Sujet :

Détail:

TP Noté : À partir d'un jeu de données fourni vous devrez par l'intermédiaire de Scikit Learn ou de Dataiku réaliser la meilleure prédiction possible.
L'utilisation et la comparaison d'au moins 3 différents modèles de Machine Learning (régressions, modèle ensembliste) est attendue Vous devrez en amont appliquer plusieurs étapes de data préparation (data cleaning, feature engineering) et de data visualisations.
Vous devrez comparer l'évaluation des différents modèles entraînées.

Equipe:

Vous pouvez travailler seul ou en équipe (2 conseillé, 3 maximum)

Contraintes:

Vous disposez d'1,5 jours de temps réservé pour réaliser ce projet.

Rendu:

Si vous choisissez de travailler directement en python nous attendons un document de tpe notebook commenté et illustré (graphique seaborn).
Pour les utilisateurs de Dataiku vous devrez fournir un document pdf expliquant votre démarche, agrémenté de screenshots du travail réalisé sur Dataiku (graphique, explication de vos étapes de data preparation, etc.)

Enfin, quelque soit les solutions que vous utiliserez un screenshot de votre note kaggle est attendu.

2 sujets proposés (au choix) :

régression :
Apprenez à prédire le prix de biens immobiliers !
https://www.kaggle.com/c/house-prices-advanced-regression-techniques

Classification:
Catégorisez les crimes de San Francisco
https://www.kaggle.com/c/sf-crime