

## Impact Textile avec Ecobalyse

**Cursus concerné :** Data Engineer

**Difficulté :** 8/10

**Description détaillée :**

*L'industrie textile est l'une des plus polluantes au monde. Au cœur des préoccupations actuelles, l'objectif de ce projet est d'utiliser l'API d'Ecobalyse, qui permet de calculer les impacts environnementaux des produits textiles, pour créer une application qui aidera les entreprises et les consommateurs à prendre des décisions plus éclairées et durables sur leurs choix de consommation. L'application pourrait, par exemple, permettre aux utilisateurs d'obtenir une évaluation de son impact environnemental basée sur les données d'Ecobalyse et/ou sur des données scrapées. Elle pourrait également fournir des recommandations sur des alternatives plus durables ou des conseils sur la manière de réduire l'impact environnemental d'un produit.*

Etape	Description	Objectif	Modules / Masterclass / Templates	Conditions de validation du projet
1	Récolte des données	<ol style="list-style-type: none"><li>1. L'objectif de cette étape est d'extraire les données nécessaires à partir de l'API d'Ecobalyse. Vous devrez écrire des scripts pour interroger l'API et récupérer les informations pertinentes sur les impacts environnementaux des produits textiles.</li><li>2. Si nécessaire, vous pouvez également utiliser des techniques de web scraping pour compléter les données de l'API comme par exemple des données sur les habitudes de consommations textiles des français.</li></ol>	133 - Web Scraping avec BeautifulSoup	<p>Fichier explicatif du traitement (doc / pdf)</p> <p>Un fichier json d'exemple de récupération.</p>

2	Architecture de la donnée	<p>Il s'agit de choisir la solution de stockage la plus adaptée. Cela pourrait impliquer le choix d'une base de données relationnelle comme PostgreSQL ou une base de données NoSQL comme MongoDB en fonction des besoins spécifiques de votre projet.</p> <p>Vous devrez également définir des schémas de données pour organiser les informations de manière logique et efficace.</p>	<p>142 - SQL</p> <p>(Architecture des données)</p> <p>Elasticsearch</p> <p>MongoDB</p>	<p>Un fichier de requêtes pour montrer que l'architecture choisie est bien fonctionnelle.</p>
3	Consommation de la donnée	<p>Dans cette étape, vous utiliserez les données pour modéliser des techniques de Machine Learning pour prédire l'impact environnemental de produits non couverts par l'API d'Ecobalyse.</p> <p>Vous pourriez également créer un tableau de bord interactif avec Dash pour visualiser les données et permettre aux utilisateurs d'explorer les impacts environnementaux des différents produits.</p>	<p>DE120 - Machine Learning</p> <p>Dash</p>	<p>Notebook de ML</p>
4	Mise en production	<p>Faire une API pour tester le modèle de ML et permettre aux autres applications d'interagir avec vos données.</p> <p>Vous devrez également dockeriser votre projet pour faciliter le déploiement et garantir la reproductibilité de l'environnement d'exécution.</p>	<p>Docker</p> <p>FastAPI</p>	<p>Fichier Yaml du docker-compose</p> <p>API FastAPI</p>
5	Automatisation des flux	<p><b>Étape Facultative</b></p> <p>Il faudra requêter l'API régulièrement via Airflow.</p>	<p>Airflow</p>	<p>Fichier python du DAG</p>
6	Soutenance	<p>Démonstration de votre application et explication de la démarche et du raisonnement effectué lors du projet.</p>	<p>X</p>	<p>Soutenance Rapport</p>

## Bibliographie :

- <https://ecobalyse.beta.gouv.fr/#/>

[DataScientest.com](https://DataScientest.com)

Agrément organisme de formation 11755665975

+09 80 80 79 49

2 place de Barcelone, 75016 Paris



[DataScientest.com](https://DataScientest.com)

Agrément organisme de formation 11755665975

+09 80 80 79 49

2 place de Barcelone, 75016 Paris