

PROJET IA

ETHIC FASHION

De la fast fashion à la mode éthique

Année Académique : 2022 – 2023



Problématique:

Comment les entreprises de la fast fashion peuvent-elles encourager la créativité et l'innovation dans la conception de vêtements éthiques pour répondre aux tendances de la mode tout en réduisant leur impact environnemental et social?

Sabine DJETRAN - M2DA1-JV / Marie Roger NOAH NOAH - M2JV-DS1/ Awa DIALLO
- M2JV-DS1 / Sorelle PENANDJO - M2DA1-JV

TABLE DES MATIÈRES

I. RÉSUMÉ

II. INTRODUCTION

III. DÉMARCHES

A. Collecte de données

B. Analyse de données

Statistique sur l'importation mondiale du textile

Statistique sur l'exportation mondiale du textile

Les matières premières

Le prix des articles

C. Modélisation

1. Random Forest

2. SVM

3. Régression logistique

IV. RÉSULTATS

A. Random Forest

B. SVM

C. Régression logistique

V. INTERPRÉTATION

VI. CONCLUSION

RÉFÉRENCES

ANNEXE

1. ANNEXE 1: Carte proportionnelle sur l'importation du textile en 2021 dans le monde

2. ANNEXE 2: Carte proportionnelle sur l'exportation du textile en 2021 dans le monde

3. ANNEXE 3 : Top 10 des marques européennes de fast fashion sur la base des unités vendues par an dans le monde au cours de l'exercice 2018/2019

4. ANNEXE 4 : Notebook ProjetIA_Datavisualisation

5. ANNEXE 5 : Notebook FastFashionScraping

6. ANNEXE 4 : Notebook ProjetIA_FastFashion_Analyse

7. ANNEXE 4 : Notebook ProjetIA_FastFahion_Modélisation

I. RÉSUMÉ

Au cours des dernières décennies, l'industrie de la fast fashion a connu une croissance exponentielle, qui s'est avérée préjudiciable pour l'environnement et les travailleurs du secteur. Ainsi, les entreprises de la fast fashion sont confrontées à un double défi : répondre aux tendances de la mode tout en réduisant leur impact environnemental et social.

Pour les aider dans cette démarche, notre projet a consisté à collecter des données sur l'industrie de la mode, et plus particulièrement sur celle de la fast fashion. Nous avons ensuite concentré notre étude sur l'un des leaders mondiaux de ce marché, en collectant des données sur leurs articles afin de développer un modèle prédictif capable de classer les vêtements en différentes catégories en fonction du prix, de la composition et des données d'éco-étiquetage. L'objectif étant d'aider les clients à trouver plus rapidement des types de vêtements spécifiques. Les résultats de l'analyse de ces données nous ont permis de développer un concept novateur pour une mode plus durable, que les autres acteurs de la fast fashion pourront adopter.

II. INTRODUCTION

Selon une étude menée par le groupe consultatif londonien sur l'environnement, The Eco Experts, les secteurs les plus émetteurs de gaz à effet de serre (GES) sur une base annuelle sont l'énergie, les transports, la fabrication et la construction, l'agriculture, la distribution alimentaire, la mode et les technologies. (La mode serait en réalité la sixième industrie la plus polluante au monde [1]. Le développement durable étant l'un des axes de l'enseignement du groupe GEMA. La thématique globale du projet IA de cette année est : « Ethical Fashion, De la fast fashion à la mode éthique. La fast fashion désigne un concept qui englobe les marques de vêtements produisant rapidement et à bas coût des collections de vêtements, souvent en lançant jusqu'à 36 collections par an, tandis qu'une marque traditionnelle n'en sort que 4. Elle est apparue dans les années 1990 dans les villes de la mode comme New York ou Londres, et depuis les années 2000, des marques telles que Zara, H&M et Topshop ont connu une croissance considérable pour devenir de véritables géants de la mode rapide [2]. Ces marques ont repris les looks et les éléments de design des grandes maisons de couture et les ont reproduits rapidement et à moindre coût, en s'approvisionnant dans les pays émergents. La fast fashion est décriée à ce jour pour ses nombreuses conséquences sociales et environnementales.

Après les examens déjà faits sur les effets environnementaux, sociaux et énergétiques de l'industrie de la mode, ainsi que les données économiques et les comportements d'achat, de conservation, de revente, de recyclage et de mise au rebut, il est crucial de proposer de nouvelles pistes de réflexion pour améliorer une situation qui est actuellement très préoccupante. C'est ainsi qu'on va se pencher sur la question : **Comment les entreprises de**

la fast fashion peuvent-elles encourager la créativité et l'innovation dans la conception de vêtements éthiques pour répondre aux tendances de la mode tout en réduisant leur impact environnemental et social?

