Process Modeling Scientist Challenge

This challenge involves developing and presenting a predictive bioprocess model using JAX.

Task: Create a predictive model of a simplified biocultivation

This assignment assesses your skills in building a probabilistic predictive mathematical model of a simplified biocultivation. We will focus on predicting metabolite concentrations as observed in the given data set and the propagation of uncertainty. Please submit your code via a Git repository and remember to add the requirement.txt.

Dataset

- You are provided with an Excel file containing bioprocess data from a simplified bioreactor.
- The file contains
    - Sheet1: time-series concentration measurements of two metabolites ("S" and "X"),
    - Sheet2: feeding over time as bolus shot volumes,
    - Sheet3: feed media concentrations of metabolite S.
    - Sheet4: start volumes
- Feeding at time t_x occurs immediately after measurements at time t_x .
- In the simplified bioreactor $1S \rightarrow 1X$ is the only chemical reaction taking place. The reaction kinetics are unknown.
- Measurement noise η can be assumed additive, independent and normally distributed, $\eta \sim N(\mu = 0, \sigma = 0.1)$ for concentration measurements of both S and X.

1) Data Preprocessing

a) Data Analysis:
   i) Load and inspect the provided data.
   ii) Analyze and visualize the dataset to understand its characteristics

b) Data Splitting:
   i) Split the data into training and validation and/or test sets. Justify your split.

2) **Model Training, Validation and Analysis**

a) Process Model:
   i) Develop a mathematical model of the simplified bioprocess that describes the changes of concentrations in time using the JAX ML framework. You may choose any appropriate modeling and inference technique.
   ii) Ensure that the model accounts for the uncertainty in the data.
   iii) Explain and justify your probabilistic model and choice of training algorithm - what alternatives did you consider?
   iv) Visualize and present the inferred kinetic function, the predicted concentrations and the accuracy of the model, including any uncertainty of the predictions.

b) Tuning and Validation:

   i) Tune the performance (accuracy on unseen data) of your JAX model using a suitable method.

   ii) Explain and visualize the performance metrics that you used.

   iii) Compare the performance of the tuned model with the untuned model.

c) Model Analysis
   i) Explain the behavior and characteristics of your model.
   ii) Explain and justify the method(s) that you chose for your analysis – what alternatives did you consider?

3) Documentation and Presentation

a) Concisely document your approach, including the rationale behind your modeling choices and how you handled uncertainty.

b) Prepare a brief presentation as if you were presenting to a customer that intends to use your model to gain process understanding and optimize process development.