

# **Project Assignment: DataOps/MLOps Concept Development and Prototype Implementation**

**Instructor: Alexandros.Nanopoulos@Mosbach.DHBW.de**

---

## **Objective:**

This project assignment is designed to provide hands-on experience in applying DataOps/MLOps principles to a real-world machine-learning-based application. You will be required to conceptualize a comprehensive DataOps/MLOps pipeline and develop a functional prototype using open-source tools. This assignment encourages you to explore the full lifecycle of a machine learning model, from development and deployment to monitoring and maintenance.

---

## **Project Phases:**

### **1. Phase 1: Application Context and Data Selection**

- **Task:** Choose an application context for your project. Examples might include predictive maintenance in manufacturing, customer sentiment analysis in e-commerce, or anomaly detection in cybersecurity.
- **Deliverables:**
  - A brief report (1-2 pages) about business and data understanding. You can explain the reasoning behind selecting the particular application, the business problem you aim to solve, the main metrics you will use, and why an ML-based solution can be beneficial. You should also describe the data and potential challenges you anticipate regarding data quality and other aspects related to your application context.

### **2. Phase 2: DataOps/MLOps Pipeline Conceptualization**

- **Task:** Develop a detailed concept for the DataOps/MLOps pipeline. This should include, but not be limited to, the following aspects:
  - **Data Quality Assurance:** Describe the main points of a plan related to data quality assurance (what issues should be checked, how, etc.).
  - **Data Management:** Outline key requirements and a data versioning, storage, and management strategy.
  - **Model Development and Versioning:** Describe an approach to developing, training, and versioning/tracking the model(s).
  - **Continuous Training:** Propose a plan for continuous training in order to keep your model(s) accurate.
  - **Model Deployment and Prediction Serving:** Describe possible deployment strategies and patterns that are suitable for the given application. Also, describe the most suitable type of serving (e.g., batch or on-demand)
  - **Monitoring and Logging:** Propose a monitoring and logging approach for your model in production, covering model performance, data drift, and error tracking.

- **Deliverables:**
  - A comprehensive document (4-8 pages) that details the aforementioned aspects of the concept. This can include architecture diagrams, flowcharts, and a list of tools you plan to use. Please consider that the document describes a concept and not a concrete implementation. In this sense, you are allowed to describe parts of a possible solution, which may not become implemented in your prototype (see Phase 3). *Note:* It is possible to include additional aspects (e.g., a feature store and especially orchestration) in your concept, according to the requirements of your application.

### 3. Phase 3: Prototype Implementation

- **Task:** Implement a working prototype of the conceptual DataOps/MLOps pipeline (Phase 2) using open-source tools. Your prototype should cover the end-to-end process as much as possible. In this sense, it has to implement at least most of the following components of your concept:
  - **Data Quality Assurance**
  - **Model Development and Versioning**
  - **Model Deployment**
  - **Monitoring and Logging**

*Note:* the combination/orchestration of the implemented components can be developed either at a simple level (e.g., by using customized code that calls the components manually) or by using a workflow-management tool. To overcome possible technical problems, you are allowed to implement components separately (i.e., without connecting them in a workflow). In the latter case you have to document clearly the problems you faced.

- **Deliverables:**
  - A working prototype of the implemented components hosted on a cloud platform or local environment (e.g., notebooks), and – if possible – accessible to instructors for testing.
  - A technical report (3-5 pages) explaining the implementation details, challenges faced, and how they were overcome. The report should include a clear set of instructions for running a demonstrating example.
  - A brief presentation (5-8 slides) showcasing your prototype, its main features, its performance, and benefits.

*Note:* the presentation is just a **written** deliverable, no presentations will be held.

### 4. Phase 4: Evaluation and Reflection

- **Task:** Evaluate your prototype and reflect on the implementation process.
  - **Deliverables:**
    - A final report (1-2 pages) evaluating the strengths and weaknesses of your prototype and the effectiveness of the tools used.
    - A reflection on the lessons learned during the project, potential improvements, and how this project might scale in a real-world scenario.
-

### **Grading Criteria:**

- **Innovation and Creativity:** Originality in problem selection and prototype design.
  - **Technical Depth:** Depth of DataOps/MLOps concepts applied, the complexity of the implementation, and usage of tools.
  - **Functionality:** Working prototype with end-to-end functionality.
  - **Documentation:** Clarity, detail, and completeness of the reports and presentation.
- 

### **Submission and Deadline:**

- You can develop the project either individually or in teams of 2-3 students.
  - Please submit your deliverables as a single ZIP (or similar) file per email.
  - **Deadline:** TBD
- 

### **Support:**

The instructor will be available to discuss project progress and challenges. Appointments can be scheduled via email.

---

**Good luck, and we look forward to seeing your innovative solutions!**