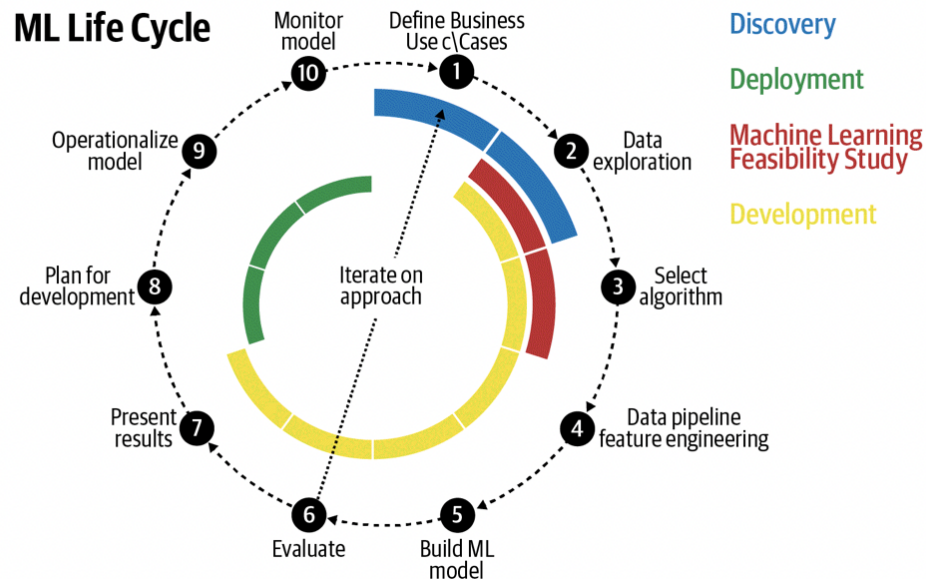


Agile Data Sprints

Bilden Sie ein Team aus zwei bis drei Teilnehmern.

Ihr Team ist ein hipbes Startup und implementiert in agilen Data Sprints: Simulieren Sie den Ablauf eines ML Lifecycles:



Analysieren Sie zu Beginn in der **Discovery** Phase eines der vier Datasets:

Schwierig (Aktive Kaggle Competition):

<https://www.kaggle.com/competitions/open-problems-single-cell-perturbations>

Schwierig (Abgelaufene Kaggle Competition):

<https://www.kaggle.com/c/tensorflow-great-barrier-reef>

Zur abgelaufenen Kaggle Competition finden Sie bereits Source Code unter:

<https://www.kaggle.com/competitions/tensorflow-great-barrier-reef/code>

Konzipieren Sie für eine dieser Submissions einen ML Lifecycle.

Mittel

https://fhstp.s3.eu-central-1.amazonaws.com/github_timeline.csv

Leicht

<https://fhstp.s3.eu-central-1.amazonaws.com/nytimes+front+page.csv>

In der **Machine Learning Feasibility Study** Phase erstellen Sie zur Präsentation ihrer Zwischenergebnisse eine Web-App mit [streamlit](#) oder [mercury](#). Setzen Sie dabei geeignete Components zur Visualisierung ein und ermöglichen Sie dem Fachbereich eine (einfache) Interaktionsmöglichkeit.

Nachdem der Fachbereich grünes Licht gegeben hat, startet die **Development** Phase. Achten Sie auf die Einhaltung der im Rahmen der Vorlesung besprochenen ML Design Patterns. Beleuchten Sie die unterschiedlichen Herausforderungen der zur Verfügung stehenden Datasets.

In der **Deployment** Phase konzipieren Sie eine ML Pipeline mit einem im Rahmen der Vorlesung vorgestellten Tools. Beleuchten Sie auch die Vorteile Ihres Deployments mit einer Antwort auf folgende Frage: Kann jeder einzelne Schritt *containerized* werden?

Rahmenbedingungen:

Wie Sie wahrscheinlich bereits gesehen haben, gibt es explizit keine Anforderungen oder andere *Business Cases*.

Lassen Sie Ihrer Kreativität - bzw. dem zu Grunde liegenden Dataset - freien Lauf! Der Fokus des Projekts liegt auf der Dokumentation des ML Life Cycles. Nicht in der Qualität Ihrer Vorhersagen. Ziel soll ein Datenprodukt sein, das auch im produktiven Einsatz belastbare Ergebnisse liefert und strukturiert im Team erarbeitet wurde. Dokumentieren Sie Ihre im Team getroffenen Entscheidungen!

Abgabe:

Erstellen Sie eine Präsentation (max. 10 min), in der Sie einen Überblick über die Applikation geben.

Laden Sie die Präsentation und den Source Code (NICHT NUR ein Jupyter Notebook, Colab, etc. ;-) Ihrer Abgabe bis zum 05.10.2023 einmal pro Team im eCampus hoch.