

PuzzleUp 2024

MLOps TechLab

Rebecca Hiller, Adrian Bader und Iwan Imsand
hiller@puzzle.ch, bader@puzzle.ch und imsand@puzzle.ch

Inhalt

- Vorstellungsrunde
- Was ist MLOps?
- TechLab
 - Vom Prototyp zum pipeline-ready Skript
 - Pipeline
 - Datenversionierung
 - Metriken und Experimente
 - Pipeline-Ausführung mit GitHub-Actions
 - Modell manuell testen
- Abschluss

Vorstellungsrunde

Wer sind wir?

Was machen wir?

Wer seid ihr?



Iwan Imsand
Software Engineer /
Member of the
Technical Board

ML Enthusiast, unser Spezialist
für ML- und MLOps-Projekte



Rebecca Hiller
Software Engineer
Umweltnaturwissenschaftlerin,
die ihren Weg über Data
Science in die
Softwareentwicklung gefunden
hat



Adrian Bader
Software Engineer
Technik Interessierter, der ML
vermehrt in der Praxis
anwenden möchte



Andreas Rava
Head of Business
Development &
Innovation
Brückenbauer - der das
Potential von AI zusammen mit
dem Kunden identifiziert



Digital Transformation & Software Development

Wir unterstützen dich von der Projektidee über die Architektur bis zur Applikation in der Cloud.



Infrastructure & Cloud

Wir entwickeln Infrastruktur- und Cloudlösungen für mehr Flexibilität und Skalierbarkeit.



Data Analytics & AI

Wir transformieren DevOps zu MLOps und begleiten beim Aufbau von MLOps-Prozessen.



IT-Security

Wir kümmern uns um die Cyber-Sicherheit der Daten und IT-Systeme in deinem Unternehmen.



Mobility IT-Services

Wir kreieren und unterhalten innovative IT-Lösungen für die Mobilitätsbranche.



Solutions

Wir bieten umfassende IT-Lösungen und spezialisierte Softwarelizenzen.

Data Analytics & AI



Beratung

Strategieanalyse und Architekturberatung



Projekte

Projektrealisierung von Analytics und AI-Vorhaben



Infrastruktur

Aufbau von Plattformen und Pipelines für Data Analytics und AI



Schulungen

Kurse & Trainings zum Thema ML und MLOps

Wer seid ihr?

- Name
- Wo arbeitest du?
- Wo siehst du Potential für ML und MLOps in deiner Arbeit?
- Was erwartest du vom heutigen Workshop?

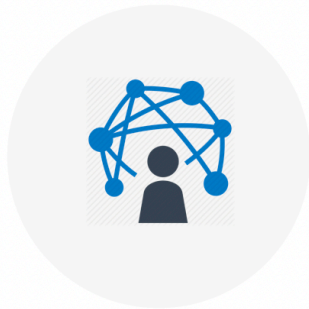
Was ist MLOps und warum ist es wichtig?

Wo kann ML zu einem Mehrwert führen?

- Large Language Modell (Content Generierung)
- Spracherkennung und -verarbeitung
- Textanalyse (Spam-Detection oder Erkennung der Gefühlslage bei Texten)
- Prediction (Erkennung von Abnutzung für optimierten Unterhalt)
- Bilderkennung / Klassifikation (Prognosen von Schadenssummen)
- Deep Learning (Autonomes Fahren)

Worauf warten wir!

