Open Source MLOps with Kubeflow and ElyraAI

Open Source @ IBM

CODAIT

Center for Open Source Data and AI Technologies

Romeo Kienzler
CTO and Chief Data Scientist, STSM
IBM Center for Open Source Data and AI Technologies (CODAIT)

Credits, thanks and kudos to
Animesh Singh, STSM and Chief Architect, Data and AI OpenSource Platform
Luciano Resende, Open Source AI Platform Architect

What is Docker?

Product using OS-level virtualization to deliver software in packages called containers

Provides...

Lightweight virtualization

Security and isolation

Super-fast startup/teardown

...on top of Linux

What is Kubernetes?

Provides...

Container Orchestration

Deployment, scaling and management

High availability

...on top of Linux Cluster Hosts

Used by (among others):
Adidas, Booking.com, Box, Google, Huawei, IBM, The New York Times, ING, ricardo.ch, Spotify, Wikimedia, Zalando

What is Kubeflow?

Provides...

AutoML

Deployment

Reproducibility

Notebooks

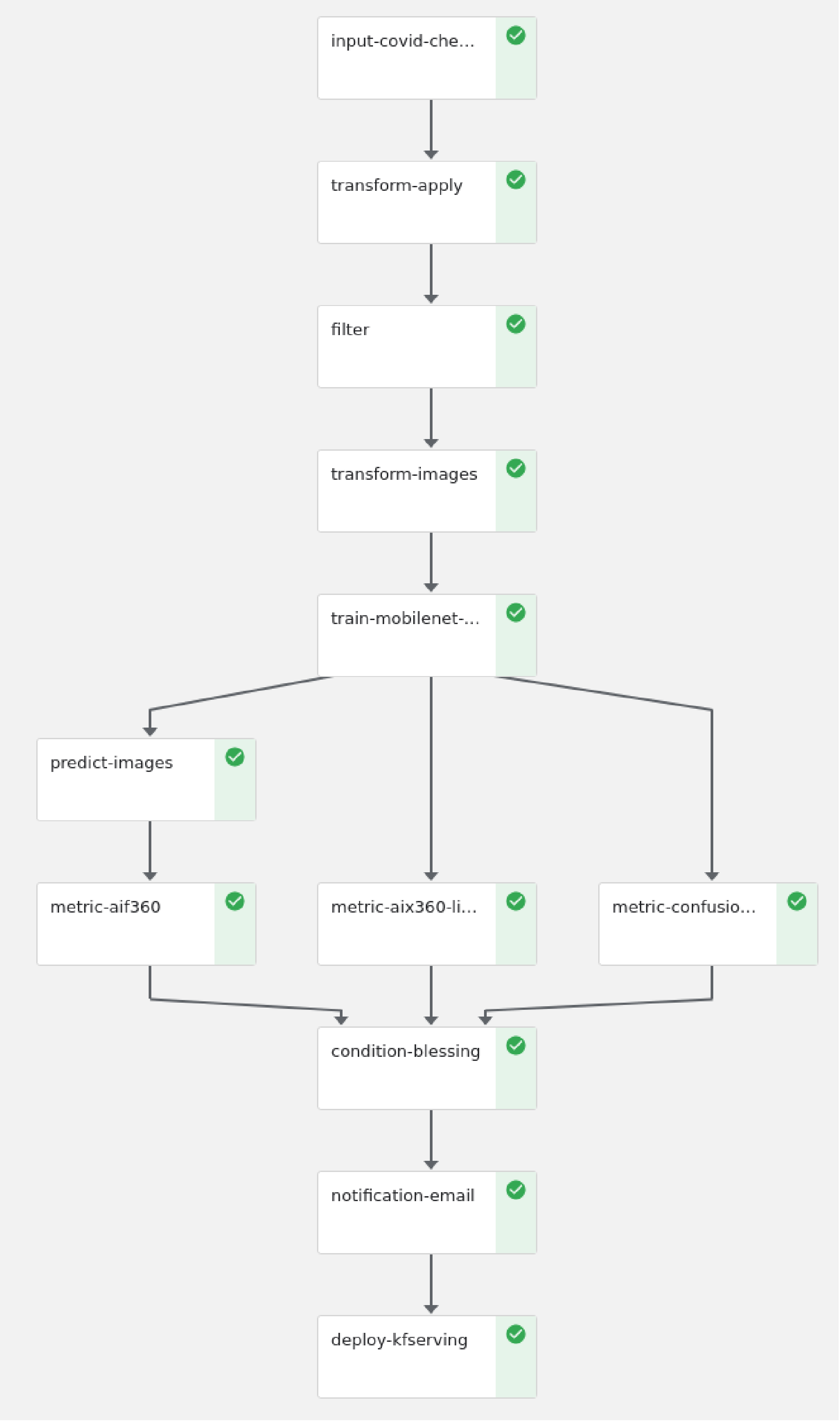
Pipelines

Serving

Training

Scale

...on top of Kubernetes



What is ElyraAI?

