

## **Master 1 Informatique: Techniques d'apprentissage artificiel**

Dans le cadre de l'évaluation du module « Techniques d'apprentissage artificiel », vous êtes invité à réaliser un projet complet d'apprentissage automatique en appliquant les notions abordées en cours.

**Le projet peut être mené individuellement ou en binôme et développé en python.**

Explorer, analyser et modéliser des données en utilisant des méthodes d'apprentissage automatique (text mining, classification, régression, segmentation, etc.) afin de répondre à une problématique concrète.

Pour ce faire, vous devrez accomplir les tâches suivantes :

1. Définir la problématique et les objectifs.
2. Identifier et collecter les données d'apprentissage.  
[Hugging Face Datasets](#)  
[Kaggle Datasets](#)  
[UCI Machine Learning Repository](#)  
[Google Dataset Search](#)
3. Préparer et diviser les données en ensembles d'entraînement, de test et de validation.
4. Choisir et implémenter des méthodes d'apprentissage adaptées à la problématique (arbres de décision, Random Forests, K-Means, CAH, etc.).
5. Évaluer les performances des modèles à l'aide de métriques pertinentes : accuracy, precision, recall, f-measure, courbe ROC, AUC, etc.
6. Discuter des résultats et proposer des pistes d'amélioration.

### **§ Quelques idées de projets**

- Détection de faux avis en ligne à partir de textes
- Classification de maladies à partir de données médicales.
- Analyse de sentiments de tweets autour d'un sujet d'actualité.
- Prédiction du prix de l'immobilier ou du risque de défaut de crédit.
- Segmentation de clients ou d'utilisateurs pour une campagne marketing.
- Reconnaissance d'émotions à partir des avis

- Recommandations de films ou produits.

### **§ Sur la forme :**

- Présentation de 10 minutes suivie de 5 minutes de questions.
- Rapport de présentation résumant le travail réalisé (contexte, problématique, livraison du projet (fichier d'apprentissage, fichier de test, readme, modèles, données, ...)) Exemple sur ce lien: <https://drivendata.github.io/cookiecutter-data-science/#nothing-here-is-binding>.

Ce travail est à rendre sur l'espace Moodle le 4 décembre avant minuit.

La soutenance est prévue pour le 10 décembre de 12h à 18h.