

SimOpt – Projekt 2021/22

Gruppe D: Datengetriebene Modellierung

Szenario

Ein kleiner Start-Up aus dem Gebiet Machine Learning, wo Sie arbeiten, hat einen Auftrag von einem großen Industrieunternehmen bekommen. Sie möchten den Störungsbetrieb und nämlich das Strömungsregime „Slug-Flow“ einer multiphasigen Anlage erkennen zu können. Die Anlage in Abbildung 1 schematisch dargestellt. Bei diesem Strömungsregime sammelt sich die flüssige Phase am Boden des Steigrohrs an und dadurch wird der Gasstrom blockiert, was die Effizienz der Anlage enorm beeinträchtigt. Da für die Beschreibung der Erscheinung kein Modell gefunden werden konnte, hat Ihre Firma entschieden einen daten-getriebenen Ansatz zu verfolgen. Deshalb wurde die Anlage mit zusätzlichen hochfrequenten Ultraschall- und Drucksensoren ausgestattet, die den Störungsbetrieb identifizieren können. Gemäß der vorbereitenden Versuchsplanung wurde eine Versuchskampagne durchgeführt. Die erhobenen und gelabelten Messdaten wurden Ihnen zur Verfügung gestellt (<https://zenodo.org/record/1341583#.YYJOem2ZNjU>). Das Ziel des Projekts ist die Untersuchung verschiedener, datengetriebenen Modellierungsansätze hinsichtlich deren Eignung für die Erkennung des Störungsbetriebszustandes „Slug-Flow“.

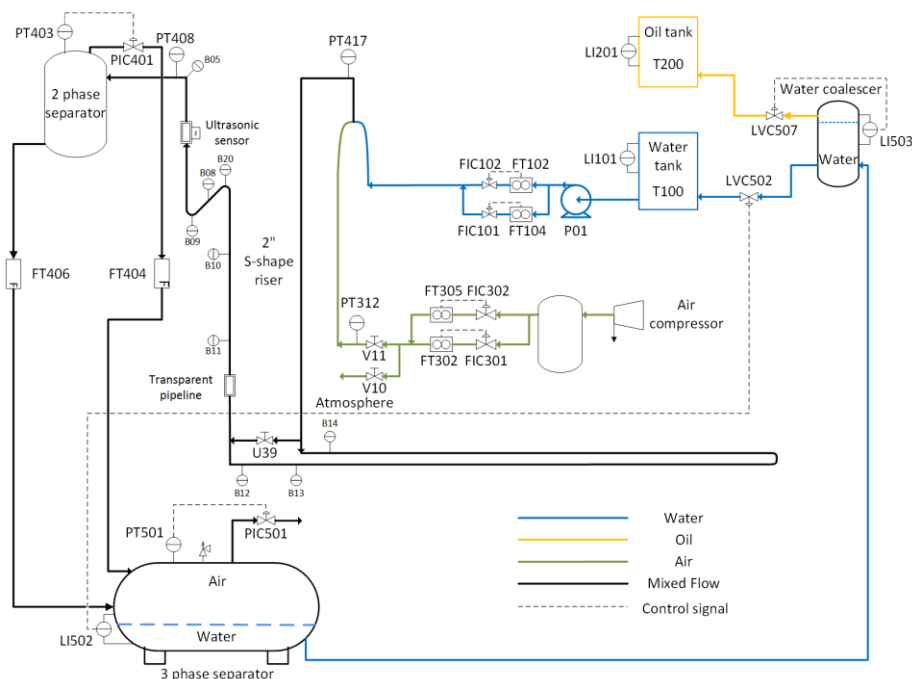


Abb. 1.: Schematische Darstellung der Anlage

Aufgabenstellung

Entwickeln Sie ein ML-Modell zur Erkennung der Störung „Slug-Flow“ in der multiphasigen Anlage. Folgende Teilaufgaben sind dazu zu bearbeiten:

1. Importieren und analysieren Sie die bereitgestellten Daten
2. Basiert auf der Problemstellung und Ergebnisse der Datenanalyse wählen Sie mehrere geeignete Modelltypen sowie die Metriken aus. Analysieren Sie Vorteile und Nachteile der gewählten Modelltypen
3. Identifizieren Sie mögliche Probleme mit Daten. Anschließend verwenden Sie geeignete Methoden um Probleme mit Daten zu beheben
4. Analysieren Sie welche aussagekräftigen Features für die Modellierung in Frage kommen

5. Trainieren Sie die im Punkt 3 identifizierten Modelle
6. Vergleichen Sie die Modelle hinsichtlich der Metriken
7. Führen Sie die Optimierung von Hyperparametern für das Model mit den besten Ergebnissen
8. Erstellen Sie eine kleine Anwendung, die die Modellinferenz über die RestAPI Schnittstelle dem Enduser anbietet (optional)

Ergebnisdokumentation

Fassen Sie alle Ergebnisse in einem Bericht zusammen und erstellen Sie einen Foliensatz, mit dem Sie Ihre Ergebnisse präsentieren können. Dokumentieren Sie ebenfalls alle getroffenen Entscheidung mit Angabe der Begründung. Der Code ist in eine Repository auf GitHub zu platzieren. MLOps-Software (z.B. DVC und CML der Firma iterative.ai) ist für die Dokumentierung und strukturierte Organisation des Modellierungsprozesses zu verwenden.

Bewertung

Die Bewertung des Projekts erfolgt auf Basis des Gesamtbilds, dass sich aus dem endgültigen Bericht, der Abschlusspräsentation und der Implementierung ergibt.