Problem verstehen



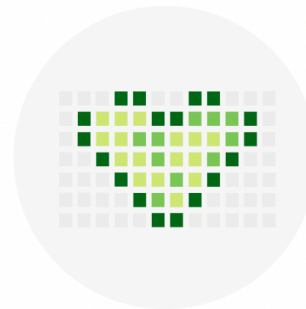
Daten sammeln &
verarbeiten



Daten analysieren

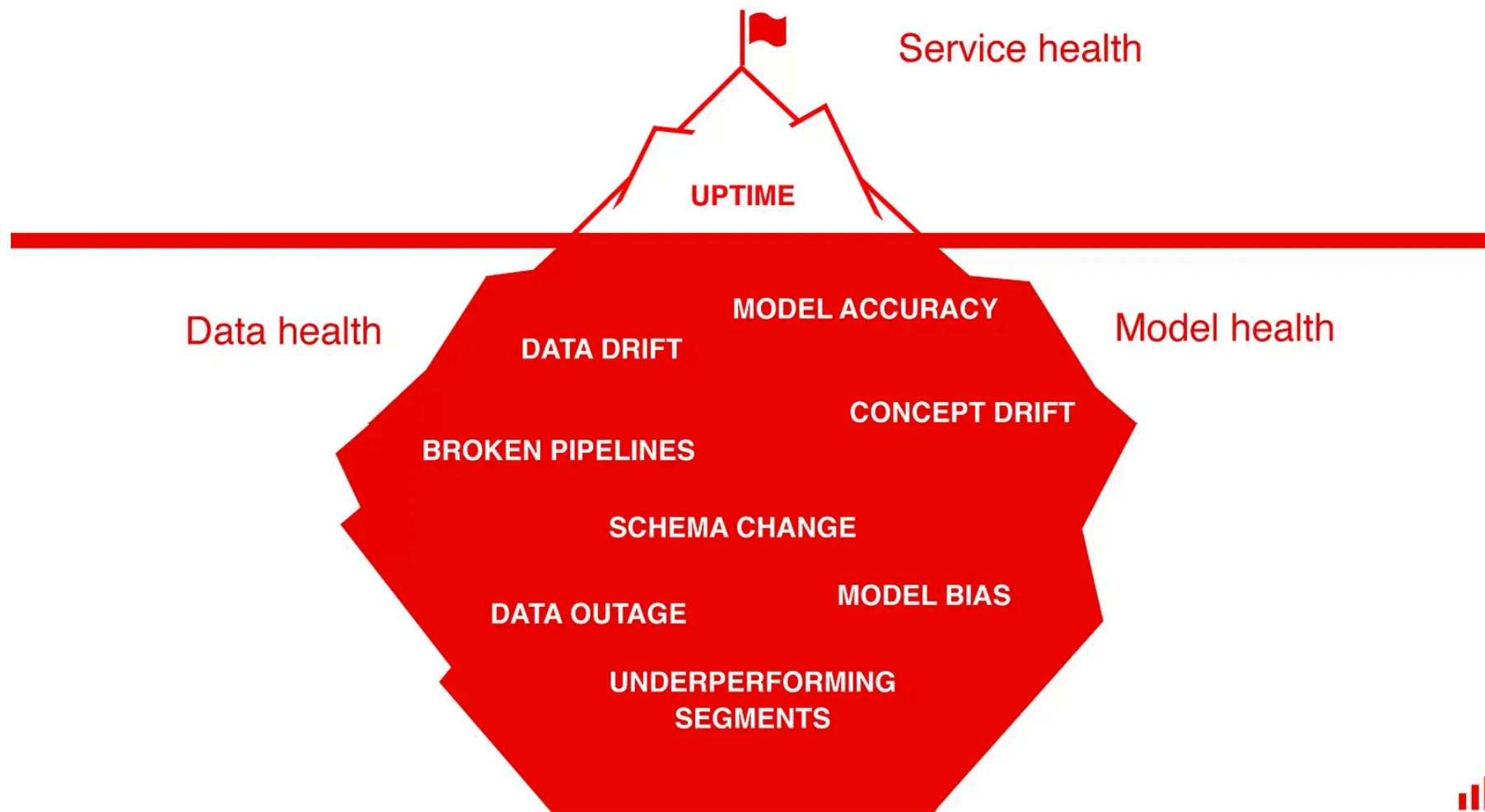


Modell trainieren



Deployen

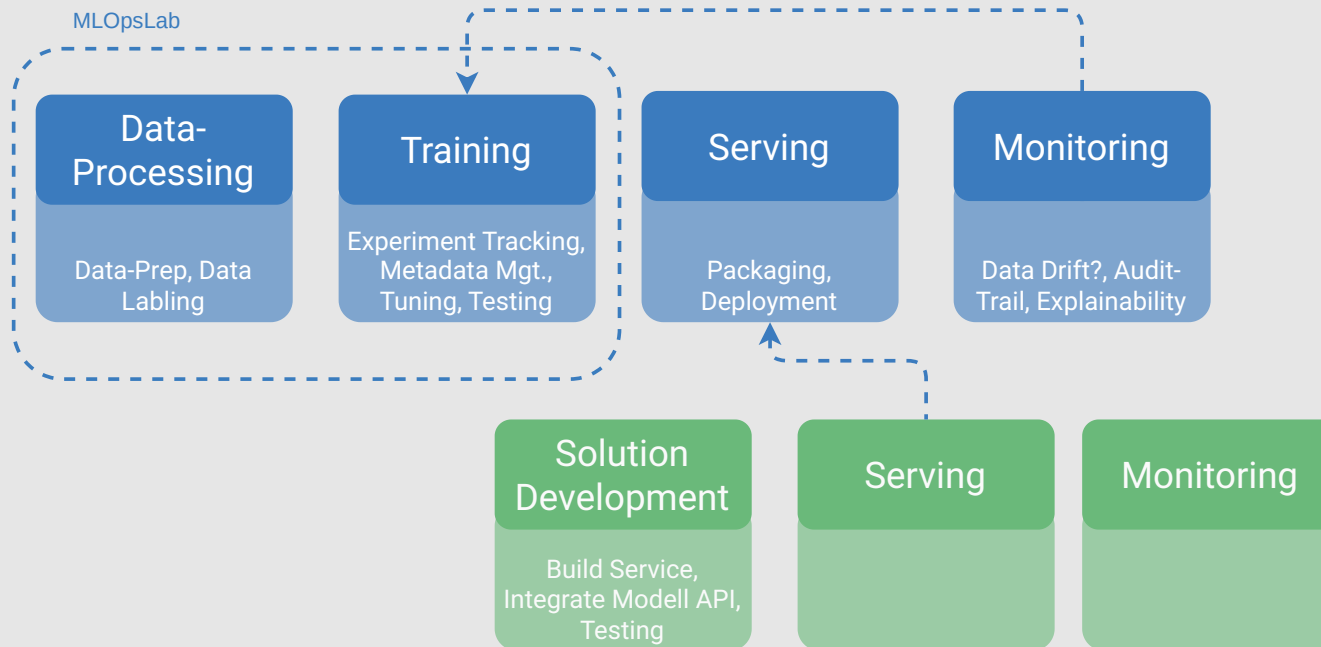




MLOps Prozess und Lab Fokus

Der MLOps Prozess

MLOps DevOps



MLOps TechLab

TechLab: Vom Jupyter Prototyp zum reproduzierbaren ML Experiment

Anwendungsfall:

Handgeschriebene Zahlen erkennen (MNIST)

Dokumentation des Modells:

https://scikit-learn.org/stable/auto_examples/classification/plot_digits_classification.html

Abgrenzung

ML KnowHow über Algorithmen & Funktionsweise ist **nicht** Teil des heutigen Workshops... 😞

-> separates Training: ML für Engineers

<https://www.puzzle.ch/trainings/ml-fuer-software-engineers>



Vorstellung Jupyter Prototyp

Vom Prototyp zum pipeline-ready Skript

Die Schritte sind im TechLab einzeln aufgeführt und beschrieben. Aufgrund der zur Verfügung stehenden Zeit werden wir sie nur zusammen anschauen.

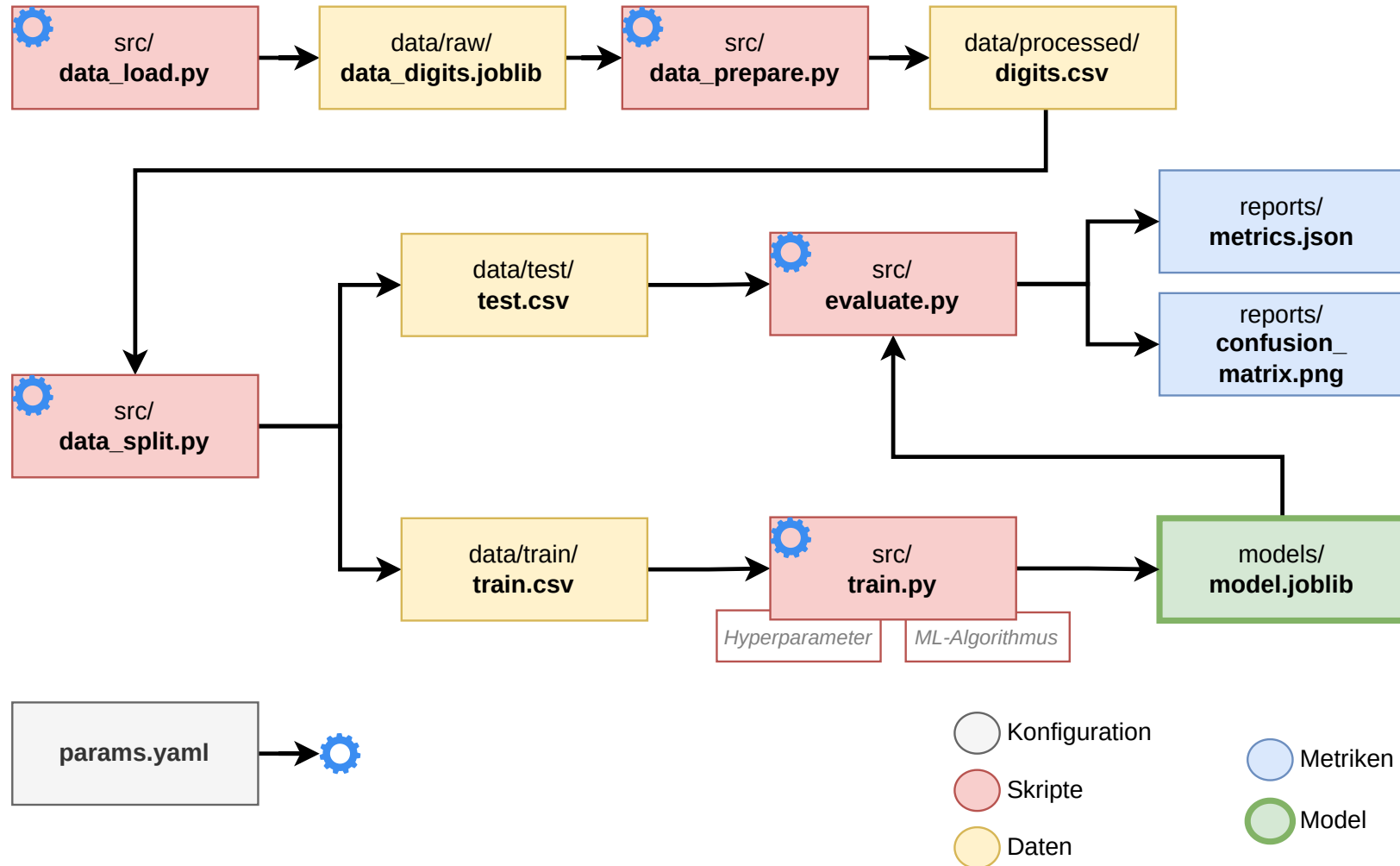
Hands-on folgt später 🤗 🙌

Vom Prototyp zum pipeline-ready Skript

1. Artefakte als separate Dateien speichern 
damit wir sie später weiterverwenden können
2. Skript konfigurierbar machen 
damit es mit verschiedenen Parametern ausführbar ist
3. Schritte aufteilen 
damit nicht immer alles ausgeführt werden muss

Pipeline

- Einzelschritte zusammenführen • Was muss ich in welcher Reihenfolge machen?
• Was muss ich bei einer Änderung alles ausführen?

