

Intelligence Artificielle 2

Cours 5 - Étapes d'un projet de ML/DL

Steve Lévesque, Tous droits réservés © où applicables

Table des matières

- 1 Les étapes d'un projet de Machine Learning
 - Les environnements
 - Exploration - La structure du partage et démonstration des découvertes (Notebook)
 - Conception - La structure du modèle/algorithme (Code dans l'IDE)
 - Rédaction de rapport - Présentation des résultats

Les étapes d'un projet de Machine Learning

Il est important d'avoir un environnement de travail organisé pour avoir :

- du code maintenable et extensible
- du code déployable, sécuritaire et testé
- des graphiques à l'appui
- des explorations et résultats partageable rapidement
- une exécution très rapide (sans perdre de temps à installer des modules)
- une gestion organisée des données et fichiers entrants/sortants

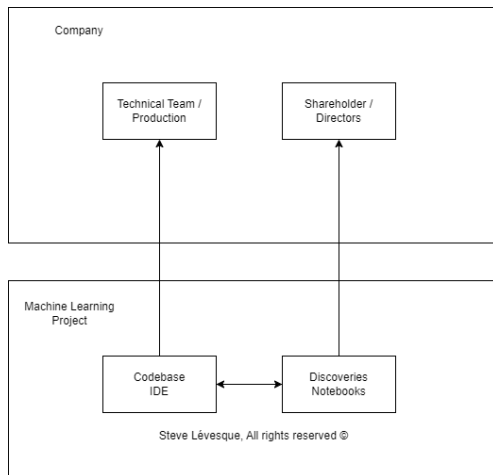
Les environnements

Certains environnements (Notebooks ou IDE) sont meilleurs que d'autres pour accomplir une tâche :

- (IDE) du code maintenable et extensible
- (IDE) du code déployable, sécuritaire et testé
- (Notebooks) des graphiques à l'appui
- (Notebooks) des explorations et résultats partageable rapidement
- (Notebooks) une exécution très rapide (sans perdre de temps à installer des modules)
- (IDE et Notebooks) une gestion organisée des données et fichiers entrants/sortants

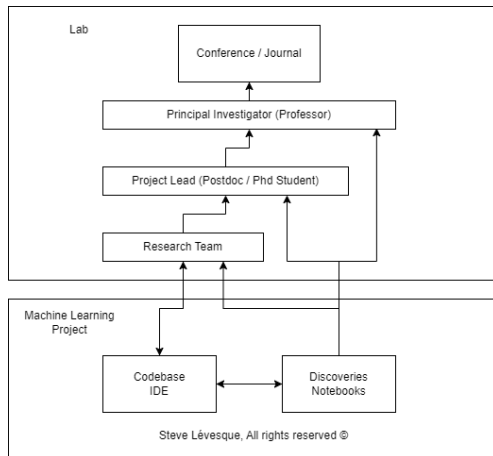
Les environnements - Profession

NB : Cela peut différer d'une entreprise à une autre



Les environnements - Académique

NB : Cela peut différer d'une institution à une autre



Exploration - La structure du partage et démonstration des découvertes (Notebook)

Un/des **Notebook(s)** avec les parties suivantes :

- **La visualisation des données** : comprendre et voir les données pour comprendre le domaine d'affaire et le but du projet.
- **Les statistiques sur les données** : avoir un compte rendu sur les données
 - Détecter s'il manque des valeurs
 - Détecter s'il y a des problèmes de mise-à-niveau ("Scaling")
 - Etc.
- **Interprétation des résultats** : si nécessaire, il est possible de prendre les résultats du programme conceptualisé dans l'IDE pour en faire des meilleures représentations graphiques
 - i.e. "Confusion Matrix" en graphique si le projet dans l'IDE ne produit que les données (structure matricielle)

Conception - La structure du modèle/algorithme (Code dans l'IDE)

Un **Projet Python dans un IDE** avec les parties suivantes :

