

SDA - sujet de projet

Anne Jeannin-Girardon, Baptiste Lafabregue

October 2025

1 Organisation

Les projets peuvent être réalisés en équipes de deux ou trois étudiants au maximum. Les différents dates à retenir sont :

- Dépôt des propositions de sujet le vendredi 05/12/25 à 23:59
- Rendu final le mardi 13/01/25 à 23:59
- Présentation lors du dernier TP (15/01/25)

2 Types de sujet

2.1 Projet applicatif

Pour ce type de projet, vous devrez créer un jeu de données. Ce jeu de données pourra se baser sur un jeu de données existant, mais devra être, a minima, enrichi de nouvelles caractéristiques (features). Il est important d'identifier vos données en entrée et votre objectif de prédiction en sortie. Vous devrez ensuite comparer plusieurs méthodes d'apprentissage automatique sur votre jeu de données.

Une attention particulière sera portée à la pertinence et à la qualité des caractéristiques choisies pour décrire une instance dans vos données, ainsi qu'au travail de préparation et de mise en forme des données. Votre jeu de données devra se baser sur des données provenant du site <https://www.data.gouv.fr/fr/>, ainsi que sur au moins une autre source de données, issue aussi du site data.gouv ou d'une autre source.

Attention, si plusieurs groupes se retrouvent à travailler sur des données similaires, ils devront se distinguer par l'angle d'analyse, ainsi que par les approches utilisées, notamment en ce qui concerne la source des données externes.

Exemple de sujet : prédiction de la langue vivante 1 et 2 en fonction du score aux élections législatives dans une commune donnée.

La base de données portant sur les accidents de la route peut être une source intéressante de sujet d'étude pour votre projet.

2.2 Projet théorique

Pour ce type de sujet, vous devrez étudier une méthode existante de l'état de l'art. Il vous faudra identifier une ou plusieurs limites de cette méthode ou analyser le comportement de l'un des composants de cette méthode en fonction des données étudiées ou d'un ou plusieurs hyperparamètres choisis. Une attention particulière sera portée à la démarche expérimentale adoptée. Il est important de noter que cette catégorie se rapproche d'un travail de recherche. Cela requiert donc une bonne compréhension du fonctionnement des méthodes étudiées en cours.

2.3 Projet ouvert

Il est toujours possible de proposer un sujet qui sorte des trois catégories ci-dessus. La pertinence du sujet (données et approche choisies) devra être clairement argumentée. Il est fortement recommandé de discuter de ce type de sujet en amont avec l'équipe pédagogique.

3 Dépôt des propositions

Le dépôt des propositions doit comprendre la liste des membres du groupe, ainsi que le type de sujet choisi. Mis à part le "projet algorithme", vous devez présenter en une ou deux phrases ce que vous comptez faire. Plus de détails sont demandés pour le "projet ouvert", qui doit comprendre une description des données étudiées, leurs sources et l'approche méthodologique envisagée. Les propositions doivent prendre la forme d'un fichier pdf, d'une page maximum, contenant la liste des membres du groupe et un résumé du projet. Le résumé devra comprendre une courte description des données choisies avec un lien vers les sources et une description rapide des méthodes envisagées.

4 Rendu écrit

Le rendu final prendra la forme d'un rapport de 8 à 12 pages (références et annexes éventuelles comprises) dont la structure s'apparentera à celle d'un article scientifique, c'est-à-dire avec une introduction exposant le contexte et la problématique spécifique que vous avez traité; une description argumentée de votre proposition pour aborder votre problématique ; une description des données ; une description de votre protocole expérimentale; une présentation de vos résultats et leur discussion; une conclusion sur le travail réalisé. Vous devrez déposer votre rapport en format pdf sur Moodle avant le 13 janvier 2025 à 23h59. Votre rapport devra inclure un passage indiquant la nature des contributions de chaque membre de votre groupe.

Vos projets seront évalués selon les critères suivants:

- La description et formalisation claire du problème traité.

- La qualité argumentative de la justification de votre proposition pour aborder le problème traité.
- La qualité de votre approche expérimentale (algorithmes et données utilisées; paramétrage de votre / vos modèles; protocole d’entraînement des modèles, *etc.*).
- La qualité de l’analyse et des discussions de vos résultats (*NB*: des résultats sont des résultats, que vous les trouviez satisfaisant ou non en terme de performance: il n’est pas attendu de vous que vous parveniez à obtenir les meilleures performances possibles dans votre projet, mais que vous montriez que vous êtes capables de formaliser un problème, d’émettre des hypothèses pour aborder ce problème, de mettre en place un protocole expérimental clair et adéquat, de discuter les résultats obtenus et de relever les points saillants de ces résultats –qu’ils soient “bons” ou non).
- La citation des sources utilisées et leur pertinence en regard de votre problématique.
- La qualité rédactionnelle en générale.

5 Présentation

Le présentation de vos projet se fera sous la forme d’une présentation orale de 12min suivi d’un séquence de questions de 5min.

Les présentations seront évaluées sur la base de :

- La qualité, en terme de contenu et de clarté, des supports.
- La qualité des réponses aux questions posées.