III. DÉMARCHES

A. Collecte de données

Après avoir établi une problématique liée à la thématique centrale du projet IA de cette année. Nous avons entrepris de recueillir des données permettant d'analyser les effets environnementaux ainsi que des données économiques. Cette analyse nous a permis de mieux orienter nos recherches vers les données qui nous permettront de dégager facilement des solutions innovantes pour l'amélioration de ce phénomène qui devient de plus en plus préoccupant.

Pour avoir un diagnostic de la situation de ces dernières années nous avons recueillir des bases de données sur :

- Les importations et les exportations mondiales de textile via le site de oec.
- Les principales marques européennes de fast fashion en termes d'unités vendues via le site de statista.

Par ailleurs, en France, Shein, Zara et H&M sont les trois principaux acteurs du marché de l'ultra-fast fashion en 2022, représentant respectivement 33,8%, 36,4% et 29,8% de parts de marché en valeur. Bien que les chiffres soient équilibrés entre les trois géants, Shein perd des parts de marché (les Français commencent doucement à délaisser la marque)tandis que Zara et H&M en gagnent. Figure phare de la fast fashion, Zara possède aujourd'hui près de 6 500 boutiques autour du globe [3] ; c'est la raison pour laquelle on a scrappé des données des articles liés à cette enseigne pour répondre à notre problématique.

Dans un premier temps, notre approche consistait à collecter les données requises en procédant au scrapping du site de la marque. Cependant, nous avons rencontré une difficulté technique due à la présence d'un pare-feu qui a empêché l'extraction des données souhaitées. Face à cette contrainte, nous avons jugé qu'il était plus opportun de chercher une plateforme alternative permettant de recueillir les informations concernant les articles de ZARA. C'est ainsi que nous avons découvert une table de données portant sur la "Fast Fashion Eco-data" sur le site zenodo.org [4] . Ce jeu de données découle d'une enquête réalisée sur les prix, la composition et l'éco-étiquetage des vêtements commercialisés par la boutique de mode ZARA. Pour ce faire, nous avons choisi d'utiliser le langage de programmation Python avec la librairie BeautifulSoup pour effectuer du web scraping. Ce choix s'est avéré être le plus adapté pour notre équipe car Python et BeautifulSoup sont souvent utilisés en web scraping puisqu'ils fournissent une combinaison puissante et flexible d'outils pour extraire des données à partir de sites web de manière automatisée.

B. Analyse de données

Statistique sur l'importation mondiale du textile

Selon Hello Planet, pas moins de 70 % des vêtements vendus en France sont fabriqués en Asie où la main-d'œuvre est peu coûteuse [7]. Certains pays de ce continent spécialisés dans la production à grande échelle, tels que la Chine, le Vietnam et l'Inde, sont souvent choisis pour la production de textiles. Cette décision est souvent prise par les marques pour répondre à la demande croissante des consommateurs occidentaux qui cherchent à consommer plus tout en dépensant moins.

Selon Greenpeace en 2021, les cinq principaux exportateurs mondiaux de textile étaient la Chine avec 286 milliards de dollars, suivie du Bangladesh avec 46,2 milliards de dollars, du Vietnam avec 43,7 milliards de dollars, de l'Inde avec 41,4 milliards de dollars et de l'Allemagne avec 40,4 milliards de dollars [8]. Ces pays ont été les plus performants sur le marché mondial du textile en termes d'exportations au cours de cette année-là (**confère Annexe 1**).

Statistique sur l'exportation mondiale du textile

En sens contraire, pour 2021, les cinq principaux importateurs mondiaux de textile étaient les États-Unis avec 135 milliards de dollars, suivis de l'Allemagne avec 66,7 milliards de dollars, du Japon avec 33,7 milliards de dollars, de la France avec 33 milliards de dollars et du Royaume-Uni avec 31,3 milliards de dollars. Ces pays ont été les plus importants acheteurs de textile sur le marché mondial au cours de cette année-là (**confère Annexe 2**).

Dans l'optique d'une meilleure compréhension de notre jeu de données, nous avons effectué une analyse descriptive afin d'extraire les informations clés qu'elle contient.

