MLOps Challenges in Multi-Organization Setup: Experiences from Two Real-World Cases

Tuomas Granlund; Aleksi Kopponen; Vlad Stirbu; Lalli Myllyaho; Tommi Mikkonen
(https://arxiv.org/pdf/2103.08937.pdf)https://arxiv.org/pdf/2103.08937.pdf

#Organisational Boundary! INTRO

- Implications of organisational boundaries have not been widely considered in the context of Artificial Intelligence(AI) and Machine Learning(ML).
  Rather, work has fostered in areas where large datasets have emerged and are available for use inside one organization, for example, Google, Amazon etc.
  However, for cases such as AuroraAI, these boundaries are an enormous obstacle for using ML techniques to their fullest potential.

- Cases, where the state would own all the data, different organizations that host it, may have their own privileges and responsibilities.

- For cases that involve legal entities, say private companies and non-governmental, non-profit organisations, and their collaborations, the boundaries can become even more apparent, as government data might not be shareable with them.

- Integration mechanisms for AI/ML, similar to integration patterns in Information Systems, but applicable to AI/ML features, need improvements. This leads us to the need for better continuous deployment(MLOps).

#BACKGROUND || Overview of MLOps!

- The goal of continuous deployment is to enable continuous flow of value adding software artifacts from the development to the actual production use with a quality assurance.

- MLOps: Amalgamation of Machine learning and Operations refers to advocating automations and advocating at all steps of ML system deployment and development. Support is needed for integration, testing, releasing, deployment, monitoring, and infrastructure management.

- #steps necessary to deploy ML models:
  - Dataset divided on Training, Testing and/or Validation sets.
  - Hyper Tuning the selected model.
  - Training the model on the Training dataset.
  - Trained model can be validated on different data.

    - Once the criteria is met, the model is deployed.

- Continuous deployment of ML features is a complex procedure that involves taking into account application code, model used for production, and data used to develop the model. Adequate monitoring facilities are needed to ensure the successful operations.

- #CD4ML(Continuous Delivery for Machine Learning): The approach formalised by ThoughtWorks for automating in an end-to-end fashion the lifecycle of Machine Learning application:
    - In CD4ML, a cross functional team produces Machine Learning applications based on code, data and models in small and safe increments that can be reproduced and reliably released in any time in short adaptation cycles.This approach contains three different steps:
        - Identify and prepare the data for training.(ML pipeline)
        - Experimenting with different models to find the best performing candidate. (ML pipeline)
        - Deploying and using the selected model in production. (deployment pipeline)

- Three artifacts, in addition to those required by devOps, that need version controlling in MLOps :
    - Different data sets used for training models and their versioning.
    - Models and their versioning.
    - Monitoring the output of the model to detect biases and other problems.

-

#Case Study || I : Integration Challenges regarding ML !
(ORAVIZIO, implemented as SaaS that is used by two organizations.)

#About Oravizo: It is a software product that provides data-driven information about patient-level risks related to hip and knee joint replacement surgery. It has been developed in co-operation between the two actors: a hospital that specialises in joint surgery and Solita, a software development company based in Finland. The three predictive models by Oravizio are:

- Risk of infection within one year from surgery.
- Risk of revision within two years from surgery.
- Risk of death within two years from surgery.

- After data-preprocessing, and finding the relation between the risk for joint replacement surgery and the selected explanatory values; AUC values and ROC curves were evaluated and XGBoost was selected, and the final model for the product was built accordingly.

- All development artifacts, including scripts that were used to create ML models, were stored in Version Control System(VCS) except the ML model.

- The use of patient records is strictly regulated. Hence, the system was designed to be deployed in the production environment with its ML model in a locked state.

- Challenges:

    - DATASET: Use of datasets is often associated with ownership issues:
        - Firstly, large scale datasets are often not possible to locate to other storage.
        - Secondly, even if it is possible, owners may not want it.
        - And, there are regulatory issues, such as in healthcare, to keep the data in one organization.

      (In this situation, the service provider does not generate new data but only uses the model. New dataset, which is generated from treating patients only, which is under the responsibility of the hospital, and is not visible to the service provider.) The downside is, this approach restricts tracking down how successful the decisions were when no operation is performed.

- MODEL: Model acts as an interface between hospital and the service provider. The model is deterministic in nature.
  - Firstly the hospital trains the model, selects the one best suited for the purpose.
  - Secondly, the software service provider uses the model as the basis for collaboration between the hospital and patients.

- MONITORING: Since the model does not evolve over time, it continues producing the same results. Therefore, no drifts can emerge overtime. Unless the hospital gains new insight from the dataset and uses it to re-train the model. However, this requires human precision and no such monitoring.

-

#Case Study || 2: Scaling of ML to multi organization Context !
(AuroraAI initiative.)

#AuroraAI is a government initiated program with an objective to create the world's best public administration. By removing organisational silos that complicate serving citizens in many ways, AuroraAI network helps determine which individuals or businesses are in need of a particular service. The model used by AuroraAI, highlights the importance of going beyond GDP or production, and tackling the most difficult task of measuring people's well being.

- Achieving the goal requires information exchange and interoperability between different products and services.
    - DATASETS:
        - MyDataGlobal is working to support collaborations between different entities, with interest in building human centric personal data, where it is the individuals who combine the data and not the society.
        - The use of Digital Twin Paradigm has also been considered leading to citizen level use of datasets and recommendations
        - Unfortunately, the data that only individuals can release in accordance to their wishes, may not be obvious which data is True which is False.

        .
    - MODEL:
        - Relying on individuals' data, may introduce considerable bias.
        - It would be possible to build systems so that the ML systems are combined, like pipes and filters architectural style, with one ML system taking as input the output of another. Unfortunately, it is

considered preferable training one, single model on a combined data set rather than two models for two different roles.

- Models originating from different sources might not be compatible.

- For AuroraAI, instead of aiming at automata that can provide recommendations for everyone, models are more targeted to individuals who can use them to determine facts about their well being.

- MONITORING:
  - Monitoring mechanism regarding data access, escapes from an individual citizen and, to gather such information, the offices responsible for hosting the data should be involved. Hence, there is a clear lack of governance in today's systems.
  - Individual offices have this system placed locally, they can monitor what takes place and who has access to what!
  - Opening such monitoring data to individuals wrt their own use would probably result in increasing confidence in the use of private, personal data.

#Discussions regarding main observations, directions for future research || PROPOSAL!

In successful multi-organization MLOps, we need patterns of integration that help us in the process. Inspiration of this can be found both from system integration and legality patterns.

- Oravizio uses the ML model as EVALUATOR. The organizations work in close cooperation and have shared interest in model development and maintenance in the Oravizio service.

- In AuroraAI, USER DELEGATION helps to combine data that can only be accessed by the user as a whole. The organizations are looking over the organizational borderlines to identify best practices that could be used in other contexts, and, furthermore, some of them have executed joint initiatives for certain specific goals, such as well-being at school in a certain region.

#Final Conclusion

In this paper, two cases were presented, where ML is used in a setup where several organizations are involved. The highlighted challenges are related to integration and scaling, with real-life cases showing how ML has been addressed. As pointed out, in both cases the teams have found practical solutions, but they are applicable to these cases only. As a more general solution, using patterns that help dealing with inter-organisational boundaries, as well as using joint operational mode across the organizations that are involved, are proposed. However,

to a large degree, this remains as a piece of future work, despite the fact they have identified such in the presented cases.