

PROJET FIL ROUGE

PRÉSENTATION DU CAS D'USAGE



Le projet

Contexte :

- Un e-commerçant, SCE GLOBAL, vient vous voir.
- Il a des données.
- Il ne sait pas comment en tirer de la valeur.
- Il pense que la datascience peut lui permettre de valoriser ses données.

Déroulement :

- Les encadrants vont jouer le rôle de l'e-commerçant tout au long du projet.
- Vous jouer en groupe jouant le rôle d'une start-up/petite équipe spécialisée dans la datascience
- Le jeu de donnée choisi est un jeu de donnée réel d'un e-commerçant et vous est transmis

Le projet

Les objectifs du projet :

- **Effectuer un projet de datascience complet en groupe.**
- **Interagir avec le métier**
- **Comprendre les besoins métiers**
- **Réaliser une(s) solution(s) apportant de la valeur au métier**

Déroulement du projet

- Constituez 8 groupes **maximum**
- Durant chaque séance, entretien de **15mn par groupe** avec l'encadrant qui jouera le rôle du métier.
- Soutenance orale la **dernière journée du module** (14 décembre). Vous aurez 20 minutes de présentations, suivis de 10 minutes de questions. Les questions seront "techniques" (sur les choix techniques posés par un datascientiste) et "métier" (posées par le e-commerçant ne comprenant pas forcément la datascience).

DEADLINES:

- Slides avant le 15 décembre 23h59. Les premiers seront les derniers à passer ... :)
- Le code n'est pas à rendre. Une démonstration lors de la présentation n'est pas obligatoire (mais peut améliorer celle-ci).

Les livrables

Les slides de présentation au format pdf.

- Ils doivent expliquer la démarche de l'équipe, justifier les choix et pistes explorée, et démontrer la valeur apportée au métier.

Si vous rendez une maquette de la solution réalisée.

- En python 3.8
- Elle doit contenir un fichier readme.md expliquant les fichiers fournis et comment utiliser la solution.
- Elle doit contenir un fichier requirements.txt détaillant l'ensemble des bibliothèques python nécessaires et leur versions
- Elle doit fonctionner en deux instructions à partir d'un nouvel environnement python 3.8 :

```
pip install -r requirements.txt (ou conda)  
python3 demo.py --data=DATA_PATH
```

DATA_PATH est le chemin du dossier contenant l'ensemble des csv fournis pour le POC **tels que fournis** (sans modifications)

- Tout argument supplémentaire est optionnel et doit être mentionné dans le readme.

Nos conseils

Appliquez ce que l'on voit dans ce module pour être capable de faire une prédiction

Trouver le ou les bonnes applications de la datascience peut être plus déterminant que l'optimisation de votre solution.

- Ne pas foncer "tête baissée" sur le premier use case mais prendre le temps de définir le ou les uses cases les plus **intéressants** (valeur apportée au métier, **faible risque de ne pas marcher d'ici la fin du projet**, ...)
- Vous pouvez constituer des sous-groupes responsables de différents uses cases si ceux-ci sont peu dépendants.
- L'utilisation du machine learning n'est pas obligatoire pour apporter de la valeur au métier.

Le projet est noté sur :

- La valeur apportée au métier (1/3 de la note)
- La présentation : qualité, clarté et à quelle point elle aurait convaincu le client de choisir votre solution (1/3 de la note).
- La qualité du travail de datascience effectué (1/3 de la note)

Ces points seront évalués lors de la présentation finale **mais aussi lors des entretiens !**



SCALIAN

A BRIDGE TO PERFORMANCE

Société SCALIAN

417 L'Occitane - CS77679

31676 Labège Cedex

Tél. : +33 (0)5 61 00 79 79

scalian.com