Les matières premières

Nous avons procédé à l'élaboration d'un graphique en barre dans le but d'identifier les différents types de tissus utilisés pour la confection des articles répertoriés. Il est ressorti que la majorité de ces articles sont conçus à base de polyester. En effet, cette fibre synthétique, dérivée du pétrole, est devenue la matière la plus largement utilisée par l'industrie textile. Par ailleurs, le coton, qui est une matière végétale, occupe la quatrième place des matières les plus employées dans la production des articles.

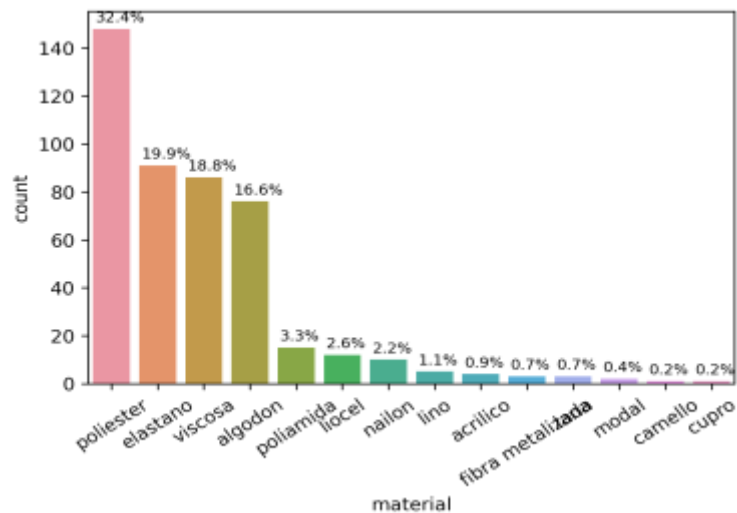


Figure 1 : La distribution des types de matières

Le prix des articles

Le prix des articles est exprimé en centimes d'euros. Il convient de noter que l'un des principaux facteurs qui ont contribué à l'essor de l'industrie de la fast fashion est l'attractivité des prix des articles proposés. Ainsi, afin de mieux comprendre cette relation entre la composition des articles et leur prix, nous avons effectué une visualisation graphique.

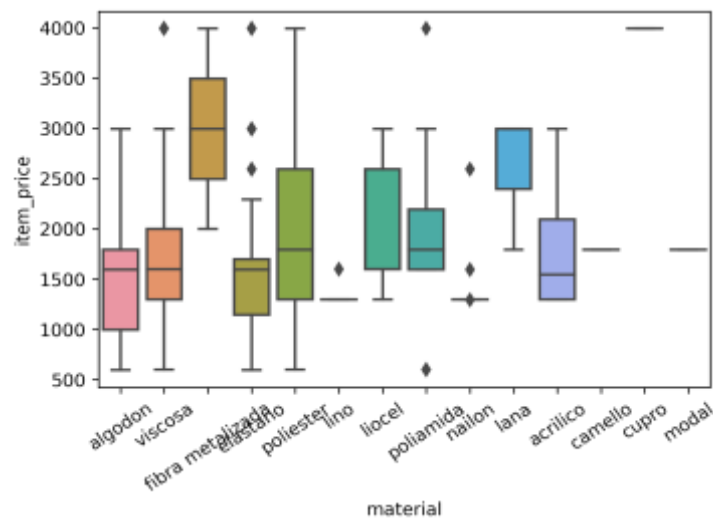


Figure 2 : La relation prix des articles en fonction des types de matières

Compte tenu de la présence de la colonne "item_code" dans notre base de données et de la possibilité pour un article d'être constitué de différents types de tissus, ce qui peut conduire à une répétition du même code dans le dataset, nous avons procédé à une visualisation graphique afin de déterminer le prix moyen des articles contenant un type de tissu spécifique.

