L2 Info.. Introduction à la Recherche.

Rédaction d'un article scientifique à partir d'une recherche synthétisée dans un notebook jupyter

Jean-François COUCHOT

couchot [arobase] femto-st [point] fr

21 mars 2025

1 Contexte

Ce TP est à réaliser en binôme et fait suite au TP 1 d'initiation aux expérimentations en recherche.

Dans ce TP, on cherche à rédiger un article de recherche synthétisant les expérimentations précédentes en respectant les standards de la recherche scientifique.

Au plus tard 5 jours après ce TP, l'article généré au format pdf sera à déposer sur l'espace de dépôt moodle. Il sera ensuite relu, critiqué et noté par un autre binôme de la promotion.

2 Export du notebook jupyter au format $\LaTeX$

On reprend le notebook jupyter déjà partagé en lecture :

<https://drive.google.com/file/d/liwOPCO9FD2qV6Dt17-3C6DM4Zma6Z3u0/view?usp=sharing>

Dans la section nommée "Export en latex du notebook jupyter" sont détaillées les étapes pour exporter ce notebook au format $\LaTeX$ qui est le langage de référence dans la publication d'article en informatique.

Exercice 2.1 (Export de votre notebook jupyter). 1. Exécuter tout le code de cette section (sauf le montage du drive pour lequel vous n'avez pas les droits) pour réaliser cet export. Constaté la production d'un fichier latex correspondant avec un dossier contenant les images.

2. Recopier cette section dans votre notebook et exporter celui-ci. A partir de cette étape, vous travaillerez sur ce document (et pas sur celui correspondant à l'export de mon notebook).

3 Prise en main d'un environnement collaboratif d'édition de fichier $\LaTeX$

La rédaction d'un article de recherche se réalise (souvent) de manière collaborative. Dans ce TP, on exploitera le service overleaf.

Exercice 3.1 (Projet initial reproduisant le notebook jupyter). 1. Créer un seul nouveau projet sur cette interface pour le binôme avec comme nom "Intro 2025 Rech Nom1 Nom2".

2. Copier tout le contenu de votre fichier latex généré dans l'exercice précédent.

3. Importer le dossier contenant les images, dossier qui a été généré à l'exercice précédent.

4. Demander une compilation du fichier et constater qu'un fichier PDF contenant l'ensemble a été généré.

5. Pour déclarer que le document est en langue française et préciser que l'encodage est de l'utf8 ajouter dans le préambule latex le code suivant :

```
\documentclass[11pt]{article}
\usepackage[french]{babel}
\usepackage[utf8]{inputenc}
\DeclareUnicodeCharacter{2501}{\rule{1em}{0.1ex}}
\DeclareUnicodeCharacter{257A}{\rule{1em}{0.1ex}}
```

6. Partager ce projet entre les 2 membres du binôme (bouton Share en haut à droite de l'interface d'overleaf).

Exercice 3.2 (Un premier nettoyage des sorties “inutiles”). C'est tout le notebook qui a été exporté : la structuration, les cellules de code et les cellules de sortie. On va enlever certaines cellules de sortie qui sont d'un trop bas niveau technique.

1. Repérer le code latex correspondant au texte suivant de sortie et le supprimer.

```
Collecting ucimlrepo
  Downloading ucimlrepo-0.0.7-py3-none-any.whl.metadata (5.5 kB)
Requirement already satisfied: pandas>=1.0.0 in /usr/local/lib/python3.11/dist-packages (from ucimlrepo) (2.2.2)
Requirement already satisfied: certifi>=2020.12.5 in /usr/local/lib/python3.11/dist-packages (from ucimlrepo) (2025.1.31)
Requirement already satisfied: numpy>=1.23.2 in /usr/local/lib/python3.11/dist-packages (from pandas>=1.0.0->ucimlrepo) (1.26.4)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.11/dist-packages (from pandas>=1.0.0->ucimlrepo) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas>=1.0.0->ucimlrepo) (2025.1)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-packages (from pandas>=1.0.0->ucimlrepo) (2025.1)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.2->pandas>=1.0.0->ucimlrepo) (1.17.0)
Downloading ucimlrepo-0.0.7-py3-none-any.whl (8.0 kB)
Installing collected packages: ucimlrepo
Successfully installed ucimlrepo-0.0.7
```

2. Supprimer toutes les sorties verbatim de sortie (mais pas les figures). On pourra se répartir le travail entre les deux membres, l'un-e commençant par le début de l'article, l'autre commençant par la fin.
3. Supprimer toute la sous-section “Export en latex du notebook jupyter”.

A l'issue de cet exercice, vous devriez avoir un article de 7 pages approximativement.

4 Structuration d'un article de recherche scientifique

Usuellement un article de recherche en informatique est structuré de la manière suivante :

1. Titre (nom des auteur-e-s, affiliations, date et résumé);
2. Introduction présentant le contexte, les enjeux, la problématique et annonçant le plan;
3. Mise en perspective de l'état de l'art avec des références et discussions autour de celles-ci;
4. Présentation des contributions;
5. Expérimentations éventuelles et analyse de celles-ci;
6. Discussion et analyse générales;
7. Conclusion et perspectives.

