

Mini projet

1. Dates du rendu

La date limite est le 5 avril pour rendre notebook

2. Le projet

Le projet sera centré sur l'analyse d'un jeu de données préparé à cet effet. Le choix du sujet est libre et vous trouverez une liste de suggestions ci-dessous.

Sources de données

1. Kaggle Datasets
2. UCI Machine Learning Repository
4. Open data de la ville de Paris
5. <https://data.fivethirtyeight.com/>
6. <https://data.worldbank.org/>
7. <https://www.fema.gov/fr>
8. <https://datasetsearch.research.google.com/>
9. <https://www.microsoft.com/en-us/research/tools/>
10. <https://registry.opendata.aws/>
11. <http://millionsongdataset.com/>
7. Etc.

Attention : Le site Kaggle référence de nombreuses compétitions intéressantes. Toutefois, avant d'utiliser les données Kaggle, il est important de lire les articles « use for teaching after competition concludes » et « Using a Kaggle contest as a term project ». Les règles peuvent varier d'un projet à l'autre, prenez soin de les lire avant de choisir un projet car on ne peut pas tout faire avec les données disponibles sur ce site.

3. Travail attendu

On vous demande de vous poser une problématique et d'y répondre en utilisant des données. Le projet est orienté analyse prédictive des données. Les critères d'évaluations sont les suivantes :

1. Notebook réalisé

2. Une introduction : quel est le problème à résoudre, les questions auxquelles on cherche à répondre, un aperçu des données.
3. Prétraitement de données
4. Présence de graphique
5. Comparaison entre plusieurs modèles prédictifs avec justification (au moins 3 modèles)
6. Présentation du raisonnement et de la démarche

4. Barème

1. Les notebooks + démarche scientifique = 15 points
2. Graphique = 5 points