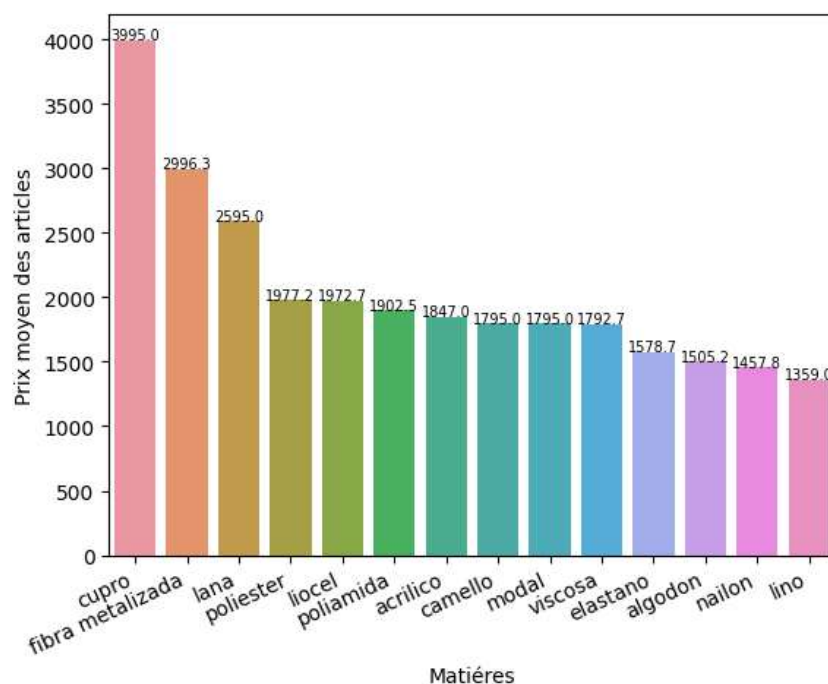


Figure 3: Le prix moyen des articles en fonction des types de matières

Le résultat de notre analyse indique que les articles comportant du cupro ont un prix moyen plus élevé que les autres. Le cupro est une fibre artificielle qui est produite à partir de la cellulose extraite du linter de coton et traitée chimiquement. Le fil obtenu est ensuite utilisé pour le tissage. Cette fibre est douce et délicate, d'où son nom « Cupro ». Par ailleurs, le cupro étant biodégradable, il est important que les teintures textiles utilisées pour le teindre le soient également, ce qui n'est malheureusement pas toujours le cas.

C. Modélisation

Dans le cadre de notre projet, nous avons opté pour l'utilisation de trois modèles d'apprentissage supervisés : SVM (Support Vector Machine ou Machine à vecteurs de support), Random Forest et Régression logistique. L'objectif étant de créer un modèle prédictif permettant de classer les vêtements en différentes catégories, en se basant sur leur prix, leur composition et leurs données d'éco-étiquetage. Le choix de chaque modèle a été effectué en prenant en compte ses avantages et limitations, et en considérant sa pertinence par rapport à notre problématique.

1. Random Forest

Dans le cadre de la problématique de l'impact environnemental et social des entreprises de fast fashion, le modèle Random Forest peut être justifié pour plusieurs raisons. Il est robuste face à des données manquantes ou des variables non pertinentes, ce qui peut être utile dans un domaine où les données peuvent être hétérogènes. Il permet également de combiner les résultats de plusieurs arbres de décision aléatoires pour améliorer la précision de la prédiction, ce qui peut être utile pour prédire notre variable cible (bool, si l'article est éco-étiqueté ou non). En somme, le choix de Random Forest est justifié pour son adaptabilité face à des

données et son aptitude à améliorer la précision de la prédiction tout en évitant le surapprentissage à condition que le modèle soit bien paramétré.

2. SVM

SVM est un modèle d'apprentissage supervisé utilisé pour la classification binaire et multiclasse. Dans notre cas, nous avons une tâche de classification binaire (article étiqueté éco-responsable ou non étiqueté), ainsi SVM est adapté pour prédire la variable 'join_life'. en outre, SVM peut également travailler avec des données non linéaires et de haute dimensionnalité, tout en maintenant la précision de la prédiction.

3. Régression logistique

La régression logistique est un modèle de classification binaire adapté à notre problématique sur l'impact environnemental et social des entreprises de fast fashion. Il permet de prédire si un article est étiqueté éco-responsable ou pas en fonction de certaines variables. La simplicité et la facilité d'interprétation de ce modèle sont des avantages pour expliquer les résultats.

IV. RÉSULTATS

A. Random Forest

Nous avons élaboré deux modèles pour notre algorithme de Random Forest. Toutefois, le premier modèle a présenté un sur apprentissage, ce qui nous a amené à effectuer des ajustements en spécifiant les valeurs des hyperparamètres, afin de réduire la complexité du modèle. Ensuite, nous avons procédé à une validation croisée, une technique statistique utilisée pour estimer la performance prédictive de notre modèle. Les résultats obtenus sont :

Best parameters: {'max_depth': 5, 'min_samples_leaf': 2, 'min_samples_split': 5, 'n_estimators': 50}

Best accuracy: 0.7972602739726027

Accuracy score : 0.7391304347826086

Rapport de classification

	precision	recall	f1-score	support
0	0.74	0.88	0.80	56
1	0.73	0.53	0.61	36
accuracy			0.74	92
macro avg	0.74	0.70	0.71	92
weighted avg	0.74	0.74	0.73	92

Le rapport montre que le modèle est meilleur pour prédire les articles qui ne sont pas éco-étiquetés (0) avec une précision de 74% . Cette facilité à prédire la classe 0 est due d'une part à la grande proportion des articles non étiquetés éco-responsables par rapport à ceux éco-responsables.

