

FORMATION : DataScientest / Data Engineer

PROJET (Nov. 2024) : Analyse des Impacts Environnementaux Textiles, avec EcoBalyse v2.4.0

Temps Partiel (9 mois) - `jan24_continu_de / cde_projet_ecobalyse`

Sommaire

1. CONTEXTE

Ecobalyse
GitHub / Equipe

2. SOLUTION

Synoptique
Architecture

3. TECHNIQUE

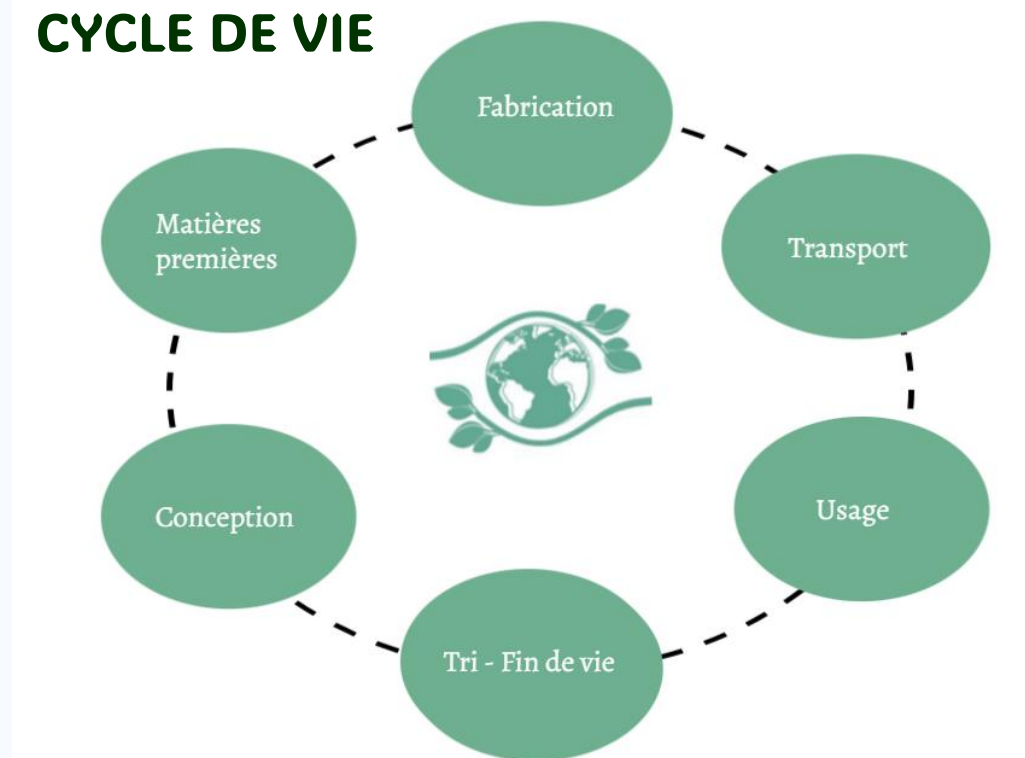
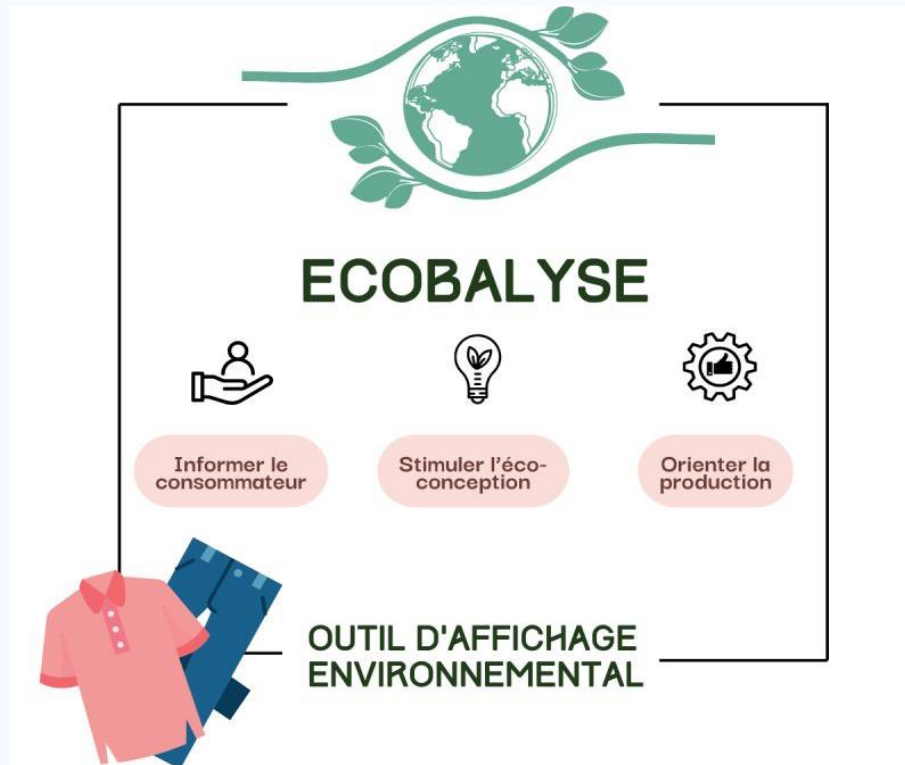
Scripts
Machine Learning