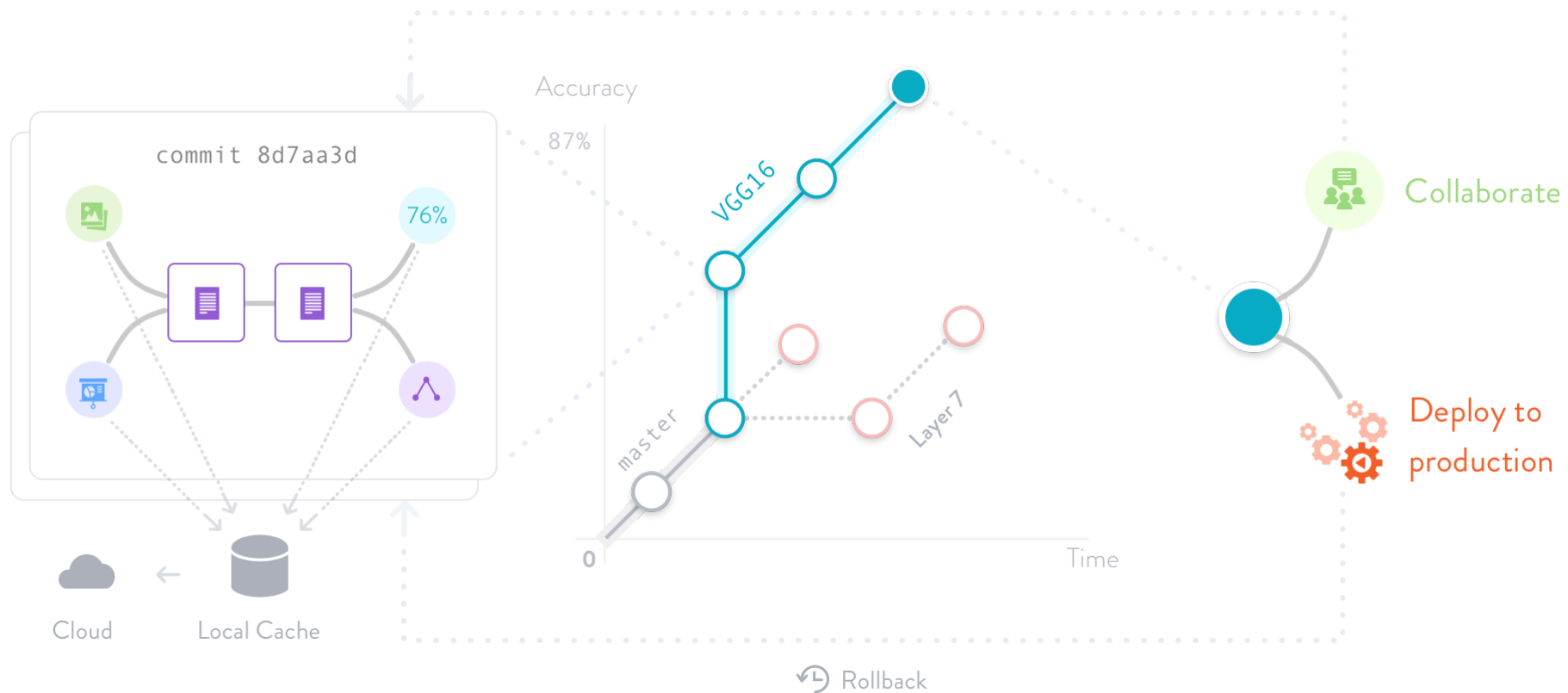


DVC (Data Version Control)

DVC ist ein OpenSource Tool von dvc.ai für:

- ML-Projekt Versionskontrolle
- ML-Experiment Management
- Unterstützung beim Deployment und der Zusammenarbeit

<https://dvc.org/>



ML project version control

ML experiment management

Deployment & Collaboration

Einführung Hands-on Labs

Umgebung aufsetzen

- Umgebung vorbereiten:
 - https://github.com/iimsand/mlops-techlab/blob/main/labs/001_lab_environment.md
- Prototyp ML-Experiment
 - https://github.com/iimsand/mlops-techlab/blob/main/labs/010_lab_initial_prototype.md

MLOps TechLab (Hands-on)

Pipeline

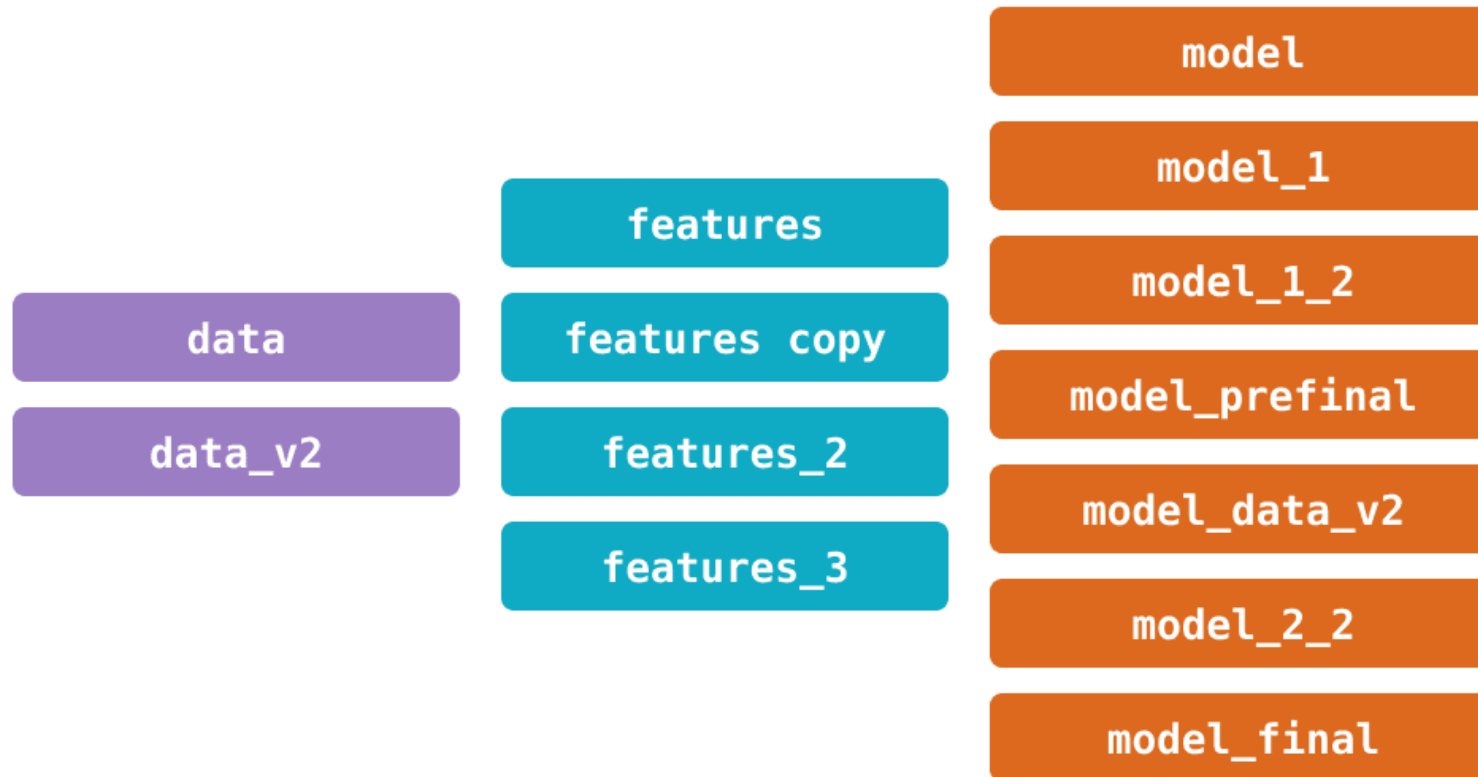
- Pipeline zusammenstellen:
 - https://github.com/iimsand/mlops-techlab/blob/main/labs/020_lab_init_pipeline.md

