Application of DevOps in the improvement of machine learning processes

Conference Paper · November 2020

DOI: 10.5281/zenodo.4318113

2 authors:

Beatriz Matsui
Universidade Federal do ABC (UFABC)
1 PUBLICATION **0** CITATIONS
SEE PROFILE

Denise Goya
Universidade Federal do ABC (UFABC)
7 PUBLICATIONS **3** CITATIONS
SEE PROFILE

Application of DevOps in the improvement of machine learning processes

Beatriz Mayumi Andrade Matsui[1], Denise Hideko Goya[2]

Federal University of ABC (UFABC), Av. dos Estados, 5001 - Bangú, Santo André - SP, Brazil

[1]beatriz.mayumi@ufabc.edu.br, [2]denise.goya@ufabc.edu.br

Abstract: This work deals with the concept of DevOps and its application in the improvement of machine learning processes. DevOps practices have been increasingly used by software development teams to automate and simplify processes ranging from integration, through testing, approval, implementation, to the final delivery of the application to users. The present study aims to focus on the possibility of applying this concept also in teams that work with machine learning and could benefit from the improvements brought with the adoption of DevOps.

1. Introduction

Software development requires some good practices to be done efficiently. In this context, the concept of DevOps is introduced – a term that represents the combination of *Development* and *Operations* – which can be defined as "a new way of thinking" in the scope of software engineering [1]. Its main objectives consist of narrowing the gap between development and operation areas, as well as ensuring that application development, deployment, and operation take place with high quality [2].

Similarly, *machine learning* (ML) is a field of study that is increasingly on the rise, which aims to improve performance and build applications capable of acquiring knowledge and reproducing actions based on previous experiences [3]. Machine learning techniques are commonly used in different areas of industry, such as health, education, and business, to automate processes and allow human activities that would take longer to be performed, to be optimized and completed more quickly and accurately [3].

The objective of this work is to investigate the applicability of DevOps to improve ML processes, through planning and practices of Continuous Integration (CI) and Continuous Delivery (CD).

2. CI/CD to improve ML practices

Continuous Integration is a DevOps practice of having the development phase of the software fully automated, validating code as soon as changes are committed and merged to a central repository in source control [4]. *Continuous Delivery* refers to the phase that comes after CI: the automated delivery of the application. When there is a new build artifact, a release is triggered and the artifact is automatically deployed in the desired environment. In Continuous Delivery, the process of triggering the deployment is manual. On the other hand, if the trigger from the development phase to delivery is automatic, then the process is called *Continuous Deployment*. [4].

Artificial intelligence (AI) and ML practices can be benefited by DevOps. AI/ML practices are essentially based on model iteration – which can take hours or even days, depending on the size of the model – until it is fully trained and tested. DevOps can assist in the stabilization and optimization of these processes, through CI/CD practices.

Also, teams that work with ML suffer from the lack of a model that adds value over time – instead, they generally use a single model from the beginning to the end of the project. DevOps practices can help in planning and implementing a lifecycle and continuous evolution of this model.

Moreover, building a machine learning model involves some characteristic steps, ranging from data collection and preparation to training, validation, and testing [5]. Teams have their own methodologies and work approaches, especially those related to data science, and many of them are not yet used to DevOps practices – which could be very useful for the development, deployment, and monitoring of these applications, in a continuous cycle of system improvement from the automated feedback of analyzed data and identified patterns.

3. Literature Review

In the literature, you can find studies such as [6], which shows how to apply machine learning + operations (MLOps) using tools and platforms that aim to automate the implementation of machine learning models. The authors develop a continuous integration system for machine learning to allow users to declare a wide range of integration conditions following strict probabilistic guarantees. In the work of [7], the authors create an application lifecycle model based on MLOps to optimize vision-based inspections in manufacturing processes. In [8], an example of MLOps applied to the development of a framework to combine machine learning with edge computing in Artificial Intelligence of Things (AIoT) applications is presented. And in [9], the study shows how to apply DevOps practices of CI and CD for machine learning applications.

Many applications use machine learning techniques and it can be seen that such techniques, when applied in isolation, are generally not enough – it is necessary to integrate with other modules and extensions that allow an automated integration. Thus, the combination of DevOps and machine learning – aiming at greater efficiency of the latter – shows itself as a field of study with wide possibilities for discoveries and applicability [10] – such as in the areas of health and safety [11], with data analysis, or in the monitoring and automation of applications [12].

4. Final Considerations

In practice, development and operations teams working in areas such as business intelligence, financial market data management, manufacturing industry, among others, have not yet achieved the desired efficiency, as their machine learning processes are often not well integrated into the continuous development and production cycle of their applications. As presented here, this study aims to contribute to the application of DevOps in machine learning practices, to optimize those processes, and promote greater integration and applicability of concepts of software engineering with machine learning. The next step of the study is to characterize the limitations and challenges of the area so that we can propose improvements.

5. References

[1] Jabbari, Ramtin, et al. “What is DevOps? A systematic mapping study on definitions and practices.” Proceedings of the Scientific Workshop Proceedings of XP2016. 2016.

[2] Agrawal, Prashant, and Neelam Rawat. “Devops, A New Approach To Cloud Development & Testing.” 2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT). Vol. 1. IEEE, 2019.

[3] Langley, Pat, and Herbert A. Simon. “Applications of machine learning and rule induction.” Communications of the ACM 38.11 (1995): 54-64.

[4] Rossel, Sander. Continuous Integration, Delivery, and Deployment: Reliable and faster software releases with automating builds, tests, and deployment. Packt Publishing Ltd, 2017.

[5] Domingos, Pedro. “A few useful things to know about machine learning.” Communications of the ACM 55.10 (2012): 78-87.

[6] Renggli, Cedric, et al. “Continuous integration of machine learning models with ease. ml/ci: Towards a rigorous yet practical treatment.” arXiv preprint arXiv:1903.00278 (2019).

[7] Lim, Junsung, et al. “MLOP lifecycle scheme for vision-based inspection process in manufacturing.” 2019 USENIX Conference on Operational Machine Learning (OpML 19). 2019.

[8] Raj, Emmanuel. “Edge MLOps framework for AIoT applications.” (2020).

[9] Karamitsos, Ioannis, Saeed Albarhami, and Charalampos Apostolopoulos. “Applying DevOps Practices of Continuous Automation for Machine Learning.” Information 11.7 (2020): 363.

[10] Schlossnagle, Theo. “Monitoring in a DevOps world.” Communications of the ACM 61.3 (2018): 58-61.

[11] Lwakatare, Lucy Ellen, et al. “DevOps in practice: A multiple case study of five companies.” Information and Software Technology 114 (2019): 217-230.

[12] Cornejo, Oscar, et al. “Field monitoring with delayed saving.” IEEE Access 7 (2019): 85913-85924.