4. DEMONSTRATION

Flask
Dash

1. CONTEXTE

ECOBALYSE: L'outil d'analyse de l'impact des textiles



1. CONTEXTE

API ECOBALYSE

L'API Ecobalyse est un outil en ligne qui permet de calculer l'impact environnemental des produits textiles selon plusieurs critères :

Données d'entrée

→

ACV

→

Eco score

Analyse du Cycle de Vie

Exemples

Tshirt coton (150g) - Majorant par défaut

Mode réglementaire

Mode exploratoire

Catégorie

T-shirt / Polo

Masse du produit fini

0,15 kg

☐ Remanufacturé

Durabilité

Coefficient de durabilité : 0,67

Prix neuf

10 €

Nombre de références

100000

Entreprise

Grande entreprise sans service de réparation

☐ Traçabilité affichée

Matières premières

446 Pts

100 %

Coton

Par défaut (Asie - Pacifique)

408 Pts

Compléments

500 km

+37,50 Pts (+1,94 Pts)

100%

+ Ajouter une matière

1800 Pts

1206 Pts hors durabilité

Pour 0,150 kg

Pour 100g : 1200 Pts

Étapes

Matières premières	36,97%
Transformation	43,95%
Emballage	0%
Transports	6,18%
Distribution	0%
Utilisation	9,45%
Fin de vie	3,46%

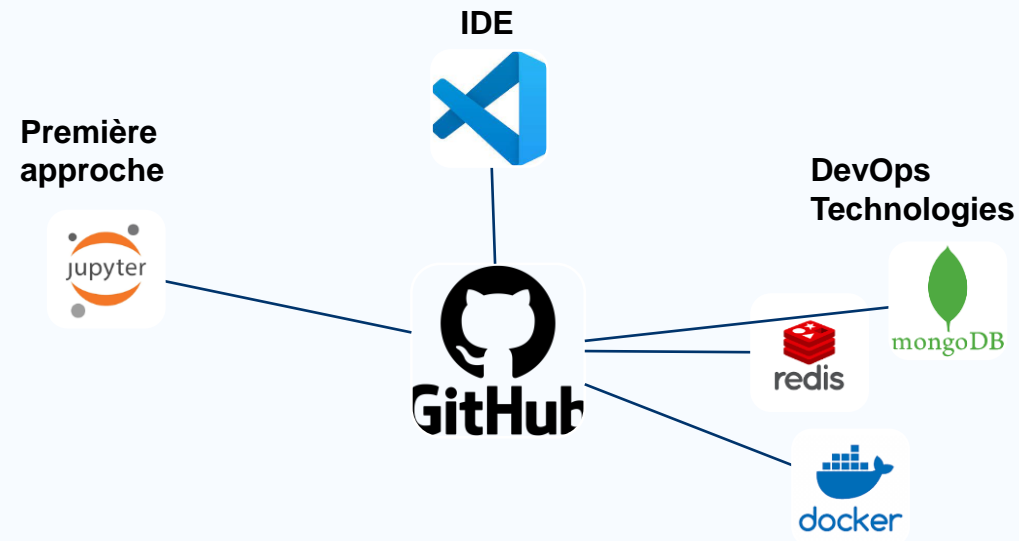
Requêtes API		
GET	/textile/countries	Liste des pays utilisables pour les simulations textiles.
GET	/textile/materials	Liste des matières textile
GET	/textile/products	Liste des types de produits textiles
GET	/textile/simulator	Calcul des impacts environnementaux d'un produit textile
POST	/textile/simulator	Calcul des impacts environnementaux d'un produit textile
GET	/textile/simulator/{impact}	Calcul d'un impact environnemental spécifique d'un produit textile
POST	/textile/simulator/{impact}	Calcul des impacts environnementaux d'un produit textile
GET	/textile/simulator/detailed	Calcul exhaustif des impacts environnementaux d'un produit textile
POST	/textile/simulator/detailed	Calcul des impacts environnementaux d'un produit textile


