WEBINAIRE

REPRODUCTIBILITÉ EN APPRENTISSAGE AUTOMATIQUE

30 OCTOBRE 2020

Faculté des sciences et de génie
Institut Intelligence et Données (IID)
UNIVERSITÉ LAVAL

OBJECTIFS DE LA PRÉSENTATION

- Sensibiliser sur les enjeux de la reproductibilité.
- Inciter l'intégration des solutions permettant une meilleure reproductibilité dans vos solutions d'affaires et académiques.
- Améliorer votre productivité.

VOTRE CONFÉRENCIER

DAVID BEAUCHEMIN
Candidat au doctorat
Département d'informatique et de génie logiciel

- Introduit à la recherche reproductible en 2016 (R Markdown et Git)
- Participation à REPROLANG de la conférence LREC [Garneau et al., 2020]
- Membre actif dans le développement d'une librairie facilitant la reproductibilité (Poutyne)

AU MENU

Gestion version

Productivité

Présenter

Réutiliser

Introduction

C'EST QUOI LA REPRODUCTIBILITÉ ?

La reproductibilité est le principe qu'on ne peut tirer de conclusions que d'un événement bien décrit, qui est apparu plusieurs fois, provoqué par des **personnes différentes**.

Toutefois, ont utilise souvent ce terme pour spécifiquement désigner la **réplicabilité**. Soit la réplication (reproduction) des résultats d'un article dans des environnements pas (toujours) différents [Drummond, 2009, Pineau et al., 2020].

POURQUOI S'Y INTÉRESSER ?

70 %[1]

1. [Baker, 2016]

POURQUOI S'Y INTÉRESSER ?

50 %[1]

1. [Baker, 2016]

POURQUOI S'Y INTÉRESSER ?

40 % [2]

2. [Raff, 2019]

MOTIVATION

Réutilisation

MOTIVATION

Réutilisation

Productivité

MOTIVATION

Réutilisation

Productivité

Transfert

MOTIVATION

Réutilisation

Productivité

Transfert

Se faire connaître

Les barrières à la réplicabilité

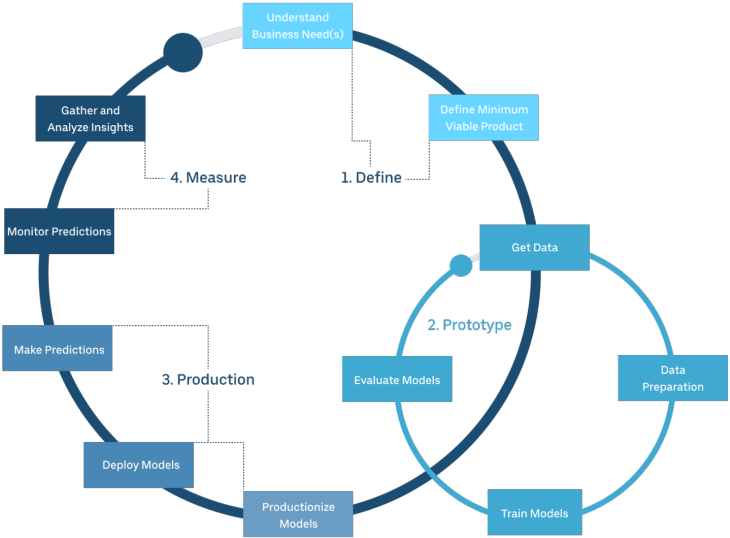

Figure 1 – From Uber Engineering

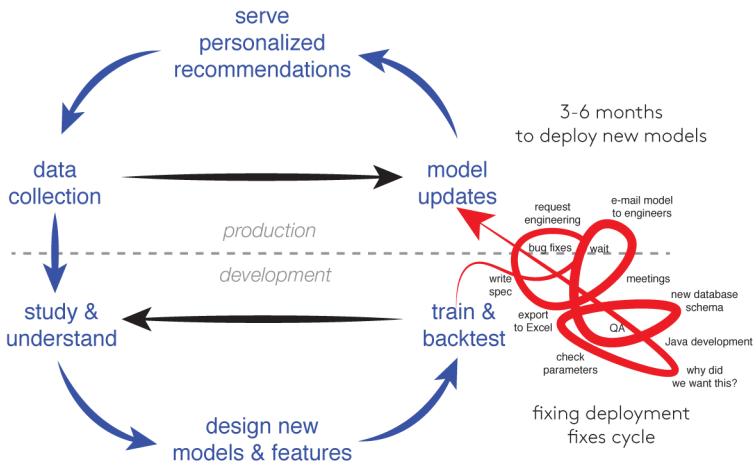

Figure 2 – The need for Agile machine learning

OK, mais comment ?