- **Importation et nettoyage des données** : les données propres de tous défauts pour aider le modèle de Machine Learning à bien performer (une des étapes les plus importantes d'un projet professionnel/académique en Machine Learning)
- **Conception/import du modèle** : un modèle pour avoir une application fonctionnelle (entraînement et évaluation)
- **Entraînement du modèle** : une phase d'entraînement (train set) avec le modèle, Loss function, l'optimizer, et metrics
- **Évaluation du modèle** : une phase d'évaluation (test set) pour tester notre modèle et voir si il est “underfitted” ou “overfitted”, en résumé pour savoir si le modèle performe
- **Exécution globale** : une fonction principale ou un “main” pour exécuter toutes les étapes ci-haut avec plusieurs hyper-paramètres différents si on le souhaite

Rédaction de rapport - Présentation des résultats

Il est nécessaire, surtout dans le domaine académique, de documenter les découvertes avec des démarches exploratives et de résultats pour prouver l'efficacité de la solution que vous proposez.

C'est une partie très peu technique, souvent détestée par la plupart des développeurs, mais celle-ci est complémentaire et nécessaire pour bien interpréter votre code/projet.

Rédaction de rapport - Présentation des résultats -
Sélection de l'outil de rédaction

Il existe deux outils principaux de rédaction : Word et L^AT_EX.

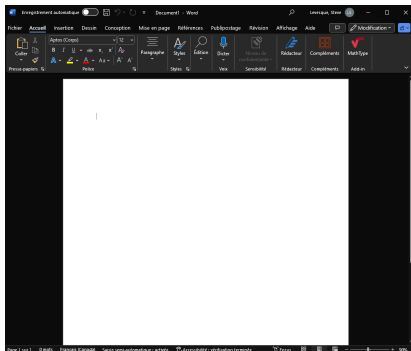


Figure: Microsoft Word, un outil facile d'utilisation et collaboratif le plus répandu au monde soutenant l'expérience utilisateur et la rédaction de manière naturelle

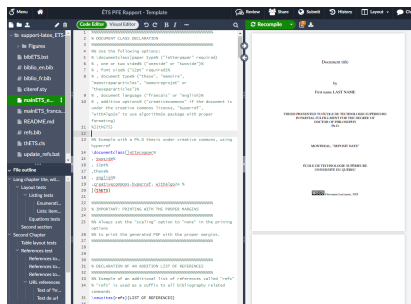


Figure: \LaTeX (Overleaf), un outil de rédaction collaboratif axé sur la programmation pour avoir une vision sur le résultat plutôt que l'apparence prise en charge automatiquement

Rédaction de rapport - Présentation des résultats - Sections du rapport

Voici toutes les sections nécessaires dans un rapport complet d'un projet de Machine Learning :

- Introduction
- Notion de base (et revue de la littérature)
- Architecture proposée
- Résultats
- Conclusion
- Annexe
- Bibliographie

Rédaction de rapport - Présentation des résultats -

Sections du rapport

Étape 1 : **Comprendre le problème du projet**

- Introduction
 - Introduire la tâche et résumé du contenu du rapport (rédaction)
- Notion de base (et revue de la littérature)
 - Résumer l'état de l'art récent des quelques années (provenance : lecture d'articles)

Rédaction de rapport - Présentation des résultats - Sections du rapport

Étape 2 : **Partie technique du projet**

- Architecture proposée
 - Expliquer la conception (provenance : Code dans l'IDE)
- Résultats
 - Montrer les graphiques exploration (provenance : Notebooks)
 - Montrer les résultats (provenance : Code dans l'IDE et
possiblement Notebooks pour traitements additionnels)

Rédaction de rapport - Présentation des résultats -

Sections du rapport

Étape 3 : **Conclusion du projet et révision globale**

- Conclusion

- Conclure avec un résumé bref du projet, résultats principaux obtenus et d'une extension future possible (rédaction)

- Annexe

- Si vous avez beaucoup trop de figures (graphiques, tableaux, etc.), les mettre en Annexe à la fin

- Bibliographie

- **OBLIGATOIRE**, il est important de citer les sources (ChatGPT, Articles, liens sur Internet, etc.)

Bibliographie

■ `https://www.overleaf.com/`