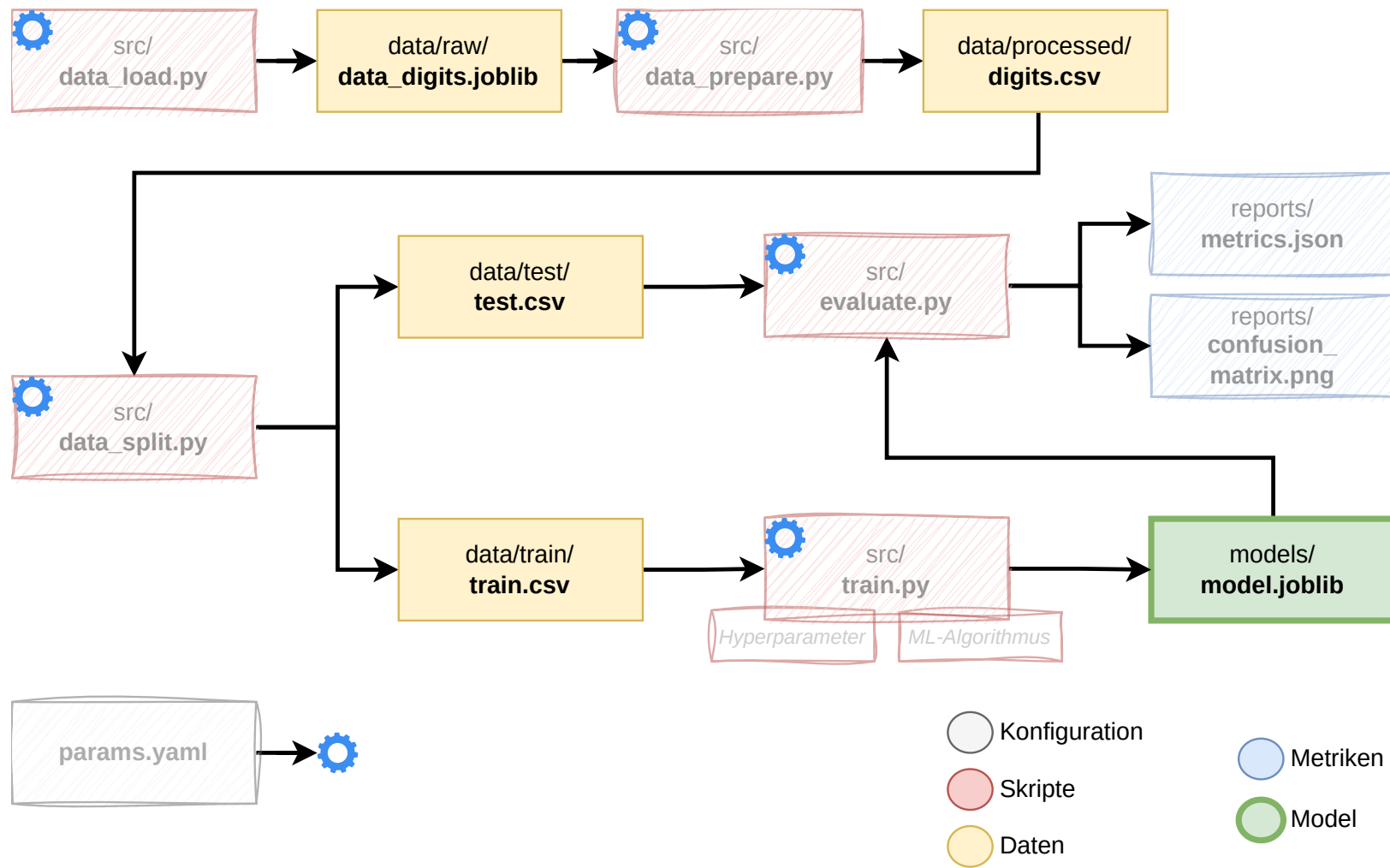
Datenversionierung

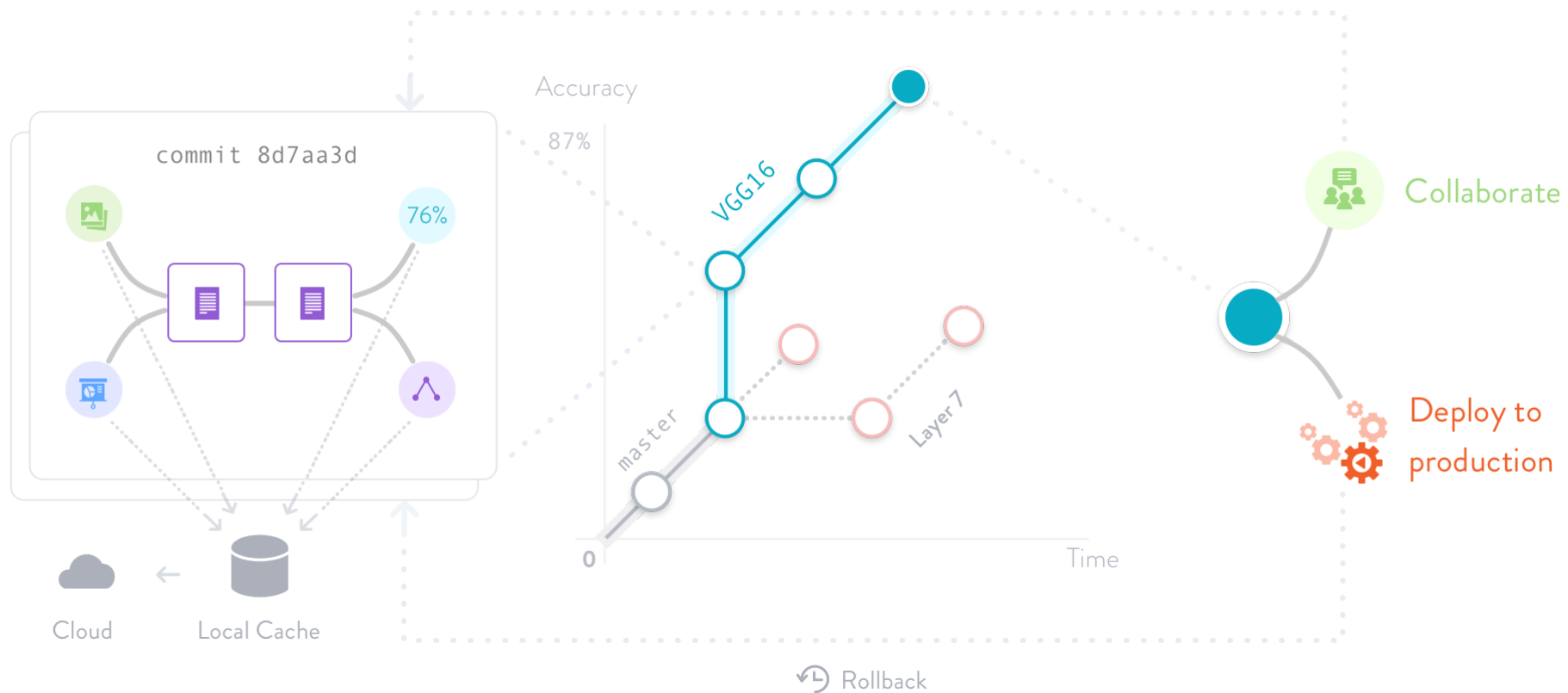
Konsistenz und Reproduzierbarkeit:

- Mit welcher Version des Datensatzes habe ich mein Modell trainiert?