AU MENU

Gestion version

Productivité

Présenter

Réutiliser

VERSION DES DONNÉES & ÉTAPES DE PRÉTRAITEMENT

Version

VERSION DES DONNÉES & ÉTAPES DE PRÉTRAITEMENT

Version

Gestion des versions

VERSION DES DONNÉES & ÉTAPES DE PRÉTRAITEMENT

Version

Gestion des versions

Étapes prétraitement

VERSION DES DONNÉES & ÉTAPES DE PRÉTRAITEMENT

Data Version Control

Dask

CODE

Version

CODE

Version

Différence

CODE

Version

Différence

Divergences

CODE

Git

GitHub

GitLab

Bitbucket

AU MENU

Gestion version

Productivité

Présenter

Réutiliser

DÉVELOPPEMENT DES MODÈLES

Réinventer

DÉVELOPPEMENT DES MODÈLES

Réinventer

Simplification

DÉVELOPPEMENT DES MODÈLES

Réinventer

Simplification

Facilite

DÉVELOPPEMENT DES MODÈLES

Poutyne

PyTorch Lightning

Scikit-learn

Gensim

Allen NLP

ENTRAÎNEMENT, CONFIGURATION ET RÉSULTATS

Version de
l'entraînement

ENTRAÎNEMENT, CONFIGURATION ET RÉSULTATS

Version de
l'entraînement

Résultats

ENTRAÎNEMENT, CONFIGURATION ET RÉSULTATS

Version de l'entraînement

Résultats

Visualisation

ENTRAÎNEMENT, CONFIGURATION ET RÉSULTATS

Version de l'entraînement

Résultats

Visualisation

Erreurs d'entraînement

ENTRAÎNEMENT, CONFIGURATION ET RÉSULTATS

MLflow

Hydra

Sacred

Notif

AU MENU

Gestion version

Productivité

Présenter

Réutiliser

RAPPORT ET ANALYSE DES RÉSULTATS

Tableau des résultats

RAPPORT ET ANALYSE DES RÉSULTATS

Tableau des résultats

Mise à jour

RAPPORT ET ANALYSE DES RÉSULTATS

Tableau des résultats

Mise à jour

Visualisation configuration

RAPPORT ET ANALYSE DES RÉSULTATS

Python2LaTeX

TensorBoard

Markdown

AU MENU

Gestion version

Productivité

Présenter

Réutiliser

ENVIRONNEMENT

Différents environnements

ENVIRONNEMENT

Différents environnements

Réutilisation

ENVIRONNEMENT

Docker

Kubernetes

La suite

Itérations d’expérimentations

POUR ALLER PLUS LOIN

- Clean code
- Continuous Machine Learning
- Faire des tests !
- Writing Code for NLP Research [Gardner et al., 2018]
- SOLID
- Cet article [Pineau et al., 2020]

PÉRIODE DE QUESTIONS

WEBINAIRE

MERCI DE VOTRE ÉCOUTE !

REFERENCES i

Baker, M. (2016).
1,500 Scientists Lift the Lid on Reproducibility.
Nature News, 533(7604) :452.

Drummond, C. (2009).
Replicability Is Not Reproducibility : Nor Is It Good Science.
Evaluation Methods for Machine Learning Workshop.

Gardner, M., Neumann, M., Grus, J., and Lourie, N. (2018).
Writing Code for NLP Research.
In *Conference on Empirical Methods in Natural Language Processing : Tutorial Abstracts.*

REFERENCES ii

- Garneau, N., Godbout, M., Beauchemin, D., Durand, A., and Lamontagne, L. (2020). **A Robust Self-Learning Method for Fully Unsupervised Cross-Lingual Mappings of Word Embeddings : Making the Method Robustly Reproducible as Well.**
- Pineau, J., Vincent-Lamarre, P., Sinha, K., Larivière, V., Beygelzimer, A., d'Alché Buc, F., Fox, E., and Larochelle, H. (2020). **Improving Reproducibility in Machine Learning Research (A Report from the NeurIPS 2019 Reproducibility Program).**
- Raff, E. (2019). **A Step Toward Quantifying Independently Reproducible Machine Learning Research.**