

Projets de SY09, semestre de printemps 2023

27/03/2023

Résumé

Vous avez un projet à réaliser au cours de l'UV, en binômes, 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. Les projets seront réalisés en binômes, déterminés au hasard dans chaque TD. Le travail réalisé sera évalué par un compte-rendu et une soutenance (séance de questions-réponses sur la base du rapport). Les compte-rendus, de six pages maximum, seront rédigés en \LaTeX (voir le document dédié) et rendus sur le Moodle *dans les temps impartis*. Une séance de revue à mi-parcours sera organisée en TD.

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. Vous pourrez appliquer les méthodes vues en cours, ou bien approfondir 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 traitera un jeu de données. L'équipe enseignante en propose plusieurs. Il est possible qu'un groupe en propose un de son choix, mais la décision

d'accepter ou non revient aux enseignants. Le projet doit être réalisé en parallèle des TD, mais les groupes peuvent néanmoins solliciter les enseignants, qui apporteront leur aide (dans la mesure du raisonnable).

Selon les jeux de données, le travail est plus ou moins aisé, et les informations disponibles (web, etc) sont plus ou moins abondantes. C'est le travail réalisé, plus que le résultat, qui est évalué et noté. Si les données sont difficiles (pas de structure de classe, problème nécessitant des classifieurs complexes, etc), il est préférable de présenter et d'expliquer des résultats mitigés, plutôt que d'exhiber des résultats miraculeux sans explications ou d'utiliser des approches mal comprises.

Chaque groupe choisira un jeu de données *spécifique* : deux groupes ne peuvent choisir le même. En cas de difficultés, vous pouvez demander à changer de données, demande soumise à l'approbation des enseignants. Compte-tenu du calendrier (le semestre passe très vite), une telle demande doit être faite rapidement.

1.3 Groupes

Le travail sur le projet sera à réaliser en binômes. Les étudiants seront répartis au hasard en groupes. Nous veillerons à ce qu'ils soient équilibrés (compétences en informatique, bonne connaissance du français).

Si vous rencontrez des difficultés à travailler avec votre binôme, 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 cette démarche le plus rapidement possible, pour nous permettre de trouver une solution rapidement.

1.4 Choix des données

Les groupes sont rendus publics la semaine du 27/03. Le choix des données doit être fait (et indiqué sur le Moodle de SY09) avant le 21/04. Le temps passe très vite : travaillez à vos projets dès votre choix entériné.

2 Évaluation

2.1 Livrables

Les résultats seront présentés par chaque groupe dans un rapport et au cours d'une soutenance. La rédaction du rapport (conseils, règles de présentation) fait l'objet d'un document disponible sur la page Moodle de l'UV. Il doit être rendu dix jours avant la date de soutenance. Celle-ci a lieu lors du dernier TD ou du dernier cours, soit entre le 20/06 et le 22/06.

2.2 Soutenances

La soutenance nous permet d'apprécier assez précisément l'implication et le travail des membres du groupe. L'objectif est de voir dans quelle mesure chacun est capable de défendre le travail et les choix faits. Pour se concentrer sur cet aspect, les soutenances consistent *uniquement en une séance de questions-réponses*.

C'est un exercice difficile qui nécessite d'être bien préparé. Chacun doit s'être approprié les choix (techniques et scientifiques) faits et avoir une très bonne connaissance *globale* du travail réalisé. Le jury évaluera cette connaissance au besoin en posant des questions poussées — l'objectif n'étant pas d'être déplaisant ou humiliant, mais de voir à quel point les étudiants se sont impliqués et ont travaillé en bonne intelligence.

Le travail de manière cloisonnée (séparation des tâches, travail de chaque étudiant en autonomie, et mise en commun à la fin du semestre) n'est pas satisfaisant, et aboutit généralement à des notes faibles.

Revue à mi-parcours Pour vous permettre de mieux appréhender les attentes en termes de livrables, nous effectuerons une revue de projet à mi-parcours. Son déroulement sera similaire à la soutenance. Vous devrez rendre, au moins une semaine à l'avance, un pré-rapport de *deux pages maximum* sur vos réalisations (il pourra être repris pour le rapport final, sans obligation).

Lors de la séance, le chargé de TD interrogera chaque groupe sur le travail réalisé (pendant 15 à 20 minutes maximum). Cette séance de travail pourra être l'occasion pour lui de commenter les choix faits et de donner une appréciation globale sur le travail et le pré-rapport.

Soutenance finale La soutenance finale se déroulera de la même manière que la revue à mi-parcours. Elle aura lieu lors de la dernière semaine, soit à l'horaire du TD soit à l'horaire du cours. Il sera nécessaire de rendre le rapport au moins dix jours avant la soutenance.

Lors de la soutenance, les groupes bénéficieront à nouveau de quinze à vingt minutes chacun, pour répondre aux questions du jury (constitué des membres de l'équipe enseignante). Il n'y aura *pas de présentation*, pour laisser un maximum de temps aux échanges.

2.3 Plagiat

Il n'est pas interdit de travailler à plusieurs groupes, tant que les contributions de chacun restent clairement identifiées. Les travaux antérieurs sur le sujet ou sur un sujet proche peuvent être pris en compte, à condition que ce soit mentionné (citez vos sources), et que le travail réalisé soit positionné par rapport à cet existant. Évidemment, si vous utilisez des travaux antérieurs, les attentes du jury seront plus importantes.

Tout plagiat (ré-appropriation d'un autre travail, quel qu'il soit) sera sévèrement sanctionné. D'une manière générale, chaque groupe sera noté sur le travail réalisé, bien plus que sur les résultats présentés.

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 par le groupe. La note d'un étudiant peut être ajustée par rapport à celle de son binôme en fonction du travail manifestement réalisé (apprécié grâce à sa prestation orale).

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 Jeux de données

Vous choisirez un jeu de données parmi ceux disponibles sur le répertoire GitHub du [Tidy Tuesday](#). Vous choisirez vos données parmi les celles mis en ligne [en 2021](#) ou [en 2022](#) (listes disponibles en bas de page). Chaque groupe doit choisir un jeu de données différent des autres. Les choix doivent être indiqués sur la page Moodle de SY09 avant le 21/04.

Si vous souhaitez traiter un autre jeu de données, vous pouvez en faire la demande aux enseignants *en la motivant*. Nous comprenons (et approuvons) que des étudiants souhaitent travailler sur des données qui leur tiennent particulièrement à cœur, mais nous refuserons probablement les choix uniquement guidés par l'existence de solutions disponibles en ligne.