Exponentielle Komplexität von Data Science Projekten





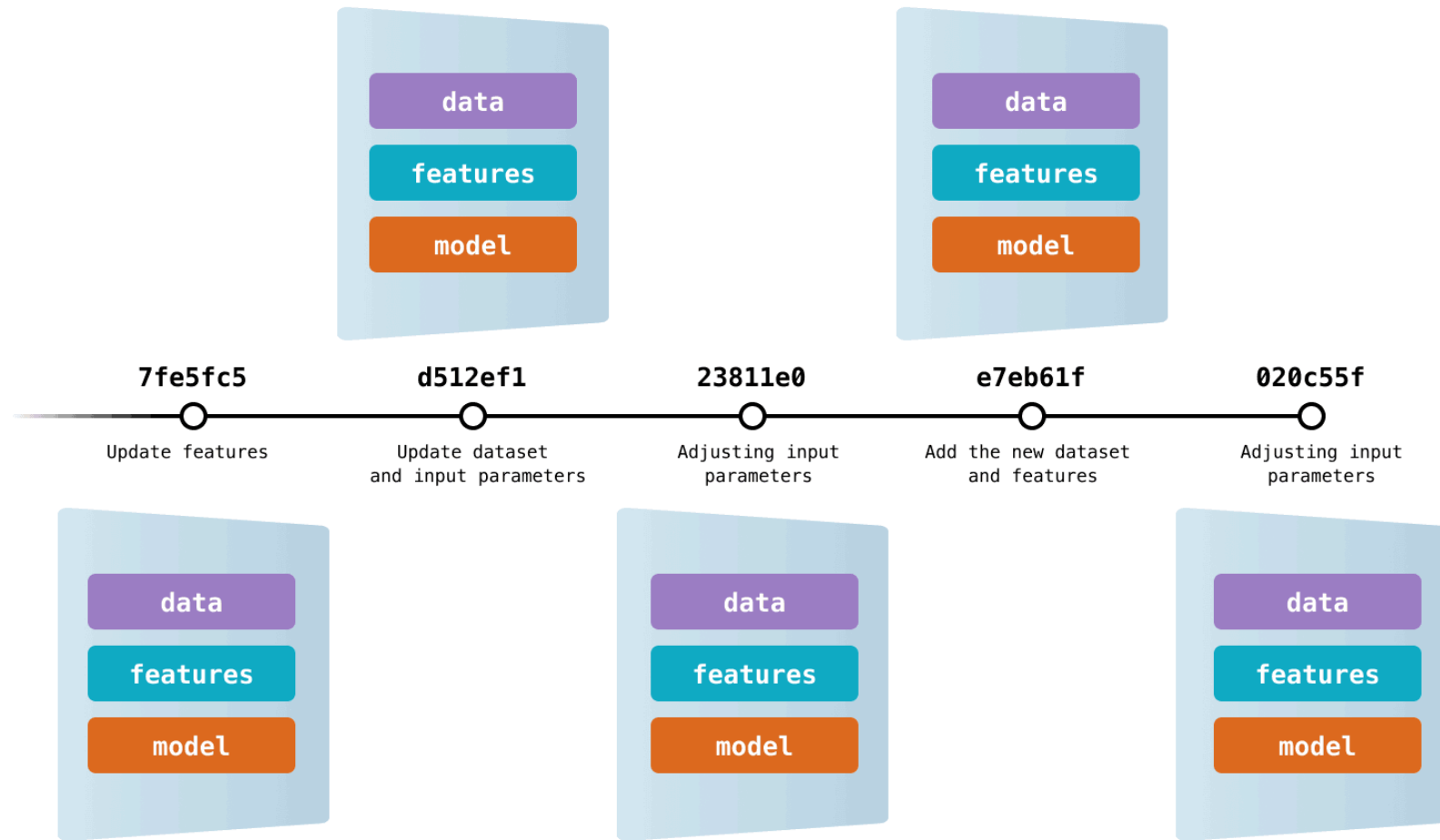


ML project version control

ML experiment management

Deployment & Collaboration

DVC matcht die richtige Version Daten, Code und Modell für dich 💖.