Matrice de confusion

Confusion matrix:

[[49 7]

[14 22]]

Le modèle a prédit correctement 49 échantillons appartenant à la classe 0 (pas d'éco-étiquetage) et 14 échantillons appartenant à la classe 1 (avec éco-étiquette).

La courbe de ROC

L'aire de la courbe ROC est 0,83, ce qui indique que le modèle a une bonne capacité à discriminer entre les deux classes.

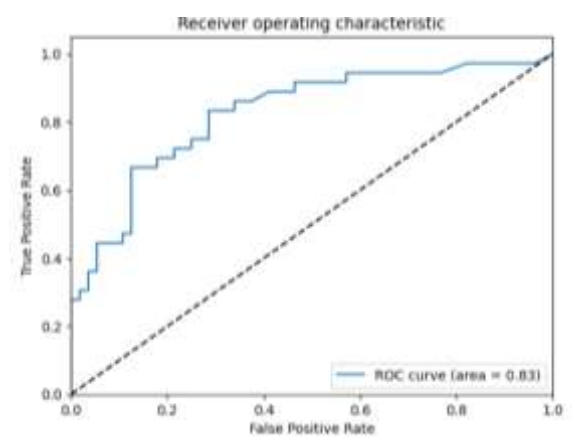


Figure 4: La courbe ROC de la forêt aléatoire

La courbe d'apprentissage

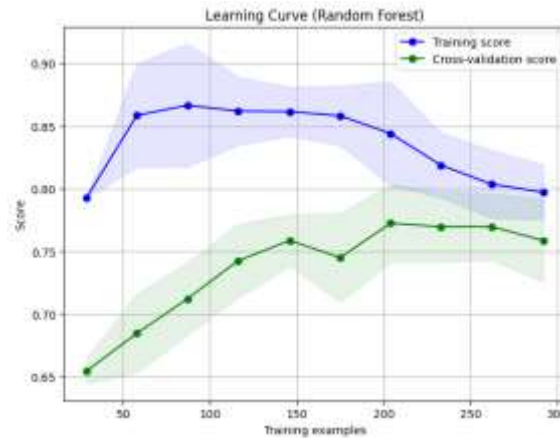


Figure 5: La courbe d'apprentissage de la forêt aléatoire

B. SVM

Nous avons créé deux modèles SVM, étant donné que le premier modèle présentait une suradaptation. Pour résoudre ce problème, nous avons réduit la complexité du modèle en diminuant la valeur de l'hyper paramètre C, qui contrôle la marge, et en utilisant un noyau RBF (Radial Basis Function).

Best parameters: {'C': 1, 'gamma': 100}

Best accuracy: 0.8246575342465754

Accuracy score : 0.782608695652174

Rapport de classification

	precision	recall	f1-score	support
0	0.78	0.89	0.83	56
1	0.79	0.61	0.69	36
accuracy			0.78	92
macro avg	0.78	0.75	0.76	92
weighted avg	0.78	0.78	0.78	92

Le rapport de classification pour le SVM (Support Vector Machine) montre que le modèle a une précision de 0.78, ce qui signifie que sur l'ensemble de données de test, le modèle prédit correctement la classe 78% du temps. Le rappel pour la classe 0 est de 0.89, ce qui signifie que le modèle a correctement identifié 89% des échantillons de la classe 0. Pour la classe 1, le rappel est de 0.61, ce qui signifie que le modèle a correctement identifié seulement 61% des échantillons de la classe 1.

Matrice de confusion

Confusion matrix:

```
[[50 6]
 [14 22]]
```

Dans cette matrice de confusion, il y a 50 vrais positifs et 6 faux négatifs pour la classe 0. Pour la classe 1, il y a 22 vrais négatifs et 14 faux positifs.

