

SY09 — P2019

Projet

1 Règles du jeu

Principe Le projet consiste à effectuer une analyse d'un jeu de données. On veillera à déployer un processus d'analyse et de traitement rigoureux. L'accent pourra être mis plus particulièrement sur un aspect (on pourra par exemple déployer une méthode « avancée » de clustering ou de discrimination).

En particulier, les données devront être présentées (analyse exploratoire) avant d'être traitées. Les pré-traitements devront être clairement détaillés et motivés. La mise en œuvre d'un algorithme ou d'une méthode n'ayant pas fait l'objet d'une étude dans le cadre de l'UV devra s'accompagner d'une présentation claire et suffisamment détaillée.

Données L'équipe enseignante propose un certain nombre de jeux de données et de méthodes « avancées » à appliquer. Il est possible de choisir soi-même ses données ou une méthode que l'on souhaite approfondir, à condition que cela n'ait pas déjà été fait par un autre groupe. Si le projet sera réalisé en dehors des séances de TD, il sera néanmoins possible de solliciter les membres de l'équipe enseignante, qui pourront apporter de l'aide dans la mesure du raisonnable.

Selon les données, des informations ont pu être rendues disponibles en nombre plus ou moins important. Nous soulignons ici que c'est *le travail réalisé, autant que le résultat*, qui sera évalué et noté. Ces informations disponibles ne devraient pas être ignorées ; il sera néanmoins primordial d'inscrire la démarche adoptée et de positionner le travail accompli par rapport à ces informations. Le plagiat (réappropriation trompeuse du travail ou des idées d'autrui) sera impitoyablement sanctionné.

La liste des jeux de données est disponible à la fin de ce document. Chaque jeu de données ne pourra être traité que par un seul trinôme dans chaque TD.

Groupes Le travail sur le projet sera à réaliser en trinômes, tirés au hasard (à l'exception de l'un des trinômes, qui sera constitué de trois étudiants appartenant à des TD différents, et pour lequel les enseignants feront appel aux bonnes volontés ; le projet étant à réaliser en dehors des séances de TD, l'appartenance à un trinôme inter-TD ne sera pas pénalisante).

2 Évaluation

Rendus Les résultats seront présentés par chaque trinôme dans un rapport (dont les règles de présentation font l'objet d'un document séparé, qui sera mis en ligne sur la page Moodle de l'UV) et au cours d'une soutenance.

Le rapport sera à rendre avant la soutenance (date à préciser). Les soutenances auront lieu la dernière semaine avant les finaux, aux horaires des TD et du cours. L'utilisation d'un créneau supplémentaire est actuellement envisagée. La soutenance durera entre 15 et 20 minutes par trinôme, et comprendra une partie d'exposé et une partie de questions.

Notes Une note sera attribuée à chaque étudiant après la soutenance. Elle tiendra compte du volume et de la difficulté du travail effectué ; elle pourra être personnalisée à chaque membre d'un trinôme, en fonction du travail qu'il aura réalisé et de sa performance lors de la soutenance.

3 Liste des jeux de données

Les données à traiter sont disponibles sur Kaggle.

1. Graduate admissions : [données](#)
2. Students Performance in Exams : [données](#)
3. Top 250 Football transfers : [données](#)
4. Credit Card Fraud Detection : [données](#)
5. Black Friday : [données](#)
6. Floodlight MS Dataset : [données](#)
7. Brewer's Friend Beer Recipes : [données](#)
8. Predict FIFA 2018 Man of the Match : [données](#)
9. Pokemon – Weedle's cave : [données](#)
10. Red Wine Quality : [données](#)
11. Biomechanical features of orthopedic patients : [données](#)
12. Deodorant Instant Liking Data : [données](#)
13. Bundesliga Results 1993-2018 : [données](#)
14. Financial Distress Prediction : [données](#)
15. Classify gestures by reading muscle activity : [données](#)
16. An Open Dataset for Human Activity Analysis : [données](#)
17. Network Intrusion Detection : [données](#)