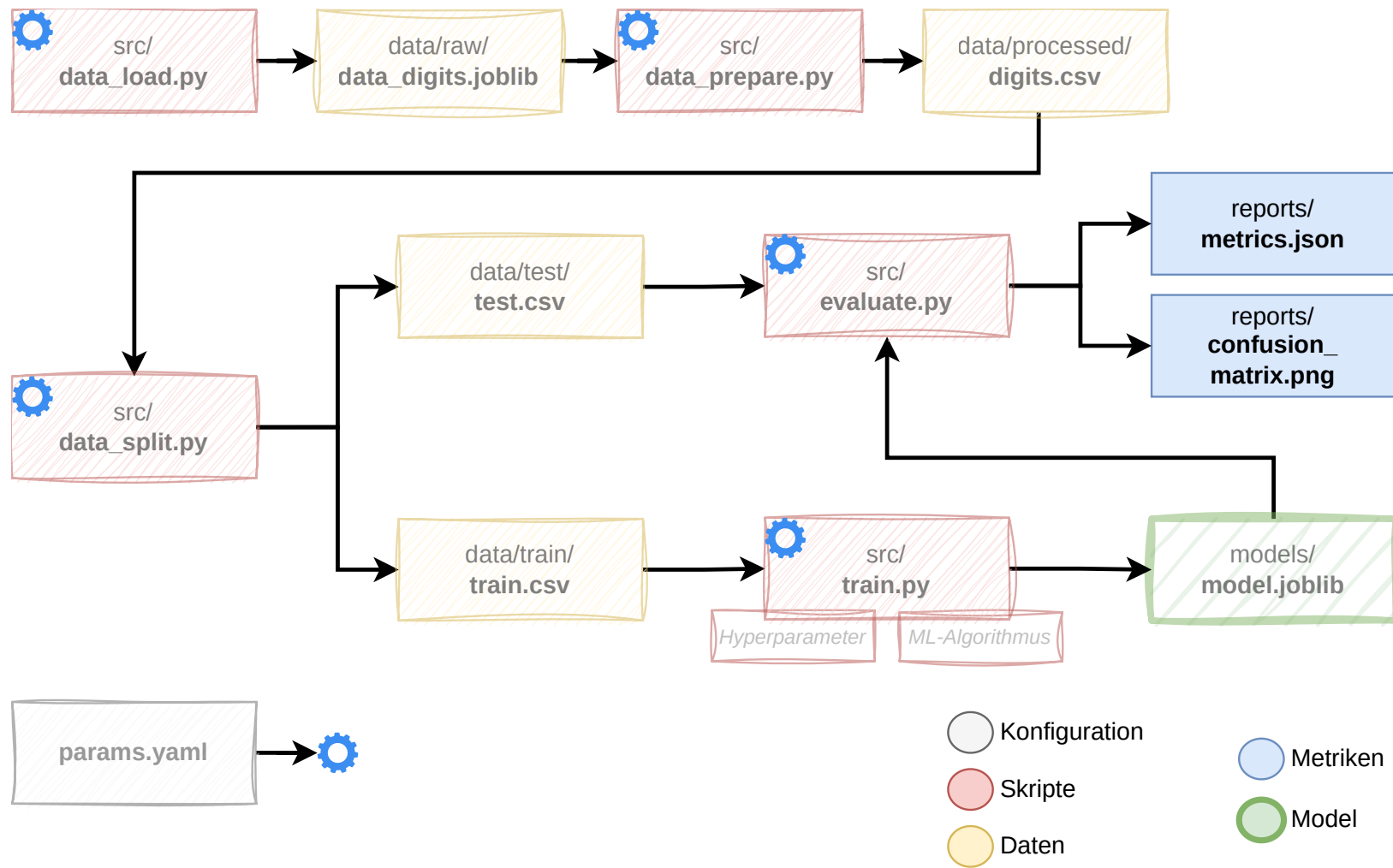
Datenversionierung mit DVC

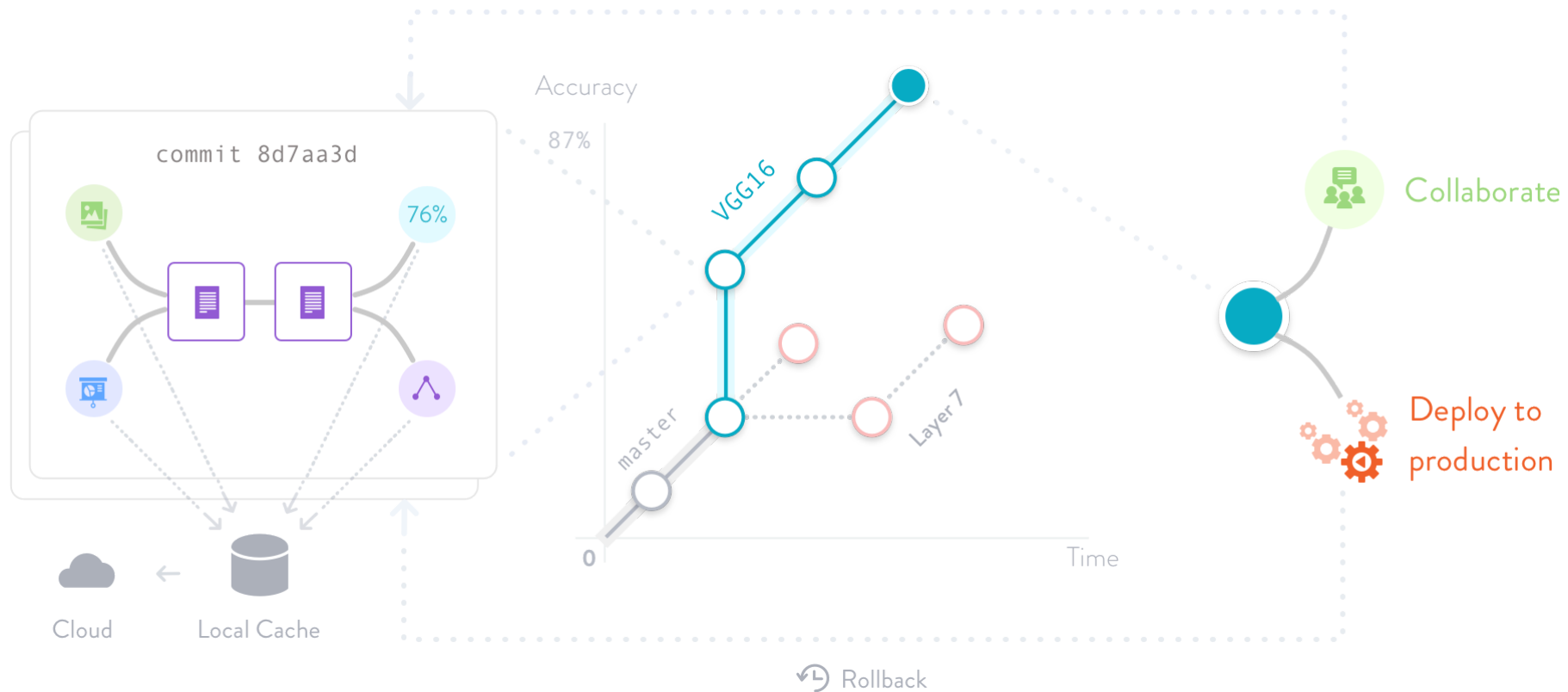
- Daten und Modell versionieren:
 - https://github.com/iimsand/mlops-techlab/blob/main/labs/030_lab_data_versioning.md

Metriken und Experimente

Modellgüte verfügbar machen und Experimente ausführen:

- Ist mein neues Modell besser als das alte?
- Verbessert sich ein Modell wenn ich Parameter verändere?



