

ML-Ops

Benjamin GUIGON

January 2021

1 Lien

Article écrit par Zakaria Rida

<https://github.com/ZakariaRida96/PSBX>

2 Introduction

Au vu du nombre de projet dans les grandes DSI, il fallait trouver des solutions afin de simplifier, fluidifier et ramifier la progression d'un projet. Si on ajoute à ces besoins l'apparition de l'agilité, nous obtenons la naissance du DevOps. Cette approche, de plus en plus indispensable, se décompose en 2 parties : l'intégration continue (CI) et la distribution continue (CD). Le ML-Ops est donc le mélange d'implémentation d'algorithmes de Machine Learning avec une approche DevOps.

3 Niveaux de ML-Ops

3.1 ML-Ops niveau 0 : processus manuel

Les ingénieurs qui s'occupent de développer des modèles de machine learning ne sont pas souvent familiés avec les approches DevOps, ils déploient donc leurs modèles manuellement. Ils préparent les données, les entraînent grâce aux modèles et partagent leur résultats une fois le modèle fonctionnel. Les inconvénients : versionning peu fréquent, pas d'intégration continue, pas de livraison continue et pas de surveillance active.

C'est souvent le cas des entités qui débutent en Data science.

3.2 ML-Ops niveau 1 et 2

Dans des environnements où le DevOps est maîtrisé, on retrouve une beaucoup plus grande autonomie : Tests rapides, entraînement continu du modèle en production, modularité des composants, déploiement continu.

Une fois que l'approche DevOps est parfaitement maîtrisée, on retrouve un environnement de plus en plus pointu et autonome : Livraison continue et déploiement continu avec une surveillance intégrée.

4 Le CI et le CD

Le CI : test unitaires de toutes sortes, test de convergence, test d'entraînement, test de validité et test d'intégration.

Le CD : Vérification de compatibilité avec l'infrastructure, vérification de réponse avec l'API, test de la performance du système, validation des données, déploiement automatisé pour le testing, déploiement automatisé vers la production.

5 Explication

Ce rapport ne comporte pas de Maths à proprement parlé, mais il utilise des sous-jacents logiques qui font eux appel à des maths. Les technologies de Machine Learning et DevOps sont des environnements complexes qui demandent beaucoup de maîtrise. L'approche DevOps atteint des degrés de complexités qui nécessitent des ingénieurs spécialisés dans le DevOps. Il existe plusieurs dizaines de logiciels qui servent uniquement au DevOps : Jenkins, Git, Docker, Kubernetes, AWS, GCP etc...

L'article résume très bien les différents environnements du DevOps avec les processus associés. Pour une personne qui ne connaît pas la méthodologie DevOps, il aurait peut être fallu commencer par détailler le CI et le CD plutôt que de le mettre à la fin. Cependant, il ne faut pas aborder un article sur le DevOps sans connaître au minimum les deux aspects fondamentaux.

6 Conclusion

Cet article n'est pas mathématique mais il servira énormément aux personnes qui voudront accéder à un environnement de Data science dans une DSI très complexe. J'apprécie énormément ce travail, que je vais conserver car il me servira beaucoup par la suite. Ces connaissances permettent d'avoir une réelle profondeur face à des Data scientist qui ne connaîtraient que les langages de programmation ML.