DevOps for Machine Learning

Supervised by – Fraunhofer IPT

Summary:

The main aim of this thesis is to develop a tailor-made, automated framework for ML Ops in Production in order to achieve a **Standardized Model Deployment** process.

Background:

A machine learning model is built for solving a problem and to do so, it must be in production and usable by specific group of end users. Therefore, the process of Model Deployment holds equal weightage as Model Building. Indeed, according to many data scientists, it is the most difficult stage in machine learning lifecycle [1]. There are several ways of deploying a model like, on-premise deployment or cloud-based deployment, chosen according to requirements and constraints. However, just deploying a model is not sufficient. It must be monitored at regular intervals for its performance. Because, several issues might occur after deployment like, model drifting, rotten models, broken component, change in data etc. [2] [3] Therefore, the model must be trained regularly (Continuous Training) on new data. And, the new model (adapted to new data) should be deployed again and integrated with the system (Continuous Integration/Continuous Deployment). Hence, this is continuous process and can be implemented in 3 ways [4]:

1. Level 0 – Manual Process
2. Level 1 – ML Pipeline Automation
3. Level 2 – CI/CD Pipeline Automation

Thesis Goals:

1. Standardized Model Deployment process
2. Faster time to Deployment by using automated methods in the process (where possible automated)
3. Higher quality in Deployment and operation of models (Model Lifecycle)

Tasks:

1. Develop framework tailored to ML Ops in Production - can also be designed from other popular Ops frameworks.
2. Develop guideline through key ML Ops phases within the framework - code versioning and standards, saving data and model versions (model management), monitoring, continuous training and validation of performance etc.
3. Integrate available tools and platforms to automate ML Ops steps - for continuous integration and continuous delivery (CI/CD) and continuous training (CT).
4. Validate on use case from production - models are available but can also start from beginning - depending on availability of time.

References:

[1] Singh P. (2021) Model Deployment and Challenges. In: Deploy Machine Learning Models to Production. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-6546-8_2

[2] Y. Zhou, Y. Yu and B. Ding, "Towards MLOps: A Case Study of ML Pipeline Platform," 2020 International Conference on Artificial Intelligence and Computer Engineering (ICAICE), 2020, pp. 494-500, doi: 10.1109/ICAICE51518.2020.00102.

[3] Baier, Lucas; Jöhren, Fabian; and Seebacher, Stefan, (2019). "CHALLENGES IN THE DEPLOYMENT AND OPERATION OF MACHINE LEARNING IN PRACTICE". In Proceedings of the 27th European Conference on Information Systems (ECIS), Stockholm & Uppsala, Sweden, June 8-14, 2019. ISBN 978-1-7336325-0-8 Research Papers. [Online] Available: https://aisel.aisnet.org/ecis2019_rp/163

[4] "MLOps: Continuous delivery and automation pipelines in machine learning," Google Cloud. [Online]. Available: https://cloud.google.com/architecture/mlops-continuous-delivery-and-automation-pipelines-in-machine-learning. [Accessed: 02-Jun-2021].

[5] Alla S., Adari S.K. (2021) What Is MLOps?. In: Beginning MLOps with MLFlow. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-6549-9_3