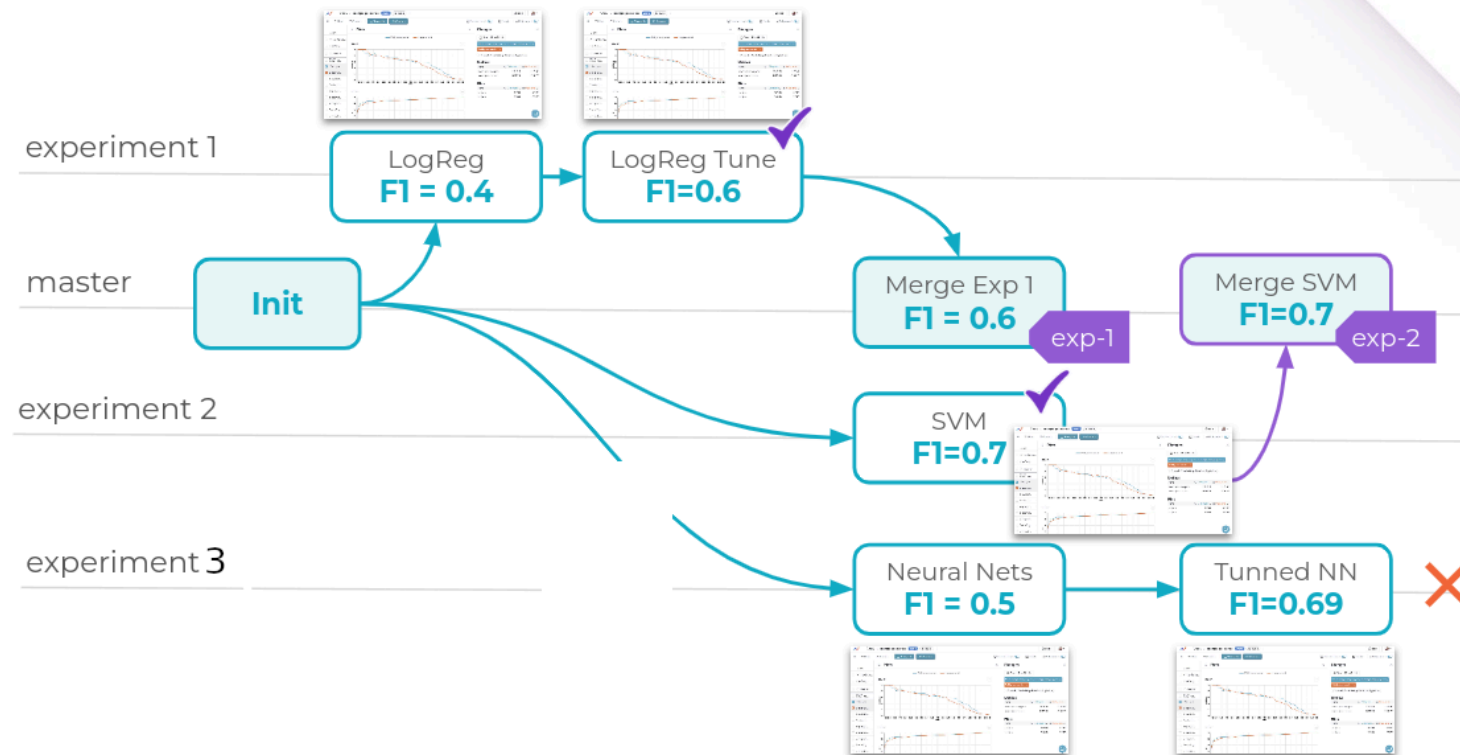


ML project version control

ML experiment management

Deployment & Collaboration

Apply Git based workflows to ML development



Metriken und Experimente

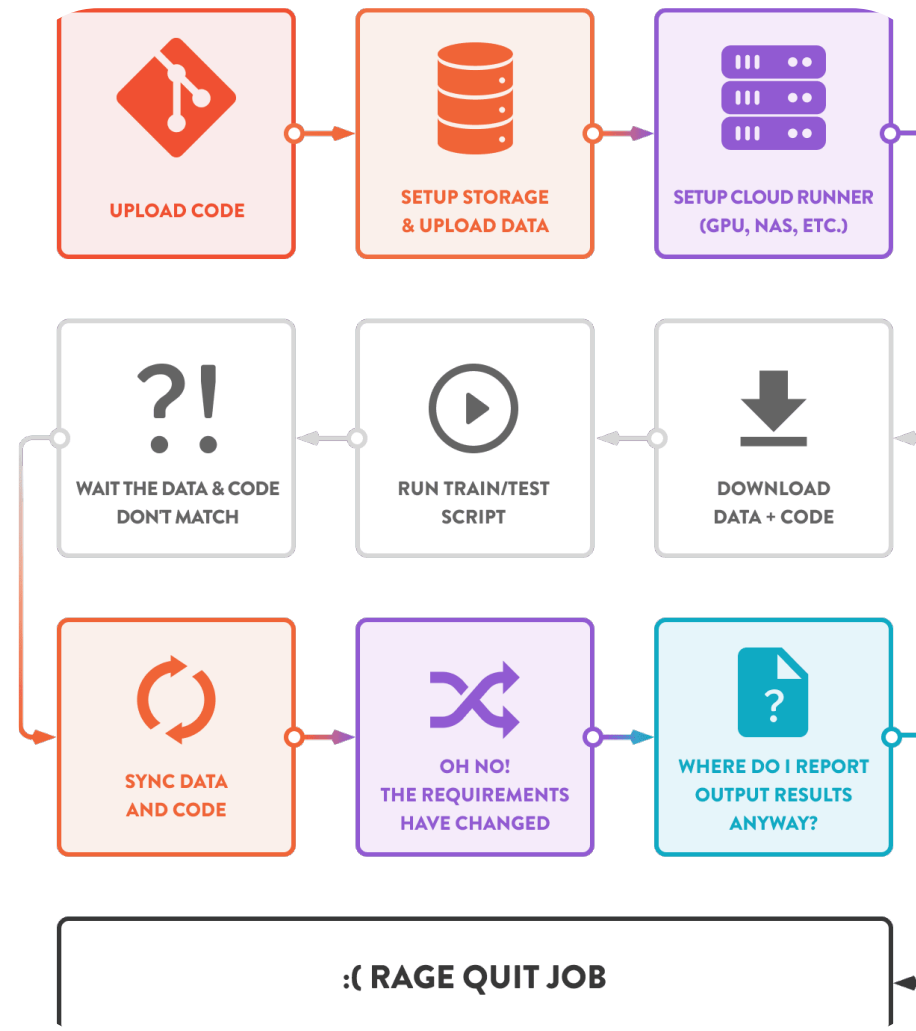
- Metriken und Experimente:
 - https://github.com/iimsand/mlops-techlab/blob/main/labs/040_lab_metrics.md

Pipeline-Ausführung mit GitHub-Actions

Nachvollziehbarkeit und Qualität:

- Hat die Änderung von meiner **Kolleg:in** das **Modell verbessert**?
- Wie hat sich das Modell durch die Anpassung **verändert**?
- **Review** mit **Team** oder **Kolleg:in** vor Integration

Basic Pipeline



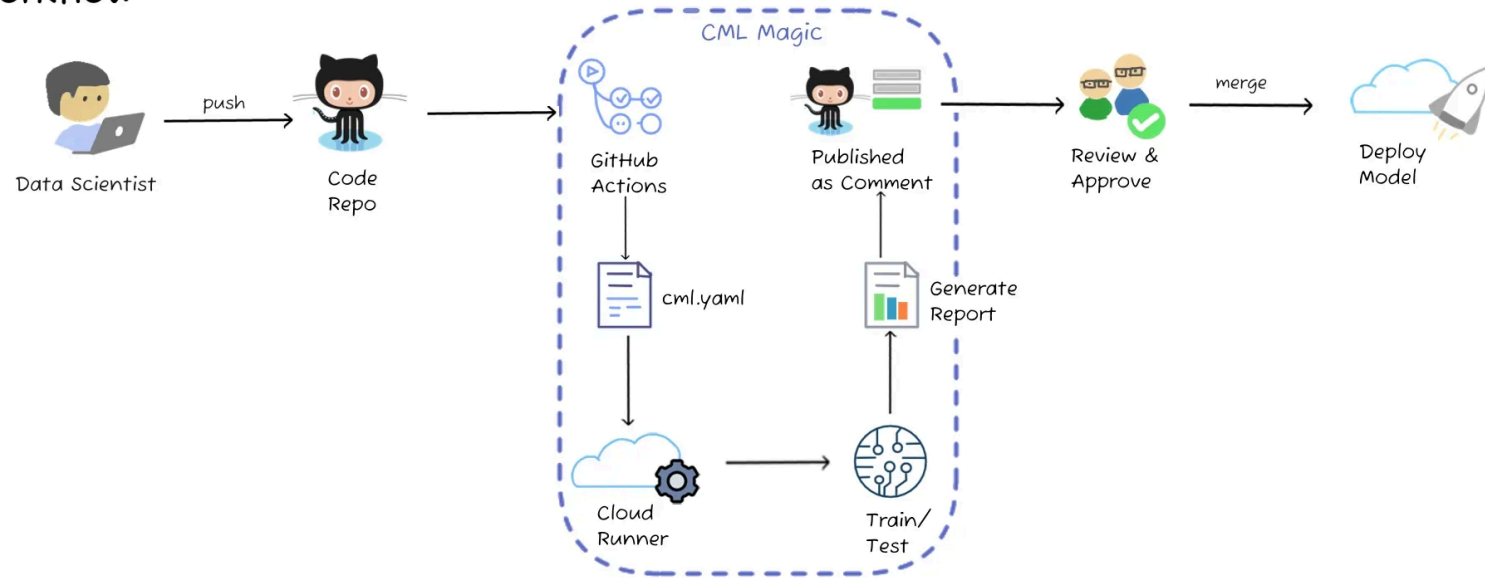
CML (Continuous Machine Learning)

CML ist ein OpenSource Tool von iterative.ai für:

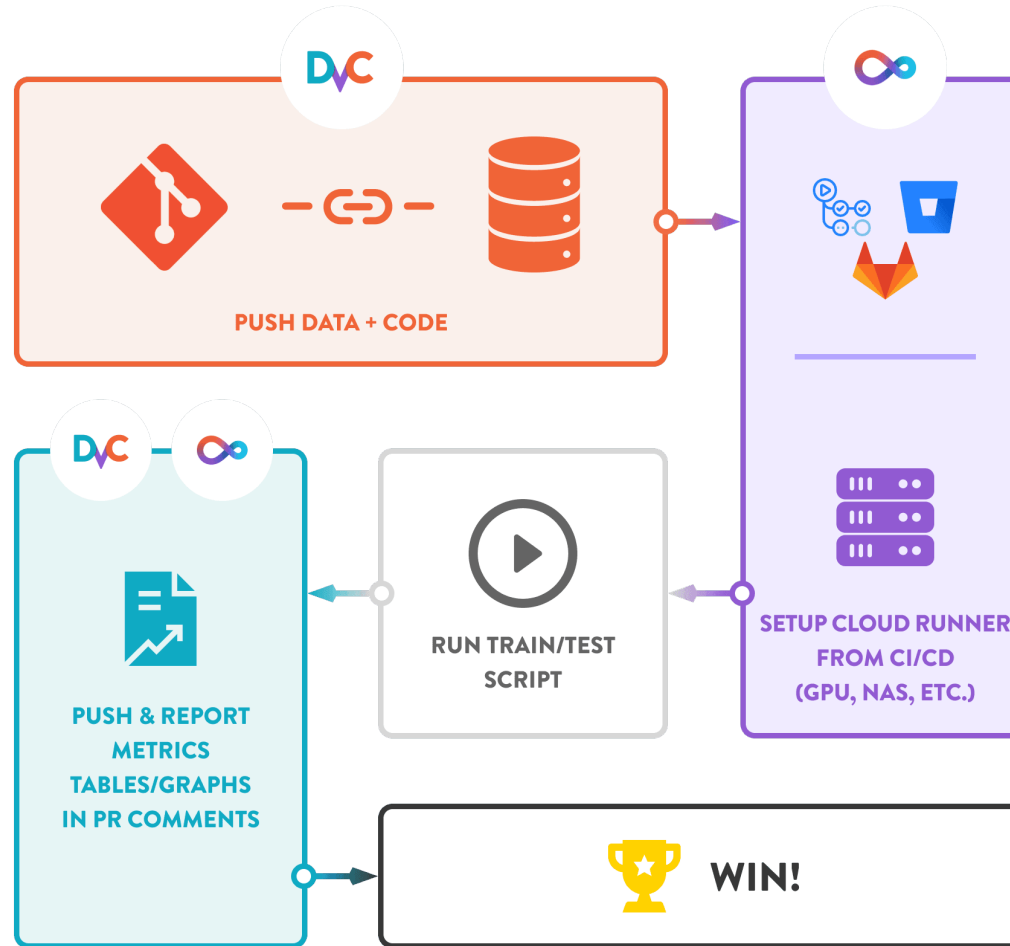
- ML in CI/CD Workflows
- Berichterstellung mit z.B. ML-Metriken
- Verwaltung von Cloud-Ressourcen für ML Experimente

