

Test technique

L'objectif de ce test à réaliser en autonomie est d'évaluer votre niveau sur quatre axes, le machine learning, le code, le traitement de données et le modèle choisi. Pour chaque partie vous trouverez un temps conseillé. Veillez à bien fournir toutes les informations nécessaires à l'analyse, compréhension et exécution de votre code.

I) Machine learning (temps conseillé : 1h)

Objectif : Faire un modèle de classification, expliquer les données et les résultats

Vous trouverez un fichier *dataset.csv* contenant un jeu de données. L'objectif est de construire un modèle qui classe les produits par activité.

Chaque ligne correspond à 1 produit, définit sur 4 caractéristiques : poids, hauteur, largeur, longueur. Un produit appartient à une seule famille : meuble ou MLP (multi-produit) ou déco.

Vous êtes libres de choisir le modèle, le feature engineering à effectuer éventuellement et la variable d'évaluation. **Le code devra être rédigé en python.**

Livrables : notebook/pdf contenant l'analyse du datasets, le modèle choisi et les résultats obtenus. Un regard particulier sera porté sur l'analyse et l'interprétation des résultats du modèle.

II) Mise en production (temps conseillé : 45min)

Objectif : mettre en production dans une API le modèle réalisé précédemment.

Cette partie est découpée en plusieurs sous parties. **L'ensemble du développement doit être fait en python. Vous êtes libres des librairies utilisées.**

2.1) Mise en place de l'api

Mettez en place une API REST qui expose votre modèle. Cette API possèdera une route qui permettra d'appeler le modèle et retournera l'activité. Voici un exemple à titre indicatif :

```
/predict?height=10&width=10&depth=10&weight=10
```

```
-> {"activity": "MLP"}
```

2.2) Package de l'api dans un container

Mettez l'api développée dans un container (idéalement Docker).

2.3) Ajout d'un contrat d'interface open API (bonus)

Ajoutez à l'API déjà développée un contrat d'interface open API qui sera accessible via une url de l'api. Cela permettra d'avoir accès aux différentes routes de l'API ainsi que les arguments et les retours.

Livrables : Ensembles des ressources et codes nécessaires à l'exécution du code de notre côté (*.py, Dockerfile, *.joblib etc.). Un regard particulier sera porté sur la qualité du code et la documentation.

III) SQL (temps conseillé : 20min)

Objectif : Réaliser des requêtes SQL claires et facilement compréhensibles

3.1) Les données

Nous avons les 2 tables suivantes :

TRANSACTIONS

Cette table contient des données transactionnelles. Avec les infos suivantes :

- date : date à laquelle la commande a été passée
- order_id : identifiant unique de la commande
- client_id : identifiant unique du client
- prod_id : identifiant unique du produit acheté
- prod_price : prix unitaire du produit
- prod_qty : quantité de produit achetée

Echantillon de la table TRANSACTIONS :

date	order_id	client_id	prod_id	prod_price	prod_qty
01/01/20	1234	999	490756	50	1
01/01/20	1234	999	389728	3,56	4
01/01/20	3456	845	490756	50	2
01/01/20	3456	845	549380	300	1
01/01/20	3456	845	293718	10	6

PRODUCT NOMENCLATURE

Cette table contient le référentiel produit c'est à dire les méta-données du produit. On y trouve les infos suivantes :

- product_id : identifiant unique du produit
- product_type : type de produit (DECO ou MEUBLE)
- product_name : le nom du produit

Echantillon de la table PRODUCT_NOMENCLATURE :

product_id	product_type	product_name
490756	MEUBLE	Chaise
389728	DECO	Boule de Noël
549380	MEUBLE	Canapé
293718	DECO	Mug

3.2) Première partie du test

Ecrivez une requête SQL simple permettant de trouver le chiffre d'affaires (le montant total des ventes), jour par jour, du 1er janvier 2019 au 31 décembre 2019. Le résultat sera trié sur la date à laquelle la commande a été passée.

3.3) Deuxième partie du test

Réalisez une requête un peu plus complexe qui permet de déterminer, par client et sur la période allant du 1er janvier 2019 au 31 décembre 2019, les ventes meuble et déco réalisées.

Livrable : Les fichiers SQL contenant les requêtes.

IV) Restitution

Vous mettrez l'ensemble des livrables dans un dépôt (gitlab/github...) avec les instructions nécessaires à la compréhension et l'exécution des codes. Si vous souhaitez utiliser un dépôt privé, nous vous communiquerons les utilisateurs à ajouter afin qu'on puisse y accéder.