Exercice 4.1 (Titre et résumé). 1. Rechercher la balise `\title` et choisir un titre adapté à ce travail sachant que celui ci présente quelques performances numériques d'une méthode de classification respectueuse de la vie privée sur un jeu de données précis.

2. Vous êtes les auteur-e-s de ce travail et faites partie de l'Université Marie et Louis Pasteur. Ajouter cette information dans votre document. On pourra exploiter l'aide d'overleaf sur le sujet. Dans ce cas, il faudra supprimer la ligne qui importe le package `titling` dans le préambule.
3. Déclarer un abstract comme le premier élément du document après le titre et le remplir avec une ligne de description de ce travail. On pourra s'inspirer de ce qui suit.

```
\begin{document}
\maketitle
\begin{abstract}
  Ce travail présente comment nous avons mis en place une méthode d'évaluation
  des performances numériques d'une méthode de
  classification respectueuse de la vie privée sur un jeu de données précis.
  Le classifieur est XXX et le jeu de données est YYY.
\end{abstract}
```

Exercice 4.2 (Structuration de l'article). La structure du document actuel ne respecte pas ce qui doit être réalisé dans un article scientifique. Cet exercice va permettre de la définir.

1. Entre l'abstract et la section déjà générée, déclarer les sections "Introduction" et "Etat de l'art";
2. Renommer le titre de la section "Reproduction d'expérimentations d'un article de Machine Learning respectueux des données" (qui a été généré) par "Tutoriel de mise en place de la classification respectueuse de la vie privée sur le dataset XXX";
3. Il y a actuellement une sous-sous-section nommée "Evaluation de l'impact du budget sur la métrique d'exactitude". Insérer juste avant une section nommée "Expérimentations"; Changer le type de cette sous-sous-section en une `\subsection`. Si vous avez d'autres sous-sous-sections pour les métriques de "rappel", de "précision" (comme cela l'était suggéré dans la dernier TP), changer de même leur type.
4. Insérer à la fin les sections de "Discussion et analyse" et de "Conclusion et perspective".

5 Ajout de contenu scientifique

On commence par compléter la section "état de l'art". En effet, ce travail de rédaction s'appuie sur plusieurs références :

- La méthode d'apprentissage machine à base d'arbre de décision [Bre+84] implantée dans Scikit-Learn [Ped+11];
- Une extension de cette méthode respectueuse de la vie privée [FI17] implantée dans diffprivlib [Hol+19];
- Le jeu de données : Adults [BK96] initialement, mais un autre pour vous.

Exercice 5.1 (Un état de l'art minimaliste). 1. Créer dans l'interface overleaf un nouveau fichier nommé "mybibliography.bib".

2. Rechercher chacune des références citées ci-dessus et référencées ci-dessous au format bibtex, par exemple sur le site <https://dblp.org/>. et les copier dans "mybibliography.bib". Repérer la clé de chacune des références.
3. Lire une documentation simple sur l'intégration d'une bibliographie dans un document latex.
4. Ajouter alors dans le préambule du document latex

```
\usepackage[backend=biber,style=alphanumeric,sorting=ynt]{biblatex}
\addbibresource{mybibliography.bib}
```

5. Dans la section "Etat de l'art", créer un texte s'inspirant de celui donné juste avant cet exercice qui donne un état de l'art autour de ce travail.
6. Tout à la fin du document, mais avant `\end{document}`, ajouter la commande suivante qui demande la construction de la bibliographie.

```
\printbibliography
```

Exercice 5.2 (Analyse des expérimentations). Écrire un paragraphe d'analyse dans la section portant le nom "Discussion et Analyse".

Exercice 5.3 (Une introduction minimaliste). Dans la section "Introduction" :

1. Écrire un paragraphe présentant le contexte, les enjeux, la problématique, les objectifs de cet article.
2. Écrire un paragraphe d'annonce du plan à la fin de cette section.

Exercice 5.4 (Une conclusion minimaliste). Dans la section "Conclusion et Perspectives" :

1. Écrire un paragraphe faisant la synthèse de ce travail.
2. Écrire un paragraphe énonçant ce qui pourra être fait, à la suite de ce travail.

Références

[Bre+84] Leo BREIMAN et al. *Classification and Regression Trees*. Wadsworth, 1984. ISBN : 0-534-98053-8.

[BK96] Barry BECKER et Ronny KOHAVI. *Adult*. UCI Machine Learning Repository. DOI : <https://doi.org/10.24432/C5XW2> 1996.

[Ped+11] F. PEDREGOSA et al. "Scikit-learn : Machine Learning in Python". In : *Journal of Machine Learning Research* 12 (2011), p. 2825-2830.

- [FI17] Sam FLETCHER et Md Zahidul ISLAM. “Differentially private random decision forests using smooth sensitivity”. In : *Expert Syst. Appl.* 78 (2017), p. 16-31. DOI : 10 . 1016 / J . ESWA . 2017 . 01 . 034. URL : <https://doi.org/10.1016/j.eswa.2017.01.034>.
- [Hol+19] Naoise HOLOHAN et al. “Diffprivlib : the IBM differential privacy library”. In : *ArXiv e-prints* 1907.02444 [cs.CR] (juill. 2019).