<https://cml.dev/>

CML Workflow



Traditional ML meets CI/CD



Pipeline-Ausführung mit GitHub-Actions & CML

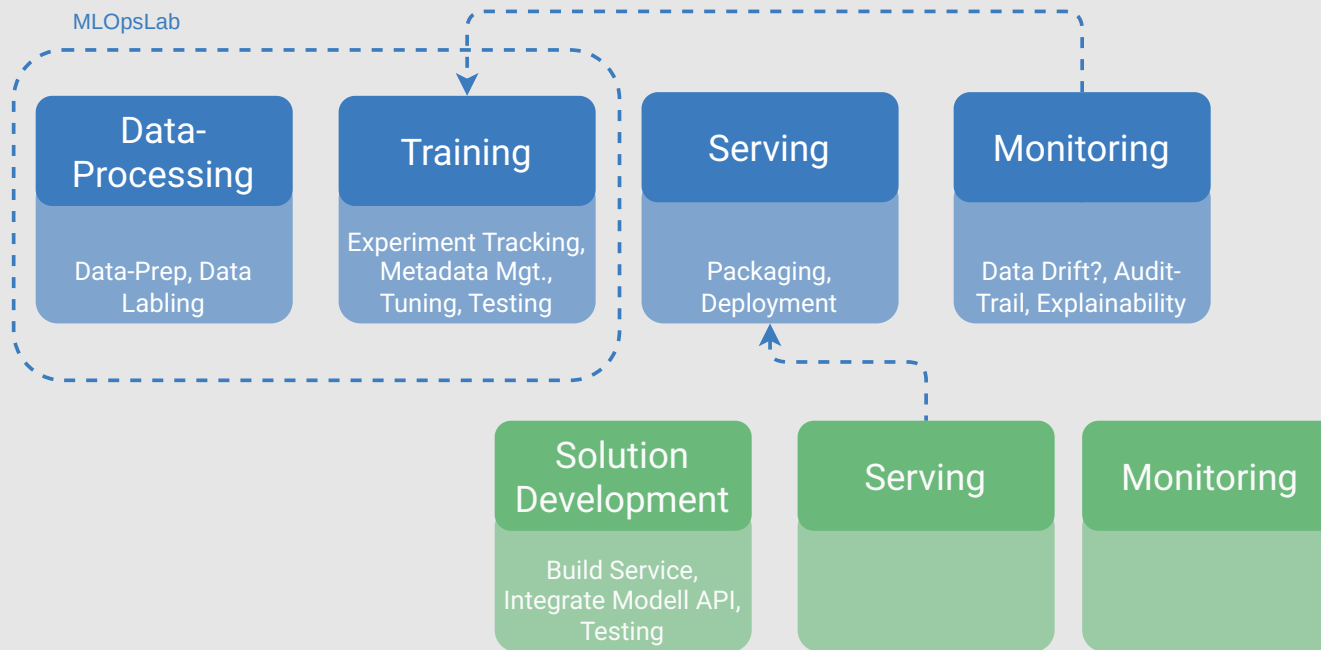
- GitHub Actions:

https://github.com/iimsand/mlops-techlab/blob/main/labs/050_lab_github_actions.md

Modell deployen

Der MLOps Prozess

MLOps DevOps



Einbindung im TechLab

Modell in der Applikation laden und einbinden.

params.yaml:

```
train:  
  model_path: models/model.joblib
```

Modell manuell testen

- Modell manuell testen:

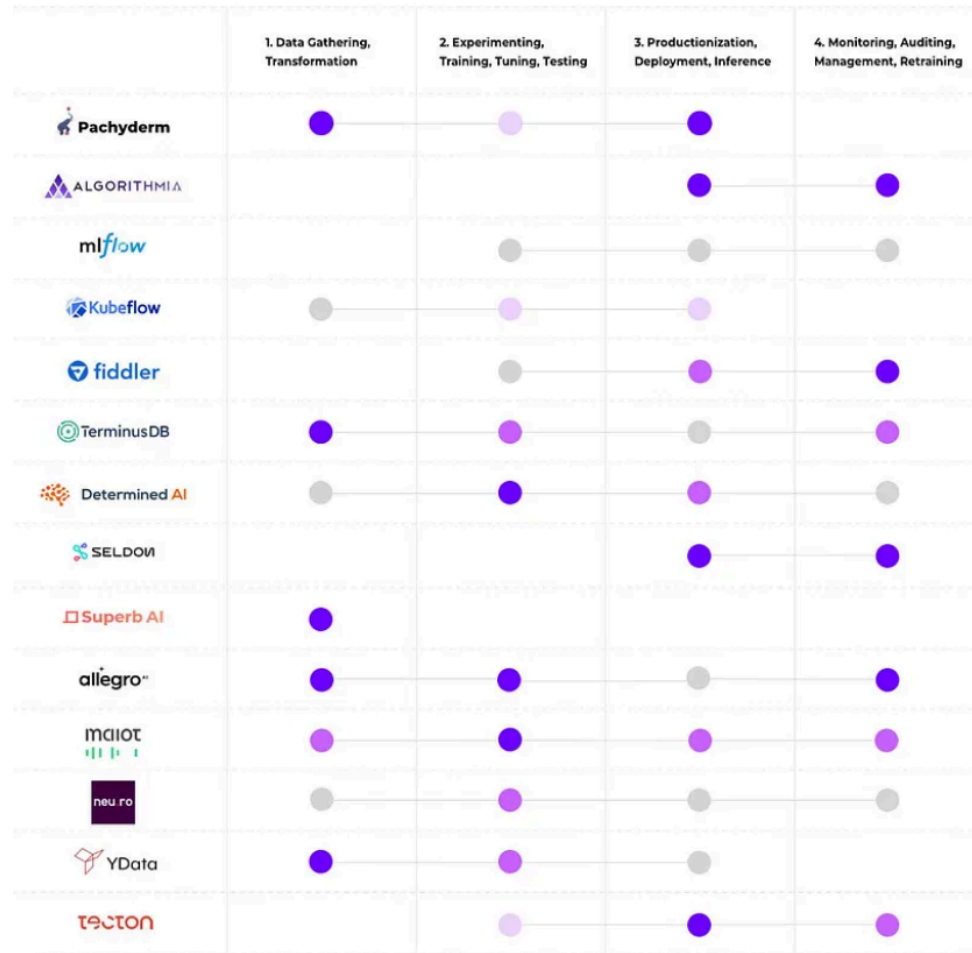
https://github.com/iimsand/mlops-techlab/blob/main/labs/060_lab_deploy_model.md

Abschluss

Tools und Plattformen

Machine Learning Life Cycle

● Highly Developed ● Moderately Developed ● Still Developing ● Lightly Developed



Quelle: Dan Jeffris, Pachyderm (Okt, 2020)

Fazit zu DVC/CML

- Leichtgewichtiges Toolset
- Direkte Integration in bestehende Git-Workflows
- Essentielle Funktionen für ML-Workflow-Management ohne komplexe Infrastruktur

Fragen

?

Feedback

- Was hat dir am Lab gefallen? Was wünschst du dir noch?
- Kannst du das neu erlangte Wissen im Arbeitsalltag einsetzen?
- Interessieren dich weitere Themen resp. Technologien rund um MLOps?

Merci

Mehr Informationen zu ML und MLOps bei Puzzle:
<https://www.puzzle.ch/service-solutions/data-analytics-ai>