1. CONTEXTE

PROJET ECOBALYSE DATASCIENTEST

Objectif :

Le projet vise à créer une application utilisant l'API d'Ecobalyse pour évaluer l'impact environnemental des produits textiles



 DataScientest

Datascientest: projet EcoBalyse (Nov. 2024)

Data Engineering End-to-End Project : AirFlow, Dash, Flask, Docker, Redis, MongoDB, Python

Dernière Mise A Jour du Document : Mar. 19/11/2024 - Version : v0.2.0

Sommaire

- [Contexte](#)
- [Présentation](#)
 - [Etapes du projet](#)
- [Mode d'emploi](#)
 - [Pré-requis \(`./info.sh` | `./mode.sh` | `./starter.sh` \)](#)
 - [\(Ré\)Initialiser \(`./init.sh` \)](#)
 - [\(Re\)Configurer \(`./setup.sh` \)](#)
 - [\(Re\)Charger \(`./load.sh` \)](#)
- [Solution technique](#)
 - [Schéma de Principe](#)
 - [Dossiers & Répertoires](#)
 - [Evolutions](#)
- [A propos d'Ecobalyse](#)

Sommaire

1. CONTEXTE

Ecobalyse
Méthodologie / Equipe

2. SOLUTION

Synoptique
Architecture

3. TECHNIQUE

Scripts
Machine Learning

4. DEMONSTRATION

Flask
Dash

2. SOLUTION SYNOPTIQUE

DATASCIENTEST – Projet Ecobalyse (Nov. 2024)

Solution proposée

4

3

2

Ecobalyse

Ecobalyse DB

Ecobalyse API

my_ecobalyse_network
(backend)

MACHINE LEARNING
docker
learn

my_ecobalyse_network
(frontend)

plotly | Dash
Flask

Dashboard

plotly | Dash

ETL

AirFlow

mongoDB

Redis

tshirt

...

jupe

my_ecobalyse_network
(backend)

GitHub Actions

Le projet se déroule en quatre étapes :

- 1. Récolte des données
- 2. Architecture et stockage
- 3. Consommation des données
- 4. Mise en production