La courbe ROC

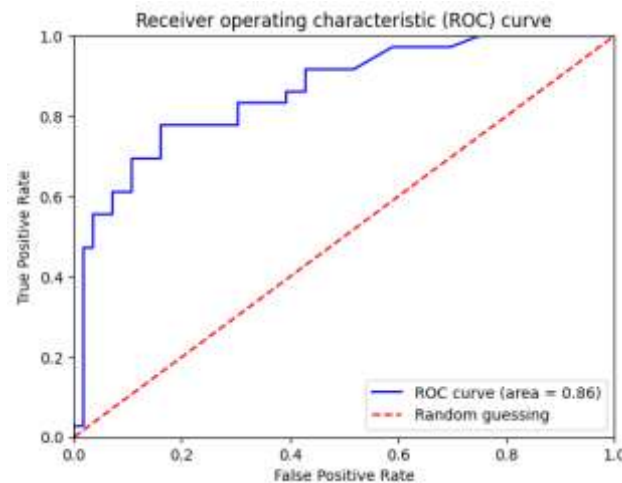


Figure 6: La courbe ROC du SVM

L'aire de la courbe ROC est 0,86, ce qui indique que le modèle a une bonne capacité à discriminer entre les deux classes.

La courbe d'apprentissage

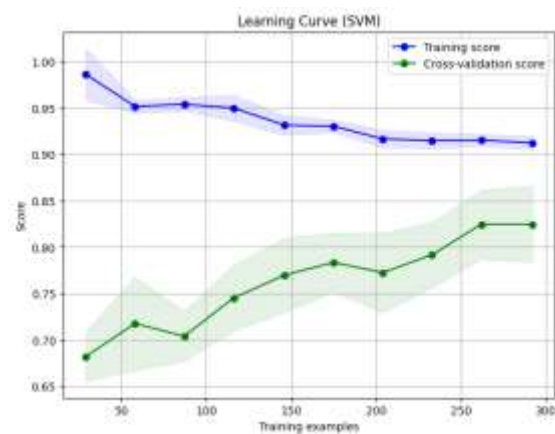


Figure 7: La courbe d'apprentissage de SVM

C. Régression logistique

Best parameters: {'C': 10, 'penalty': 'l2'}

Best accuracy: 0.6986301369863013

Accuracy score : 0.6630434782608695

Le rapport de classification

	precision	recall	f1-score	support
0	0.68	0.86	0.76	56
1	0.62	0.36	0.46	36
accuracy			0.66	92
macro avg	0.65	0.61	0.61	92
weighted avg	0.65	0.66	0.64	92

La précision est de 0,68, ce qui signifie que 68 % des prédictions de la classe 0 étaient correctes. Le rappel est de 0,86, ce qui signifie que le modèle a correctement identifié 86 % des instances de la classe 0. Pour la classe 1 (étiquetée), la précision est de 0,62, ce qui signifie que 62 % des prédictions de la classe 1 étaient correctes. Le rappel est de 0,36, ce qui signifie que le modèle a correctement identifié 36 % des instances de la classe 1.

La matrice de confusion

Confusion matrix:

```
[[48 8]
 [23 13]]
```

La matrice de confusion pour la régression logistique montre que le modèle a correctement identifié 48 vrais négatifs et 13 vrais positifs. Cependant, le modèle a également classé à tort 8 observations négatives comme positives (faux positifs) et 23 observations positives comme négatives (faux négatifs). Cela peut être dû à un déséquilibre des classes

La courbe ROC

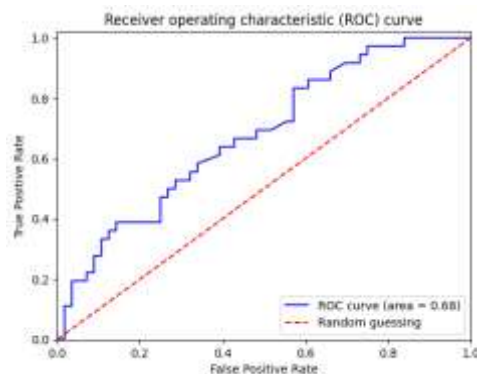


Figure 8: La courbe ROC de la régression logistique

L'aire sous la courbe ROC de la régression logistique est de 0.67, cela signifie que le modèle a une performance moyenne pour la classification binaire.

