



Développez une preuve de concept (option stage)

MISSION COURS ÉVALUATION



Mis à jour le mercredi 12 mai 2021

Vous pouvez réaliser un stage à la place de ce projet. Consultez les **conditions** à remplir.

Introduction

Le domaine du machine learning, et plus généralement de la data science, évolue très rapidement. Il est donc important que vous développiez l'habitude de vous tenir au courant des avancées dans le domaine en effectuant **une veille thématique**. Vous devez aussi être capable de **monter rapidement en compétence** sur une nouvelle thématique, en sachant effectuer une recherche et mettre **en pratique un nouvel algorithme de façon autonome (POC)**.

Votre mission

Votre supérieur vous demande de trouver une méthode plus récente pour améliorer la méthode utilisée en production chez un client. Vous procéderez en deux temps :

- En début de projet (1 à 2 jours de travail), après avoir fait un rapide tour d'horizon de l'état de l'art du domaine, vous soumettrez à votre supérieur (votre mentor) un plan prévisionnel de votre travail, avec l'algorithme envisagé, le dataset sur lequel vous pensez évaluer les performances, un ou deux arguments justifiant votre choix, deux ou trois références bibliographiques (post de blog ou article de recherche) vous permettant de présenter un état de l'art sur le problème étudié et sur lesquels votre travail futur s'appuiera.
- Une fois la démarche validée, vous approfondirez votre compréhension de cette nouvelle approche, et l'implémenterez, ainsi qu'une méthode baseline simple (algorithme déjà utilisé sur un projet précédent, qui jouera le rôle de ce qui est actuellement en production).

Vous vous appuierez également sur les sources sélectionnées pour présenter l'état de l'art et les avancées récentes (un exemple très récent, <18 mois, et rien de plus de 5 ans).

Contraintes

1) Choix des données

Pour ce projet, vous pourrez vous appuyer au choix :

- Sur la méthode finale et le dataset d'un des projets du parcours, pour améliorer les performances.

Par exemple, choisissez un projet qui vous a particulièrement intéressé, ou sur lequel vous vous sentez moins à l'aise et que vous voulez retravailler.

- Sur un autre jeu de données et une problématique de votre choix.

2) Choix de l'algorithme implémenté en prototype

- L'algorithme doit provenir d'un article de recherche publié au cours des 18 derniers mois. Si aucun code (package, répertoire github, etc.) n'est disponible pour l'algorithme décrit, vous pourrez mettre en œuvre une méthode précurseur de l'algorithme décrit.
- La méthode effectivement utilisée doit avoir été publiée dans les 5 dernières années.

3) Réutilisation de code

Vous pouvez réutiliser du code (d'un tutoriel, par exemple) à condition de le préciser **explicitement**. Dans ce cas, l'application devra être originale, c'est-à-dire reposant sur un dataset différent de celui du tutoriel ou de l'article d'où provient le code.

4) Sources bibliographiques

Il est recommandé de sélectionner avec attention vos sources, en privilégiant les blogs sérieux. De bons critères de sélection peuvent être la longueur des articles, ou la présence de citation d'articles de recherche.

Sources conseillées :

- [fastml](#), [machine learning mastery](#), [kdnuggets](#), [import AI](#), [MIT tech review](#), [MIT news ML](#).
- Newsletters de qualité comme [data elixir](#) et [data science weekly](#)...
- Twitter, en suivant de grands noms de la discipline.
- Articles de recherche : <https://arxiv.org/>, <https://scholar.google.fr/>...

Il est obligatoire de s'appuyer sur au moins un article de recherche parmi les 2 à 3 sources du projet.

Si lire un article de recherche vous intimide, choisir l'article de recherche illustré dans un article de blog peut vous aider ! Dans ce cas, ils compteront ensemble pour une référence bibliographique. Si vous aimez les vidéos, beaucoup de conférences proposent des tutoriels (NIPS, ICML, ICCV...) qui sont des revues du domaine et peuvent vous aider à identifier des sources pertinentes.

Livrables attendus

- **Un plan de travail prévisionnel** d'une page présentant la thématique choisie, les sources bibliographiques identifiées et le prototype à implémenter (dataset, méthode baseline pour comparer les performances, méthode mise en œuvre) à fournir à votre mentor en début de projet.
- **Le code** (baseline + nouvelle méthode) et le jeu de données du prototype.
- **Un rapport rédigé** (max 10 pages) présentant la thématique, l'état de l'art, le jeu de données choisi, la méthode implémentée et ses performances, en comparaison avec la méthode baseline. Une analyse des résultats, et des éventuels cas d'échec est attendue.
- **Un rapport de stage** (le cas échéant).
- Un **support de présentation** pour la soutenance.

Pour faciliter votre passage au jury, déposez sur la plateforme, dans un dossier nommé "P7_nom_prenom", tous les livrables du projet. Chaque livrable doit être nommé avec le numéro du projet et selon l'ordre dans lequel il apparaît, par exemple "P7_01_plan", "P7_02_code", et ainsi de suite.

Soutenance

- 25 minutes de présentation de votre recherche ou de votre rapport de stage
- 5 à 10 minutes de questions-réponses

Ressources complémentaires

Le cours OpenClassrooms [Réalisez des rapports statistiques clairs et impactants](#).

Compétences évaluées



Identifier des sources d'informations fiables et pertinentes



Identifier les méthodes "état de l'art" pour résoudre un problème de Data science



Développer une preuve de concept pour résoudre un problème de Data science



Réaliser une veille sur les évolutions de la Data Science

OPENCCLASSROOMS



POUR LES ENTREPRISES



AIDE



EN PLUS



Français



Télécharger dans
l'App Store