

Travail sur le dataset du Titanic :

Pour votre premier projet en autonomie de Machine learning, vous allez travailler avec le fameux dataset du Titanic! C'est un peu le passage obligé de chaque étudiant en data science et souvent le premier vrai projet !

Dans ce projet on doit prédire si un passager en fonction de ces caractéristiques serait parmi les survivants ou les morts.

Pour ce projet vous pouvez faire des recherches sur ce dataset afin de mieux comprendre les données. L'important est que vous compreniez ce que vous faites et de comprendre le principe de l'algorithme utilisé ici une régression logistique .

Il faudra faire de l'exploration de vos données avec pandas et de la data visualisation.

Visualiser les variables.

Faire du nettoyage de données,

Ensuite travailler sur le modèle de machine learning.

Faire des recherches sur la classification et la régression logistique et comment traiter ce problème en machine learning.

Comprendre les metrics pour la classification : l'accuracy , la precision , le recall et le f1_score

Ce projet sera une manière de voir ce que vous avez retenu et ce que vous êtes capable de faire sans être guidé par des questions mais par votre curiosité, votre intuition et également par vos recherches sur le sujet.

Le travail sera à mettre sur le drive avant la fin de la dernière séance avec vos noms, cela me permettra d'évaluer un peu votre niveau à la fin de ces 3 jours de formations.

Certains d'entre vous feront une restitution de leurs résultats la prochaine séance afin que nous puissions en discuter ensemble.