Project
ID2223 / HT2022

Task: Build a scalable Prediction Service using a Real World Data Source

1. Choose a real-world data source that is updated regularly (once min/hour/day/week).
2. Define the prediction problem that you will solve based on data from the data source.
3. Define a data model with the features and label(s) or your data.
4. Build a feature pipeline to create the features/label(s) from the raw data
5. Build a training pipeline to train a model using your features/label(s)
6. Build an inference pipeline to make predictions using new data and build a User Interface to communicate the results to a user

You are free to use any of the technologies we covered in the labs: Hugging Face, Hopsworks, Modal, Colab, Github. You can also use other services - please vet them with us first, though.

Deliverables

- By Dec 9th, propose the project to us. If you cannot come up with a project, you can do a variant of the air quality project (next slide)

- Deliver your source code and documentation as a **Git Repository URL**

- Deliver your project architecture and description as a README.md file in the root of your Git repository.
- Include in the README.md file:
  - a description of your prediction service its the data source
  - a public URL for the UI for your project (e.g., a Hugging Face Spaces URL).

Deadline midnight 15th January.

Air Quality Project

- If you cannot come up with your own project, you can work on this project to predict air quality.

- You need to

  (1) identify a location for air quality predictions (not Stockholm or Kyiv)

  (2) collect historical weather and air quality data for that location

  (3) train a model with acceptable performance (don't use logistic regression - use XGBoost as a baseline) that uses both historical weather data and air quality measurements to predict future air quality (using future weather predictions)

  (4) build a Hugging Face space where you can see the predictions of air quality for that location for the next 7 days

Sample code:
https://github.com/logicalclocks/hopsworks-tutorials/tree/master/advanced_tutorials/air_quality