2. SOLUTION ARCHITECTURE

L'architecture et l'arborescence du projet sont structurées autour de **différents conteneurs Docker**, afin de faciliter les mises à jour, indépendamment et facilement.

```
prj_ECOBALYSE
├── data
│   ├── mongo
│   ├── redis
│   ├── params01_T-shirt.txt
│   ├── params02_Pull.txt
│   ├── params03_Pantalon.txt
│   ├── params04_Manteau.txt
│   ├── params05_Maillot-de-bain.txt
│   ├── params06_Jupe.txt
│   ├── params07_Jean.txt
│   ├── params08_Chemise.txt
│   ├── params09_Chaussettes.txt
│   ├── params10_Calecon.txt
│   └── params11_Slip.txt
├── dag
├── dash
├── etl
├── flask
├── img
├── logs
├── mongo
├── notebooks
├── redis
├── PRJ-ECOBALYSE-00-FICHE_PROJET.pdf
├── LICENSE
├── CONVENTIONS.md
├── README.md
├── docker-compose.yml
├── lisezMoi.txt
├── info.sh
├── load.sh
├── setup.sh
├── init.sh
└── starter.sh
```

```
etl
├── Dockerfile.etl
├── constants.py
├── extract1.py
├── get_constants.py
├── requirements.txt
├── utils01.py
├── utils02.py
├── utils03.py
```

```
flask
├── Dockerfile.flask
├── constants3.py
├── get_constants3.py
├── init_flask.sh
├── mongo_queries.py
├── redis_queries.py
├── requirements.txt
├── stylesheets
│   ├── listMongoBDD.css
│   └── styles.css
├── templates
│   ├── bienvenue.html
│   ├── index.html
│   └── listMongoBDD.html
├── test_flask.py
└── utils.py
```

```
mongo
├── Dockerfile.mongo
├── constants1.py
├── get_constants1.py
├── init_mongo.js
├── init_mongo.sh
├── mongo.conf
└── test_mongo.py

redis
├── Dockerfile.redis
├── constants2.py
├── get_constants2.py
├── init_redis.sh
├── redis.conf
└── test_redis.py
```

```
dash
├── Dockerfile.dash
├── assets
│   ├── info_countryDyeing.txt
│   ├── info_countryFabric.txt
│   ├── info_countryMaking.txt
│   ├── info_countrySpinning.txt
│   ├── info_masse.txt
│   ├── info_matiere.txt
│   ├── info_mode.txt
│   ├── info_pays.txt
│   ├── info_product.txt
│   ├── linear_regression_model_basic.pkl
│   ├── page0.jpg
│   └── styles.css
├── components.py
├── constants4.py
├── get_constants4.py
├── init_dash.sh
├── old_page13.py
├── page0.py
├── ...
├── page31.py
├── requirements.txt
└── test_dash.py
```



(Reste à faire)

Sommaire

1. CONTEXTE

Ecobalyse
Méthodologie / Equipe

2. SOLUTION

Synoptique
Architecture

3. TECHNIQUE

Scripts
Machine Learning

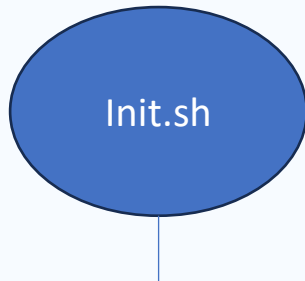
4. DEMONSTRATION

Flask
Dash

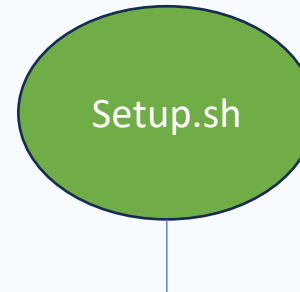
3. TECHNIQUE SCRIPTS

- [\(Ré\)Initialiser \(./init.sh \)](#)
- [\(Re\)Configurer \(./setup.sh \)](#)
- [\(Re\)Charger \(./load.sh \)](#)

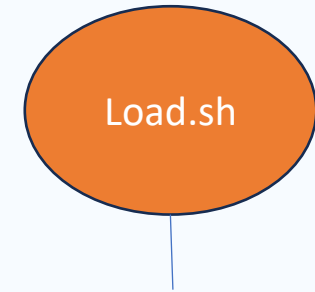
Trois scripts permettent de s'assurer du bon fonctionnement du projet :



- Nettoie l'environnement pour une nouvelle exécution du setup
- Supprime les images des conteneurs préalablement créés
- Supprime les réseaux préalablement créés pour les échanges entre dockers
- Supprime les données préalablement créées lors de l'exécution du setup



