaia, table full creative, everybody pantsless met, strike per
Cluster 2 : crochet tote bag, designer tote bag, tote bag crew, tote bag designer, bag designer tote, bag d
Cluster 20 : adut akech met, inside adut akech, szymon brzówska stalker, barbaro got ready, theme met diotin
Cluster 21 : kaia gerber model, tank kaia gerber, beyoncé cowboy carter, chanel friend motherhood, transce
Cluster 22 : fisherman sandal balance, bitton establishment production, frame acetate sunglass, reformation
Cluster 23 : dreamy dress mango, model dreamy dress, dress mango swear, sheer dress met, embrace sheer dres
Cluster 24 : superfine tailoring black, tems bold met, accessorizing tems bold, primer mature skin, skin ad
Cluster 25 : long sleeved shirt, polo ralph lauren, ralph lauren chino, white button shirt, long sleeve shi
Cluster 26 : accessory industry expert, working world ready, gave lesson accessorizing, lesson accessorizi
Cluster 27 : help young people, free people revolve, free people midnight, free people sweet, free people b
Cluster 28 : prada miu miu, prada tory burch, dress prada floral, persol sunglass classic, prada anemone p
Cluster 29 : zara top model, dream zara top, living dream zara, industry expert zara, expert zara top, acco
Cluster 3 : faithfull marais midi, marais midi dress, dress polka dot, suede ballet flat, shoulder midi dre
Cluster 4 : mango swear cardigan, boater hat everybody, gala boater hat, biggest hair met, curl biggest hai
Cluster 5 : linen mesh shoe, mesh shoe long, shoe long crochet, ready footwear accessory, world ready footw
Cluster 6 : mango zebra print, animal print accessory, animal print flat, dot print scarf, polka dot print,
Cluster 7 : mené pick gift, gift forget thrifting, forget thrifting gen, pick gift forget, sleeved shirt je
Cluster 8 : straight form penske, top straight top, sign straight top, top straight form, del rey got, rey
Cluster 9 : material leather measurement, material leather dimension, leather flip flop, material suede lea
```

FIGURE 22 – Clusters obtenus

Ces regroupements permettent de dégager des **macro-tendances** esthétiques. En sélectionnant les n-grams les plus représentatifs de chaque cluster, nous générons des **étiquettes qualitatives** utilisables :

- pour **annoter les collections**,
- indexer les produits,
- ou créer des **résumés de tendances mensuels**.

Pour 50 clusters sur les données globales regroupés en bigrams, on distingue les clusters significatifs qui donnent des résultats concrets : le cluster 13 notamment qui semble représenter un axe des couleurs tendances de la saison : *butter yellow, powder pink, chocolate brown, olive green, cherry red, navy blue, pop red, espresso brown, mint green, shade blue...*



FIGURE 23 – t-SNE sur les biigrams globaux, nbsamples=5000, 50 clusters

## 5 Recommandation de produits issue des prédictions

### 5.1 Dédution des tendances

Les tendances ont été extraites via un clustering **KMeans** appliqué sur des vecteurs TF-IDF construits à partir des descriptions produits. Chaque cluster représente un groupe sémantique cohérent de produits partageant des caractéristiques stylistiques similaires. L'analyse des termes discriminants a permis de nommer les tendances, par exemple « vintage sportif » ou « minimalisme technique ».

### 5.2 Association trends-produits

Chaque produit est associé à un cluster (tendance) selon sa proximité avec le centroïde. Les produits les plus représentatifs, situés près du centroïde, illustrent bien la tendance à laquelle ils appartiennent. Cette association facilite la compréhension des regroupements stylistiques et leur exploitation commerciale.

### 5.3 Recommandations finales

La recommandation finale s'appuie sur une représentation vectorielle riche de chaque produit, construite à partir de plusieurs informations clés : le titre, la description, le type de vêtement, les matières et les couleurs. Ces données textuelles et catégorielles sont encodées via un modèle *SentenceTransformer* multilingue, qui génère des **embeddings globaux** capturant la sémantique et le contexte stylistique de chaque produit.