La courbe d'apprentissage

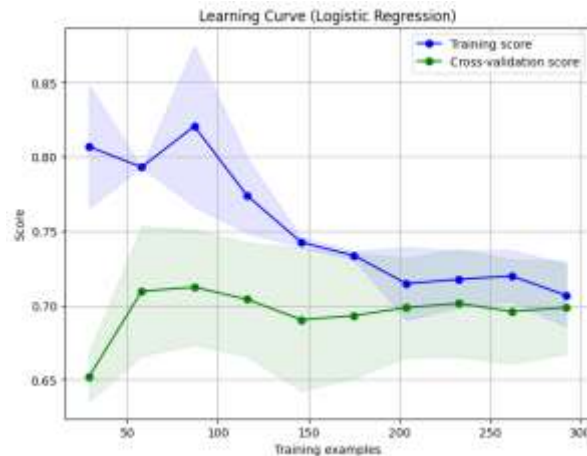


Figure 9: La courbe d'apprentissage de la régression logistique

V. INTERPRÉTATION

Zara Second Life : un concept innovant pour une mode plus durable

Après avoir analysé et modélisé les données sur les matières de Zara, nous avons proposé une solution à notre problématique sous forme d'une plateforme que nous avons nommée "Zara Second Life". Cette plateforme permettra aux clients de Zara de donner leurs vêtements usagés en échange de crédits ou de réductions sur leurs prochaines commandes. Ces crédits ou réductions inciteront davantage les clients à donner leurs vêtements usagés, contribuant ainsi à la réduction des déchets textiles.

Pour que cette initiative soit efficace, nous proposons également la mise en place de points de collecte Zara. En outre, cette dernière pourrait encourager ses clients à utiliser les transports en commun ou les vélos pour se rendre aux points de collecte, afin de réduire leur empreinte carbone.

La plateforme Zara Second Life présente de nombreux avantages pour Zara, l'environnement et les consommateurs de la fast fashion. Tout d'abord, elle permet à Zara de mieux gérer son cycle de production en réutilisant les matières premières de ses anciennes collections. Ensuite, cette initiative réduit l'impact environnemental de l'industrie de la mode, qui est l'une des plus polluantes au monde. Enfin, les clients de Zara bénéficieront de crédits ou de réductions sur leurs achats, ce qui les incitera à acheter des produits durables et écoresponsables.

La plateforme nous l'avons créée avec les langages de programmation Html, Css et Js

En somme, Zara Second Life est une initiative qui peut bénéficier à tous : Zara, l'environnement et les consommateurs.

VI. CONCLUSION

En conclusion, la problématique de l'impact environnemental de l'industrie de la fast fashion peut être résolue en adoptant des initiatives durables telles que la plateforme Zara Second Life. Cette plateforme permet de réduire les déchets textiles, de mieux gérer le cycle de production de Zara, de réduire l'empreinte carbone des clients et de favoriser l'achat de produits durables et écoresponsables. En encourageant les clients à donner leurs vêtements usagés en échange de crédits ou de réductions, cette initiative peut être bénéfique pour toutes les parties impliquées. En somme, Zara Second Life représente une solution innovante et prometteuse pour une mode plus durable.

RÉFÉRENCES

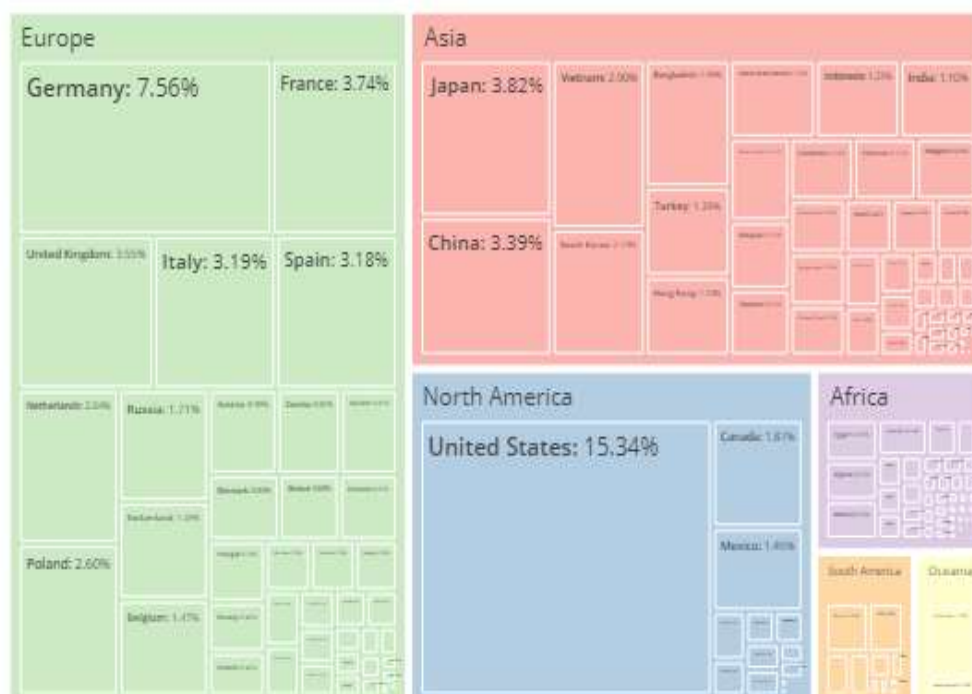
- [1] : [fashionnetwork](#)
- [2] : [WeDressFair](#)
- [3] : [foxintelligence.io](#)
- [4] : [Nos données](#)
- [5] : [Observatory of Economic Complexity](#)
- [6] : [Statista](#)
- [7] : [Hello Planet](#)
- [8] : [Green Peace](#)
- [9] : [Notre Plateforme: Zara Second Life](#)