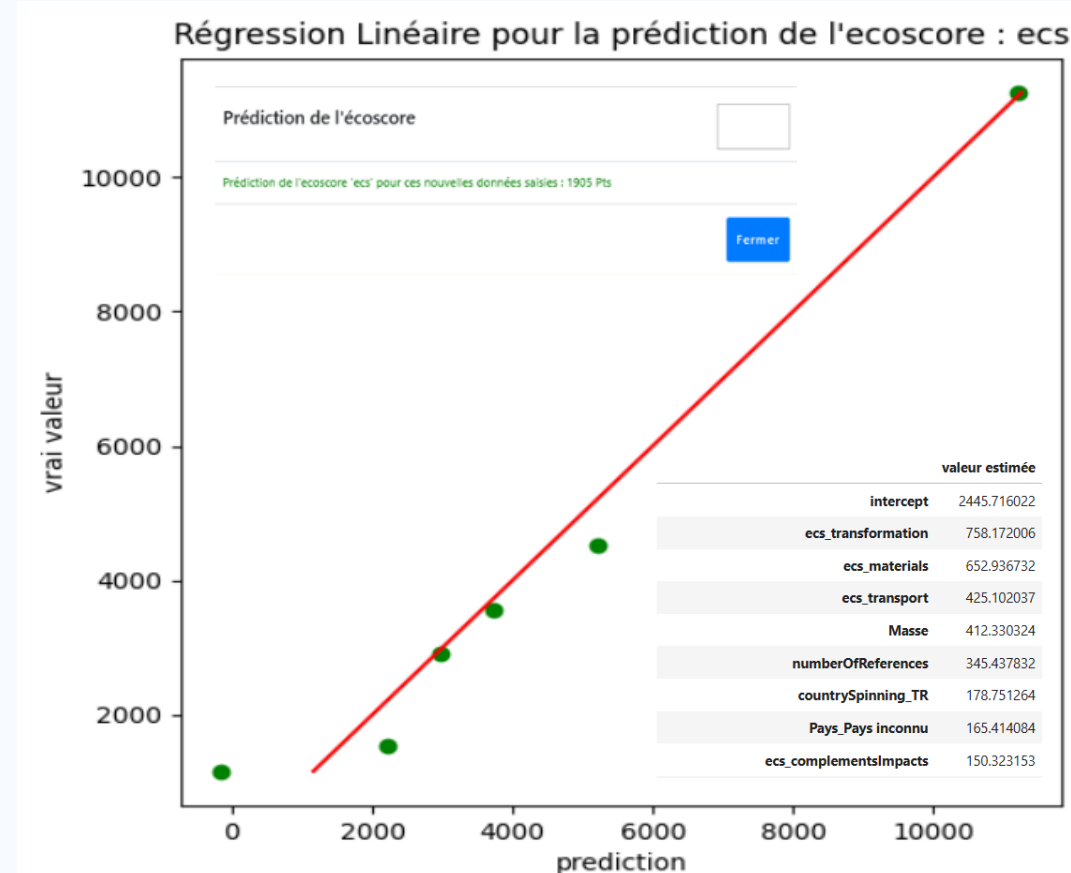
- Supprime les fichiers logs si déjà existants
- Mets en place les conteneurs nécessaires à l'extraction et la transformation des données (*ecblwebscraping* , *ecblmongodb* , *ecblredis*)



- Permet d'avoir une visualisation graphique des données traitées à l'étape précédente via le setup
- Fournit des résultats d'ecoscore basés sur le modèle de Machine Learning choisi pour tester les données fournies en entrée (27 échantillons)

3. TECHNIQUE MACHINE LEARNING

Suite à l'entrainement de plusieurs modèles de Machine Learning, la Régression Linéaire a été retenue, car les résultats fournis sont ceux qui se rapprochent le plus de ceux de [l'API Ecobalyse](#).



Objectif : reproduire le calcul de [l'API Ecobalyse](#), par prédiction

Visualisation Des Données : Listing & Description

Charger les données

T-shirt / Polo

Libelle	Categorie	ecs
T-shirt lin 150g France Mode ethique	T-shirt / Polo	446
T-shirt coton bio 150g France Mode ethique	T-shirt / Polo	507
T-shirt coton 150g France Mode traditionnelle	T-shirt / Polo	751
T-shirt coton 150g Chine Mode fast fashion	T-shirt / Polo	1118
T-shirt synthétique 150g Asie Mode ultra fast fashion	T-shirt / Polo	1547
T-shirt coton 150g Pakistan Mode traditionnelle	T-shirt / Polo	1714
T-shirt coton 150g Majorant par défaut	T-shirt / Polo	1855

