

**Projet :** Analyse de données en apprentissage automatique pour résoudre des problèmes de classification et de clustering dans le domaine de la santé / de la finance / des risques liés à l'agriculture.

### **Introduction :**

Dans ce projet, vous allez travailler en équipe de deux à trois étudiants pour analyser des données dans le domaine de la santé / de la finance / des risques liés à l'agriculture en utilisant des techniques d'apprentissage automatique. Vous allez explorer une variété de techniques de classification et de clustering pour résoudre des problèmes spécifiques dans ces domaines. Les étapes principales du projet sont la recherche d'une dataset, l'analyse exploratoire des données (EDA), la construction de modèles d'apprentissage automatique, l'évaluation des performances du modèle et la présentation des résultats.

### **Objectifs :**

- Identifier une dataset riche en fonctionnalités dans le domaine de la santé / de la finance / des risques liés à l'agriculture
- Effectuer une analyse exploratoire des données (EDA) pour comprendre les variables et les distributions des données
- Construire des modèles de classification et de clustering en utilisant des techniques d'apprentissage automatique
- Évaluer les performances du modèle en utilisant des métriques appropriées
- Interpréter les résultats et proposer des améliorations pour améliorer les performances du modèle

### **Dataset :**

Vous pouvez trouver des datasets dans le domaine de la santé / de la finance / des risques liés à l'agriculture sur plusieurs sources en ligne. Voici quelques suggestions :

Kaggle : <https://www.kaggle.com/datasets>

UCI Machine Learning Repository : <https://archive.ics.uci.edu/ml/index.php>

OpenML : <https://www.openml.org/search?type=data>

### **Analyse exploratoire des données :**

Avant de commencer à construire des modèles, vous devez effectuer une analyse exploratoire des données pour comprendre les données. Vous pouvez utiliser des bibliothèques Python telles que Pandas, Numpy, Seaborn et Matplotlib pour visualiser les données et effectuer des statistiques descriptives. Voici un lien utile pour apprendre l'analyse exploratoire des données : Datacamp : <https://www.datacamp.com/community/tutorials/exploratory-data-analysis-python>

### **Modèles de classification et de clustering :**

Vous pouvez utiliser une variété d'algorithmes de classification et de clustering pour résoudre les problèmes dans le domaine de la santé / de la finance / des risques liés à l'agriculture.

### **Évaluation des performances du modèle :**

Après avoir construit des modèles, vous devez évaluer leurs performances en utilisant des métriques appropriées telles que la précision, le rappel et la F-mesure. Vous pouvez utiliser des bibliothèques Python telles que Scikit-Learn pour évaluer les performances du modèle. Voici un lien utile pour apprendre comment évaluer les performances du modèle :

Scikit-Learn : [https://scikit-learn.org/stable/modules/model\\_evaluation.html](https://scikit-learn.org/stable/modules/model_evaluation.html)

**Interprétation des résultats :**

Après avoir évalué les performances du modèle, vous devez interpréter les résultats et proposer des améliorations pour améliorer les performances.