

Projet

Instructions :

- * Formation des équipes :
 - Le projet est réalisé en équipes de 3 à 5 étudiants
 - Les équipes peuvent être différentes de celles des devoirs et peuvent comporter des étudiants des deux versions du cours (GIF-4101 et GIF-7005)
 - Formation des équipes dans monPortail, au plus tard le mercredi 17 octobre à 9h30
 - * Programmation :
 - Utilisez Python autant que possible
 - * Remise :
 - Proposition de projet dans monPortail, au plus tard le vendredi 26 octobre à midi
 - Fichier zip avec rapport et code source dans monPortail, au plus tard le vendredi 21 décembre à midi
 - * Pondération :
 - Ce devoir compte pour 15% de la note finale
-

Présentation et organisation

Pour le projet, vous avez la possibilité de définir et exécuter le projet de votre choix sur l'application de l'apprentissage machine à des données spécifiques. Le projet doit être effectué en équipes de 3 à 5 étudiants, pouvant provenir des différents cycles d'études (le mélange entre étudiants inscrits à GIF-4101 et GIF-7005 est possible). Les équipes peuvent être différentes de celles formées pour les devoirs.

Le projet se déroule en trois étapes :

- Formation des équipes dans monPortail, au plus tard le mercredi 17 octobre à 9h30;
- Proposition de projet, transmise dans monPortail au plus tard le vendredi 26 octobre à midi;
- Réalisation du projet, transmission dans monPortail d'un rapport de projet sous la forme d'un article scientifique¹, accompagné du code source, au plus tard le vendredi 21 décembre à midi.

Proposition de projet

Vous devez préparer une proposition de projet dans un document de deux pages maximum. L'objectif est d'y faire une présentation précise et concise du problème que vous voulez résoudre, des méthodes envisagées à cette fin ainsi que des jeux de données que vous voulez utiliser, avec références complètes sur ceux-ci. Votre proposition devra démontrer que votre projet est d'une envergure adéquate pour un projet de fin de cours. Vous devrez transmettre la proposition de projet (fichier PDF) dans monPortail.

Vous devez prévoir quelques jours avant de recevoir une décision. Il est possible que des précisions ou une bonification du projet soient demandées avant que vous ailliez l'autorisation de poursuivre.

1. Vous pouvez utiliser à cette fin un patron d'articles tel que celui de la conférence ICML : <https://icml.cc/Conferences/2018/StyleAuthorInstructions>.

Réalisation du projet

Nous vous demandons d'utiliser le langage Python pour réaliser votre projet, utilisant des bibliothèques comme scikit-learn, PyTorch et/ou d'autres bibliothèques si nécessaire. Vous devez fournir tout le code que vous avez produit pour votre projet dans le dépôt dans monPortail, incluant le code nécessaire pour lire et prétraiter les jeux de données que vous utilisez. Également, fournir les résultats de vos expérimentations dans votre dépôt si ceux-ci sont de taille raisonnable (quelques Mo au plus).

Vous devez préparer un rapport ayant la forme d'un **article scientifique**, d'un maximum de 6 pages (police de 10 points, double colonne), sur votre projet. Nous vous demandons de fournir les éléments suivants dans ce document :

- Présentation du problème et état de l'art
- Approche proposée
- Méthodologie expérimentale pour analyse et comparaison des résultats
- Résultats expérimentaux, avec tableaux, figures et tests statistiques
- Analyse des résultats et discussions

Soyez brefs mais rigoureux dans votre rapport et évitez tout verbiage inutile. Selon les circonstances, le contenu de l'article peut être ajusté pour refléter l'essentiel du travail proposé.

Suggestions de projets

Des suggestions de projets proposés par des collaborateurs seront communiquées dans Piazza d'ici le 20 octobre. Vous pouvez contacter les collaborateurs directement, en mettant le professeur en copie, pour obtenir plus de détails et vous entendre sur la tâche précise qui sera proposée.

Également, nous vous encourageons à participer au *ICLR Reproducibility Challenge*² comme projet, où vous devez reproduire des expériences présentées dans des articles soumis à la conférence ICLR 2019 et en cours d'évaluation. Si vous suivez cette piste, une participation complète au challenge est demandée. En particulier, vous devez soumettre les résultats de la reproduction des résultats de l'article dans OpenReview. Pour les équipes ayant choisi un tel projet, le rapport pourra insister plus sur les aspects techniques associés à reproduire un travail scientifique et en analyser les résultats, et moins sur la présentation du problème et l'état de l'art.

Autrement, vous pouvez consulter les différents portails de données ou engins de recherche de données³. L'obtention de jeux de données probants est généralement centrale au bon succès des projets.

Barème d'évaluation

- Qualité et rectitude de l'approche proposée (6pt)
- Qualité et rectitude de la méthodologie expérimentale (4pt)
- Qualité et analyse des résultats expérimentaux (6pt)
- Qualité de la langue et de la présentation (4pt)

Un bonus de 5 points sur 20, pour un maximum de 20/20, sera donné à l'équipe ayant présenté globalement le meilleur projet.

4/10/2017

CG

2. https://reproducibility-challenge.github.io/iclr_2019/

3. Voir par exemple : <https://toolbox.google.com/datasetsearch>.