CLAIMED – Component Library for AI, Machine Learning, ETL and Data Science

A Kubeflow low code architecture

A scalable open source (Trusted AI)
Computer Vision pipeline for health care

Romeo Kienzler
Data Scientist
IBM

Ivan Nesic
ML Engineer
University Hospital Basel
Requirements

- Rapid prototyping using visual editing and notebooks
- Seamless scaling during development and deployment
- GPU support
- ML tools: PyData stack, TensorFlow, PyTorch, …
- Life science tools: DICOM input, DICOM output, …
- Reproducibility
- Data lineage
- Reference implementation in open source
- Collaboration support
Cloud annotations provides...

Browser based image labeling: Classification / Object recognition training data

IBM Cloud Annotations

...in Open Source
Kubeflow provides...

AutoML, Deployment, Reproducibility, Notebooks, Pipelines, Serving, Training, Scale

Some prominent users:
Amazon Web Services
IBM Watson Services
IBM’s top clients

Kubeflow...on top of Kubernetes
Elyra provides...

No Code / Low Code ML Pipeline Design
Re-usable pipeline components
Interchangeability of Engines
(Kubeflow, Airfow, ...)

...on top of JupyterLab, VSCode, ...
CLAIMED...

Component Library for AI, Machine Learning, ETL and Data Science

Portability
No Code / Low Code
Pipeline Components
Jupyter Notebooks
Sample Pipelines

...on top of Elyra and Kubeflow
Example Pipeline Components

**Category**: Training  **Group**: Distributed  **Name**: TFJob

The TFJob operator supports parallel training on multiple nodes and GPUs.
Example Pipeline Components

**Category:** Processing  **Group:** Distributed  **Name:** SparkJob

The SparkJob operator supports parallel processing on multiple nodes.
Example Pipeline Components

**Category:** Tuning

**Group:** Hyperopt

**Name:** Katib

Visualization of a hyper parameter optimization result
Example Pipeline Components

**Category:** Metric  **Group:** Explainability  **Name:** AIX360/LIME

Example on how LIME helps to identify classification relevant areas of an image.
Example Pipeline Components

Category: Metric Group: Adversarial Robustness Name: ART

Example on how Adversarial Attacks happen
Example Pipeline Components

Category: Metric Group: AI Fairness Name: AIF360

Dataset: German credit scoring
Mitigation: Adversarial Debiasing algorithm applied

Protected Attribute: Sex
Privileged Group: Male, Unprivileged Group: Female
Accuracy after mitigation changed from 75% to 76%
Bias against unprivileged group unchanged after mitigation (0 of 5 metrics indicate bias)

Protected Attribute: Age
Privileged Group: Old, Unprivileged Group: Young
Accuracy after mitigation changed from 75% to 66%
Bias against unprivileged group was reduced to acceptable levels for 1 of 4 previously biased metrics (3 of 6 metrics still indicate bias for unprivileged group)

Example on how the AIF360 toolkit computes fairness metrics and mitigates bias
IBM Watson OpenScale

IBM Watson OpenScale uses the same Open Source components on top of Kubernetes, Kubeflow and KFServing

Did you know?

Model Monitoring and visualization of bias (IBM Watson OpenScale)
Component creation and cataloging...

...using Elyra and Data Science Asset Repository
Component consumption...

...using Elyra or Watson Orchestrator
The final pipeline in Elyra...

...and KubeFlow
Summary

- Rapid prototyping using visual editing and notebooks
- Seamless scaling during development and deployment
- GPU support
- ML tools: PyData stack, TensorFlow, PyTorch, …
- Life science tools: DICOM input, DICOM output, …
- Reproducibility
- Data lineage
- Reference implementation in open source
- Collaboration support