

# Projet de SY09, semestre de printemps 2020

6 avril 2020

## Résumé

Vous avez un projet à réaliser au cours de l'UV, en trinômes, qui a une double vocation pédagogique et d'évaluation. Il a pour objectif d'appliquer les méthodes étudiées au cours du semestre sur des jeux de données réelles. Vous rendrez compte du travail effectué et de la démarche adoptée au travers d'un compte-rendu puis d'une soutenance.

Les projets seront réalisés en trinômes, tirés au hasard dans chaque TD, à l'exception d'un trinôme inter-TD pour lequel les volontaires sont appelés à se manifester. Le travail réalisé sera évalué au travers d'un compte-rendu et d'une soutenance.

Les compte-rendus, de douze pages maximum, seront rédigés en L<sup>A</sup>T<sub>E</sub>X — voir le document dédié — et rendus sur le Moodle *dans les temps impartis* (les retards sont pénalisés). Les soutenances seront organisées à la fin du semestre, lors de la dernière semaine de cours.

## 1 Objectifs du projet

### 1.1 Objectifs

Le projet consiste à effectuer une analyse d'un jeu de données, en veillant à déployer un processus d'analyse et de traitement rigoureux. On pourra appliquer les méthodes vues en cours, ou bien mettre plus particulièrement l'accent sur un aspect (par exemple en déployant une méthode « avancée ») : vous êtes libres, tant que vous assumez vos choix.

Nous accorderons de l'importance à la démarche adoptée : les données devront être présentées (analyse exploratoire) avant d'être traitées. Les pré-traitements éventuels devront être clairement détaillés et motivés. La mise en œuvre d'un algorithme ou d'une méthode pas étudiée en SY09 devra s'accompagner d'une présentation claire et suffisamment détaillée.

Il est préférable de faire un travail solide et complet à partir du contenu de SY09, que d'appliquer de nouvelles méthodes en omettant des étapes importantes et de rendre un travail incomplet ou bancal.

### 1.2 Données

Chaque groupe se verra attribuer un jeu de données parmi ceux proposés par l'équipe enseignante. Il est possible pour un groupe de proposer ses propres données, mais la décision d'accepter ou non nous reviendra. Le projet doit être réalisé en parallèle des TD, mais vous pourrez néanmoins solliciter les enseignants, qui vous apporteront leur aide (dans la mesure du raisonnable).

Selon les jeux de données, le travail sera plus ou moins aisé, et les informations disponibles (sur le web) pourront être plus ou moins abondantes. C'est le travail réalisé, plus que le résultat, qui sera évalué et noté. Si les données sont difficiles (pas de structure de classe, problème de prédiction difficile nécessitant des outils complexes, etc), il est préférable de présenter des résultats mitigés et d'expliquer pourquoi, plutôt que de présenter des résultats miraculeux qu'on ne peut pas expliquer ou obtenus à l'aide d'approches (traitements ou méthodes) mal comprises ou pas maîtrisées.

En cas de difficultés à traiter les données qui vous auront été attribuées, vous aurez un délai pour nous demander d'en changer (mais le choix des données de remplacement nous reviendra de toute manière).

### 1.3 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).

Si vous rencontrez des difficultés à travailler avec les membres de votre trinôme, par exemple parce qu'il ne répond pas à vos sollicitations (après un nombre raisonnable de tentatives de contact et un temps raisonnablement suffisant pour lui permettre de répondre), vous êtes priés d'en faire part à votre chargé de TD. Le cas échéant, faites-le le plus rapidement possible, pour que nous trouvions des solutions si possible.

## 2 Évaluation

### 2.1 Livrables

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.

Les rapports devront être rendus dix jours avant les soutenances (date à préciser). Les soutenances auront lieu la dernière semaine avant les finaux, sur les créneaux 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 un exposé et une période de questions.

### 2.2 Soutenances

Nous rappelons ici que les soutenances, lors desquelles vous devez défendre votre travail et vos choix, nous permettent d'apprécier assez précisément l'implication et le travail de chacun. L'objectif est de voir dans quelle mesure vous êtes capable de défendre le travail et les choix réalisés. La séance de questions qui suit la présentation peut être assez intense, et donner lieu à des questions techniques ou scientifiques pointues.

Cet exercice est difficile, et il nécessite d'avoir les idées très claires quant aux choix stratégiques (techniques et scientifiques) faits par le trinôme ainsi qu'une très bonne connaissance *globale* du travail réalisé. L'objectif n'est en aucun cas d'être déplaisants ou humiliants (on n'est pas là pour « casser »), mais de voir dans quelle mesure chacun s'est impliqué dans la réalisation du projet et approprié le travail du groupe, *dans son intégralité*.

Pour être encore plus explicites, il est nécessaire de pouvoir expliquer et défendre les choix faits, quand bien même les tâches seraient réparties au sein du trinôme.

### 2.3 Plagiat

Les travaux antérieurs peuvent être pris en compte, à condition que vous positionniez votre travail par rapport à eux. Vous n'en restez pas moins notés sur le travail que vous avez réalisé. Il n'est pas interdit de travailler à plusieurs groupes, tant que les contributions de chacun restent clairement identifiées. Tout plagiat (ré-appropriation trompeuse du travail d'autrui, qu'il s'agisse d'un travail de l'année en cours, de travaux antérieurs, ou de travaux réalisés à l'extérieur de l'UTC) sera impitoyablement sanctionné.

### 2.4 Notes

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

Nous ferons tout notre possible pour adopter la notation la plus juste possible, de manière à tenir compte de l'ampleur du travail réalisé, de la difficulté des données, mais aussi de l'implication de chacun.

## 3 Liste des jeux de données

Les jeux de données suivants sont disponibles dans la base de données de l'UCI ou sur Kaggle :

- Census income : [données](#)
- Classification of satellite images : [données](#)
- Default of credit card clients : [données](#)
- Drug consumption : [données](#)
- QSAR oral toxicity : [données](#)
- Students performance in exams : [données](#)
- Brewer's friend beer recipes : [données](#)
- Pokemon – Weedle's cave : [données](#)
- Red wine quality : [données](#)
- Biomechanical features of orthopedic patients : [données](#)
- An open dataset for human activity analysis : [données](#)
- Network intrusion detection : [données](#)

Le jeu de données suivant est disponible sur la page Moodle de l'UV :

- Recettes