

Automate your data science workflows using Elyra, Kubeflow pipelines

Operate First Data Science Meetup
Jan 25, 2022

Oindrilla Chatterjee
Senior Data Scientist
Open Services

What are AI/ML Pipelines?

A reproducible workflow that can be used to automate repetitive steps in a machine learning workflow.

How do AI/ML Pipelines make a data scientist's life easier?

How do AI/ML Pipelines make a data scientist's life easier?

- Automating Manual Tasks

How do AI/ML Pipelines make a data scientist's life easier?

- Automating Manual Tasks
- Better Reproducibility

How do AI/ML Pipelines make a data scientist's life easier?

- Automating Manual Tasks
- Better Reproducibility
- Monitoring and Tracking

Tools to create automated pipelines

- **Elyra** – Notebook pipelines visual editor
- **Kubeflow Pipelines** – Build and deploy scalable ML workflows

Pre-requisites

1. Jupyterlab instance with Elyra
2. Kubeflow Pipelines Deployed
3. Jupyter notebooks/scripts to automate
4. Container images containing notebook dependencies

Demo



Resources & Q/A

1. Detailed Guide/Tutorial:
<https://github.com/AICoE/elyra-aidevsecops-tutorial>
2. Track your metrics using Kubeflow Pipelines:
https://www.operate-first.cloud/data-science/ds-workflows/docs/develop_collaborate/track-metrics-using-kubeflow.md
3. OS Climate Repo:
<https://github.com/os-climate/aicoe-osc-demo>