ANNEXE

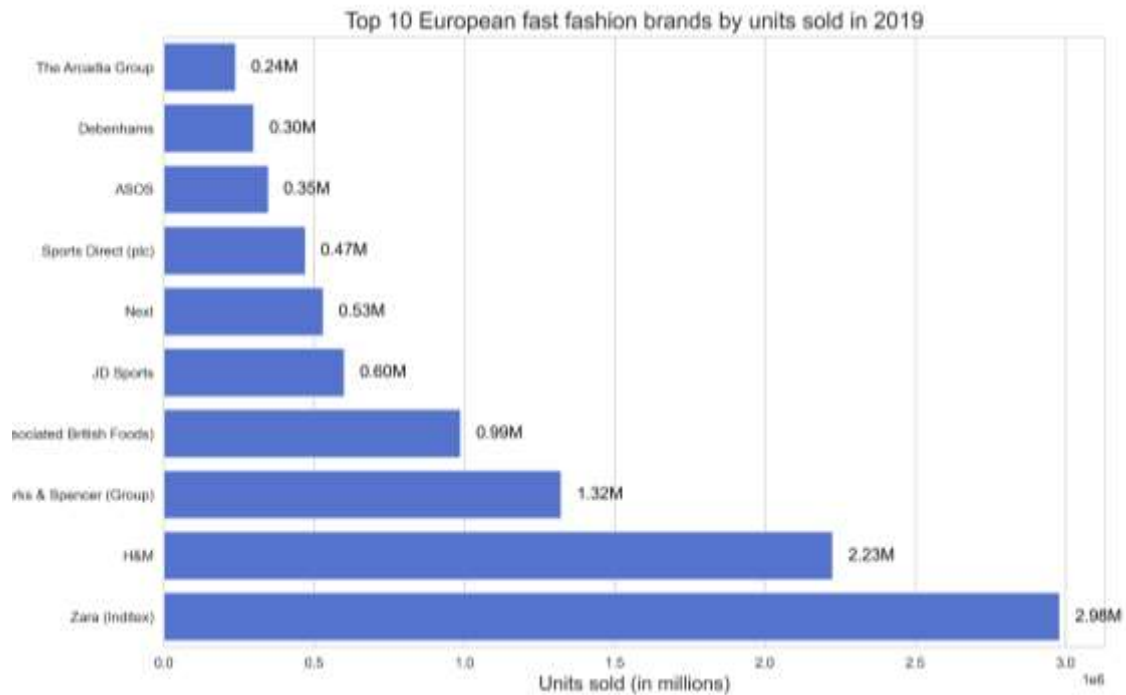
1. ANNEXE 1: Carte proportionnelle sur l'importation du textile en 2021 dans le monde



2. ANNEXE 2: Carte proportionnelle sur l'exportation du textile en 2021 dans le monde



3. ANNEXE 3 : Top 10 des marques européennes de fast fashion sur la base des unités vendues par an dans le monde au cours de l'exercice 2018/2019



4. ANNEXE 4 : Notebook ProjetIA_Datavisualisation

[Diagnostic de l'industrie textile](#)

5. ANNEXE 5 : Notebook FastFashionScraping

[Web Scraping](#)

6. ANNEXE 4 : Notebook ProjetIA_FastFashion_Analyse

[Analyse](#)

7. ANNEXE 4 : Notebook ProjetIA_FastFashion_Modélisation

[Modélisation](#)