Pour enrichir ces représentations, nous avons intégré des éléments issus du **dataset des articles de mode**, qui reflètent les tendances et le vocabulaire courant dans la presse spécialisée (comme *Harper's Bazaar* ou *Women's Wear Daily*). Cette fusion permet au modèle de mieux saisir les expressions et concepts liés à la mode actuelle, renforçant ainsi la pertinence des embeddings.

La similarité cosinus est ensuite calculée entre ces vecteurs pour identifier les produits les plus proches stylistiquement dans cet espace vectoriel. Pour garantir la qualité des recommandations, seules les références marquées comme “*tendance*” (variable `is_trendy` = 1), sur la base de l'analyse combinée des données produits et éditoriales, sont retenues.

Ainsi, l'association des deux datasets — produits et articles — enrichit considérablement la compréhension du style et des tendances, offrant un système de recommandation plus robuste, précis et aligné avec l'univers de la mode contemporaine.

- Produit cible : *Cardigan En Cachemire Et Soie À Maille Torsadée Bleu*
- Produits recommandés :
  - *Short En Tissu Technique Marron Cacao*
  - *Bracelet En Cuir Avec Boucle Bruyère*
  - *Chemise En Popeline À Carreaux Céleste+bleu*
  - *Jupe Courte À Carreaux Marron Cacao*
  - *Bracelet En Cuir Avec Boucle Noir*

Cette méthode permet de proposer des alternatives ou des compléments stylistiques cohérents, fondés sur la similarité sémantique des descriptions.

## 5.4 Évaluation et visualisations

Afin de mesurer la performance du modèle de classification *XGBoost* prédisant si un produit est tendance, plusieurs métriques et visualisations ont été produites :

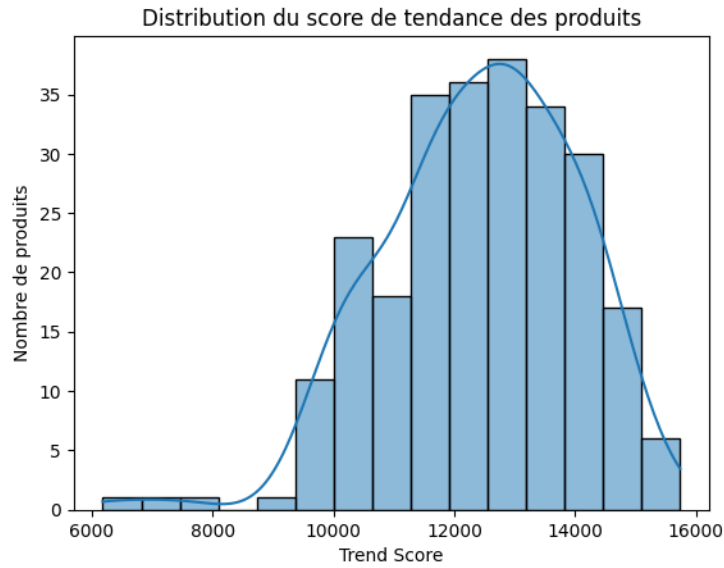


FIGURE 24 – Distribution des scores de similarité cosine entre produits.

On observe une concentration autour de scores élevés ( $>0.7$ ), indiquant que les produits tendance partagent des caractéristiques sémantiques fortes (même description, matière ou style). Cela valide l'approche par embeddings pour capter les similarités stylistiques.

