Project Specification Document

Title: Forecasting and Drift Detection in Pasteurization Sensor Data
Project Type: Machine Learning and Software Deployment
Version: 1.0

This project focuses on applying online learning and drift detection techniques to sensor data from a synthetic pasteurization process. The target is to analyze time-series signals from multiple sensors, implement forecasting models to predict future values, and integrate drift detectors to identify changes in process behavior.

Learning Objectives

-Understand the characteristics of streaming sensor data in industrial processes.
-Implement forecasting models suitable for continuous signals (e.g., temperature, pH, pressure).
-Apply and evaluate concept drift detectors (e.g., ADWIN).
-Interpret the behavior of the system before and after a detected drift.

Dataset Description

The stream represents synthetic signals from a pasteurization process, including six primary sensors and one derived feature:

- T — Temperature (°C)
- pH — Acidity
- Kappa — Electrical conductivity (mS/cm)
- Mu — Viscosity (cP)
- Tau — Turbidity (relative/NTU)
- Q_in, Q_out — Inlet and outlet flow (L/s)
- P — Pressure (bar)
- dTdt — Rate of temperature change (°C/s)

Each observation corresponds to one second of process operation, covering different states such as Idle, Fill, HeatUp, Hold, Cool, and Discharge, as represented in Figure 1.
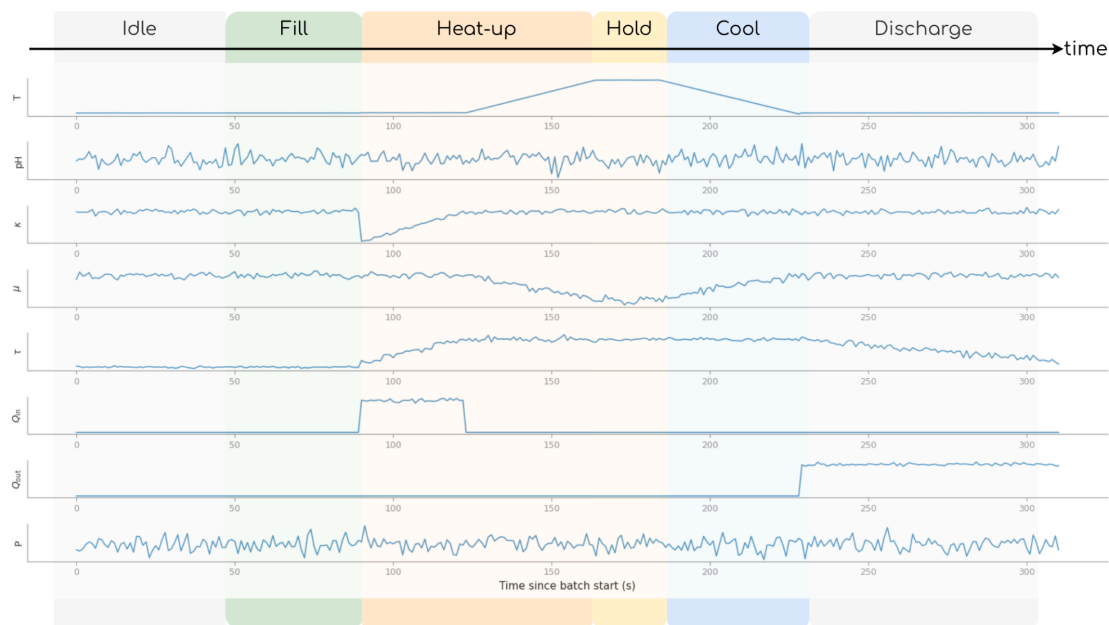
Figure 1. Sensor signals and pasteurization steps.

Tasks

1. Using a Flask server, provide access to sensors generated by the simulation function.
2. Choose one or more key sensors (e.g., Temperature, pH, or Conductivity) and implement a forecasting model to predict their next value in the stream. Recommended options:
   - Linear regression or Ridge regression (Online Algorithms)
3. Implement a concept drift detector using ADWIN (Adaptive Windowing)

4. Integrate the drift detector into your streaming pipeline. When a drift is detected:
   - Log the timestamp and affected variable
   - Optionally, reset or retrain the forecasting model

5. Visualize the results in a Streamlit app:
   - Actual vs. predicted values
   - Detected drift points