Exemples

T-shirt coton (150g) - Majorant par défaut

Mode réglementaire

Mode exploratoire

Catégorie

T-shirt / Polo

Masse du produit fini

0,15 kg

Remanufacturé

Durabilité

Coefficient de durabilité : 0,67

Prix neuf

10 €

Nombre de références

100000

Entreprise

Grande entreprise sans service de réparation

Traçabilité affichée

Matières premières

100 %

Coton

Par défaut (Asie - Pacifique)

446 Pts

408 Pts

+37,50 Pts

1800 Pts

1206 Pts hors durabilité

Pour 0,150kg

Pour 100g : 1200 Pts

Étapes

Étapes	Pourcentage
Matières premières	36,97%
Transformation	43,95%
Emballage	0%
Transports	6,18%
Distribution	0%
Utilisation	9,45%
Fin de vie	3,46%

Sommaire

1. CONTEXTE

Ecobalyse
Méthodologie / Equipe

2. SOLUTION

Synoptique
Architecture

3. TECHNIQUE

Scripts
Machine Learning

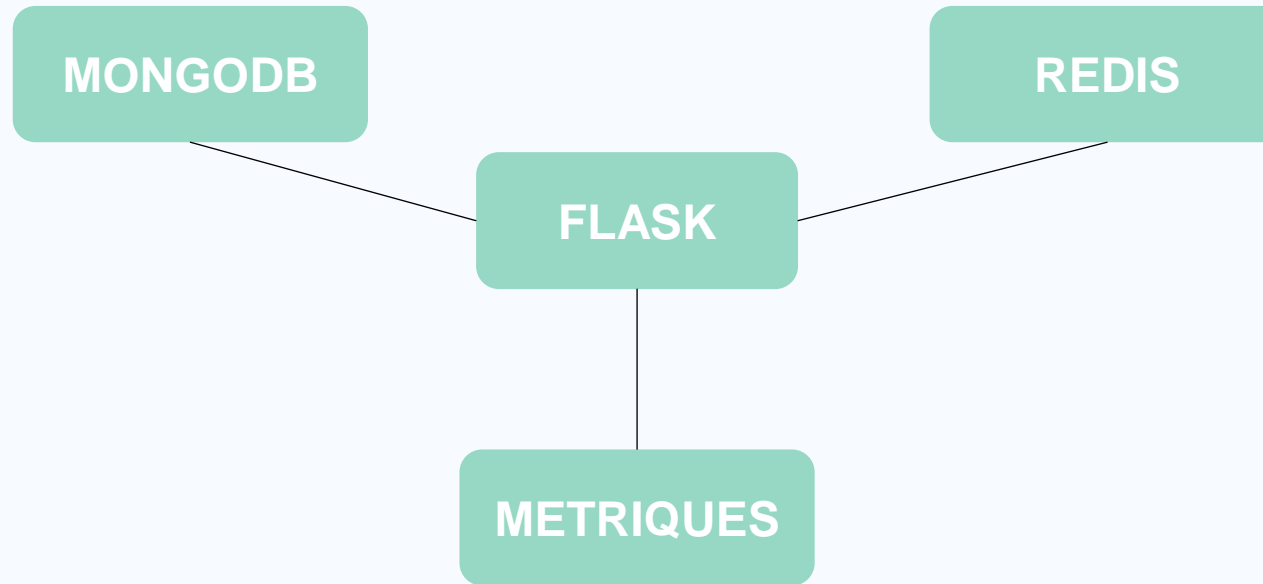
4. DEMONSTRATION

Flask
Dash

4. INTERFACE FLASK (ACCUEIL)

Lancer Flask

- lancer le script `./load.sh -adm` pour lancer Flask via un *navigateur Web*.



Métriques Environnementales

Métriques Économiques

Métriques de Durabilité

Métriques de Transport

Métriques de Traçabilité

Métriques de Transformation

- FLASK permet la communication entre les différents modules du projet et permet d'afficher quelques métriques, telle que :
 - Métriques environnementales : Calcul score pour chaque catégorie de produit, pays, mode
 - Métriques économiques : Calcul le prix moyen pour chaque produit selon la catégorie
 - ...