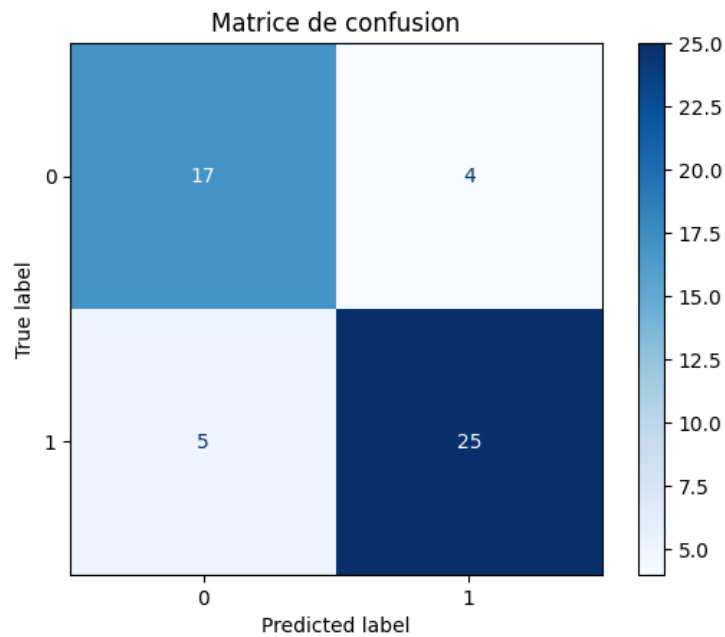


FIGURE 25 – Matrice de confusion du classificateur XGBoost : évaluation des prédictions tendance / non tendance.

Les vrais positifs (TP) et vrais négatifs (TN) dominent, nous indiquant qu'on a une bonne précision et un bon recall

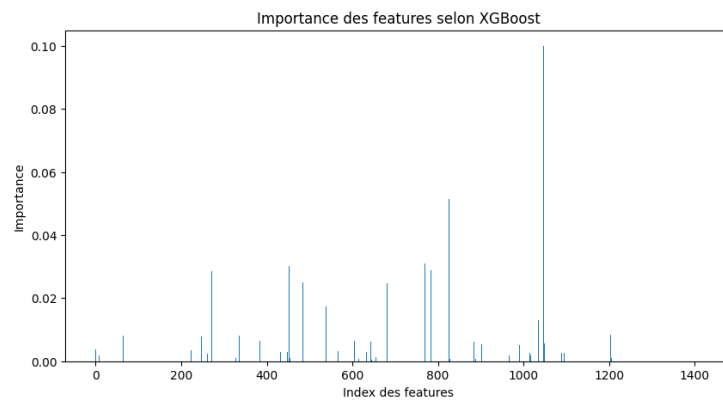


FIGURE 26 – Importance des features utilisées par le modèle XGBoost, avec une prédominance des embeddings globaux.

On constate que les embeddings globaux (textuels, couleurs, matières) sont les plus discriminants. Cela montre que les informations issues du NLP et du traitement d'image ont une valeur prédictive significative pour l'identification de produits tendance.

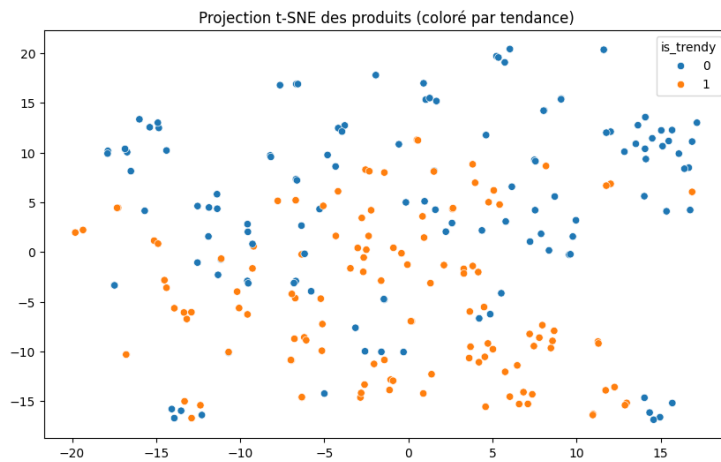


FIGURE 27 – Projection t-SNE des embeddings produits, montrant la séparation entre produits tendance et non tendance.

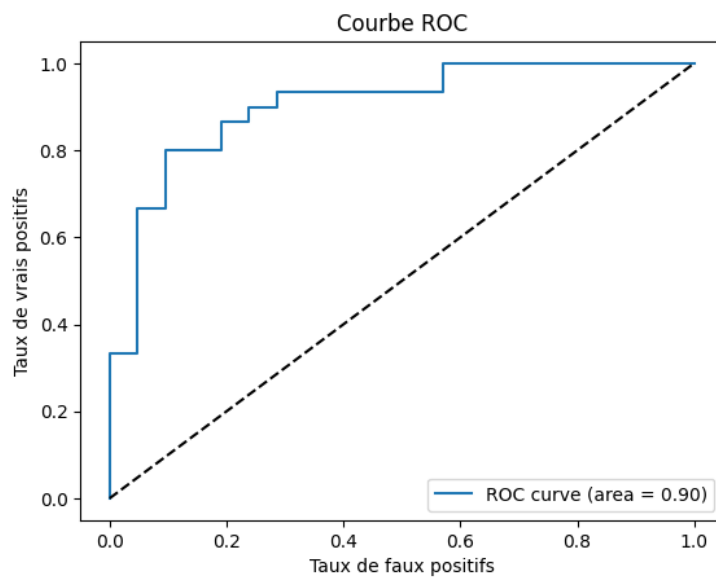


FIGURE 28 – Courbe ROC du classificateur XGBoost, illustrant la capacité du modèle à discriminer les produits tendance.

Le fait que la courbe ROC est au-dessus de la diagonale, montre que le modèle est fiable.

Ces analyses confirment la pertinence du modèle pour identifier les produits tendance et proposer des recommandations basées sur la similarité sémantique.



## 6 Limites et biais

Dans ce projet, plusieurs biais potentiels peuvent influencer les résultats et doivent être pris en compte lors de l'interprétation des analyses et des recommandations.

### 6.1 Biais de représentativité des données

Les articles analysés ne couvrent pas nécessairement toutes les tendances de manière uniforme. Certaines périodes temporelles sont surreprésentées, ce qui peut biaiser l'analyse des évolutions des tendances. En effet, certains mois comportent beaucoup plus d'articles que d'autres ce qui impactent l'analyse des articles.

### 6.2 Biais liés au traitement du texte

Le nettoyage des données textuelles, notamment la suppression des stop-words et la lemmatisation, peut entraîner la perte d'informations pertinentes. Par ailleurs, le mélange de langues, principalement l'anglais et le français dans les descriptions des produits et articles, complique la vectorisation TF-IDF. Cela peut engendrer une dispersion des termes similaires dans différentes langues, réduisant ainsi la qualité du clustering et la cohérence des tendances extraites.

### 6.3 Biais dans le clustering

Le choix du nombre de clusters influence la granularité des tendances détectées. Certains produits peuvent être mal regroupés, ce qui impacte l'interprétation des clusters et la pertinence des tendances extraites.

### 6.4 Biais dans la recommandation

Le système de recommandation, basé sur la similarité cosinus, tend à privilégier les produits similaires à ceux déjà connus par l'utilisateur, renforçant ainsi un effet de « bulle » d'intérêt (biais de filtre).

### 6.5 Biais temporels

Les tendances et préférences évoluent rapidement, mais notre modèle ne prend pas pleinement en compte cette dynamique temporelle.

Reconnaître ces biais nous permet de savoir que les résultats doivent être interprétés avec prudence, et que des améliorations futures pourraient intégrer une meilleure prise en compte des aspects linguistiques et temporels.

## 7 Conclusion

À travers cette étude, nous avons analysé les tendances stylistiques présentes pour la saison printemps-été 2025 à partir de sources éditoriales spécialisées et de fiches produits. Cette double approche nous a permis d'identifier les grandes orientations de la mode à venir, tant sur le plan esthétique que commercial.

En croisant les prédictions issues des articles avec les produits disponibles ou émergents, nous avons pu établir des correspondances concrètes entre discours de tendance et offres vestimentaires. Ce travail contribue à une meilleure compréhension des dynamiques de la mode saisonnière, et offre des repères à la fois pour les consommateurs en quête d'inspiration et pour les marques souhaitant ajuster leurs collections.

En somme, ce projet illustre comment l'analyse de contenus textuels et de produits peut alimenter une réflexion sur les choix stylistiques individuels et les orientations stratégiques du secteur de la mode.