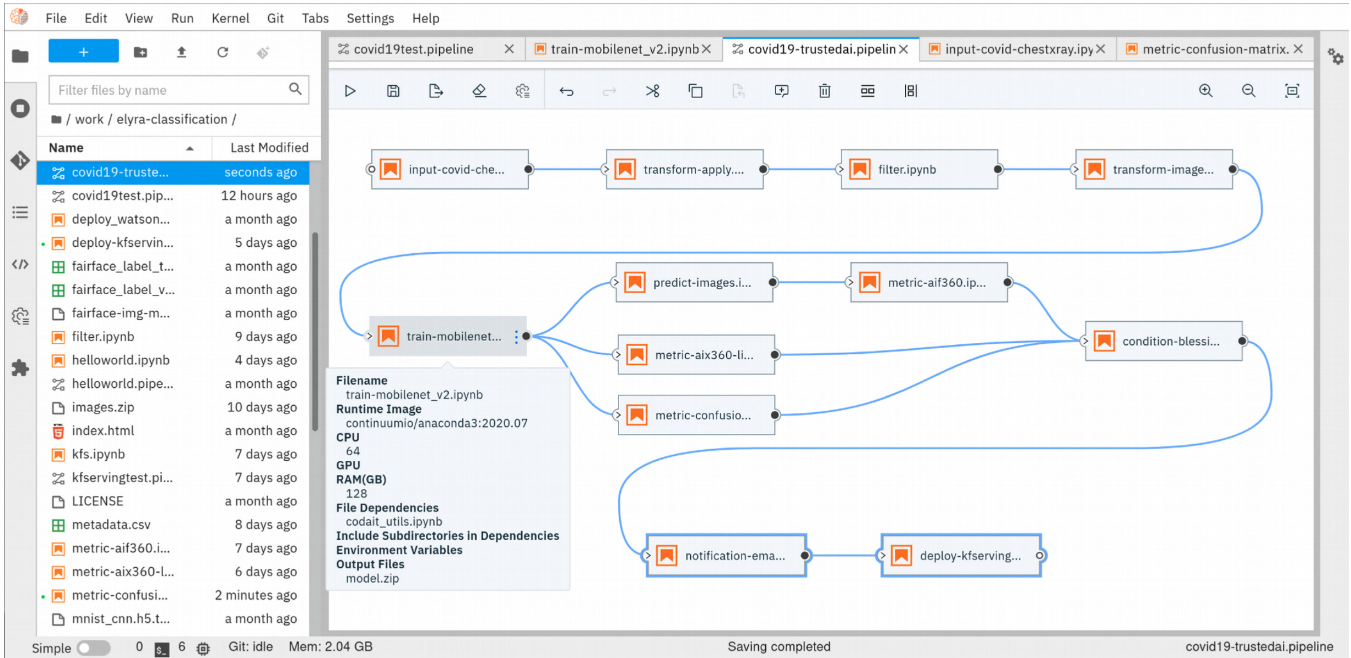
Provides...

No Code / Low Code ML Pipeline Design

Re-usable pipeline components

Interchangeability of Engines (Kubeflow, Airfow, ...)

...on top of JupyerLab, VSCode, ...



Name	Last Modified
covid19-truste...	seconds ago
covid19test.pip...	12 hours ago
deploy_watson...	a month ago
deploy-kfservin...	5 days ago
fairface_label_t...	a month ago
fairface_label_v...	a month ago
fairface-img-m...	a month ago
filter.ipynb	9 days ago
helloworld.ipynb	4 days ago
helloworld.pipe...	a month ago
images.zip	10 days ago
index.html	a month ago
kfs.ipynb	7 days ago
kfservingtest.pi...	7 days ago
LICENSE	a month ago
metadata.csv	8 days ago
metric-aif360.i...	7 days ago
metric-aix360-l...	6 days ago
metric-confusi...	2 minutes ago
mnist_cnn.h5.t...	a month ago

Filename
train-mobilenet_v2.ipynb
Runtime Image
continuumio/anaconda3:2020.07
CPU
64
GPU
RAM(GB)
128
File Dependencies
codait_utils.ipynb
Include Subdirectories in Dependencies
Environment Variables
Output Files
model.zip

What is CLAIMED?

***C**omponent **L**ibrary for **AI**, **M**achine Learning, **E**TL and **D**ata Science*

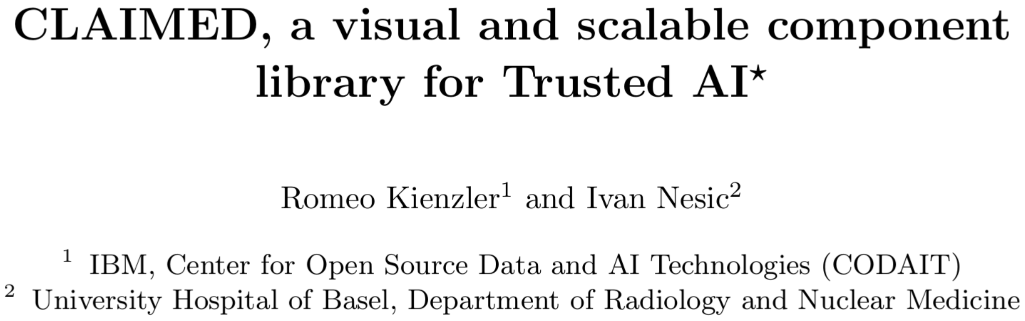
Provides...

Portable No Code / Low Code Pipeline Components

Jupyter Notebooks

Sample Pipelines

...on top of ElyraAI and Kubeflow



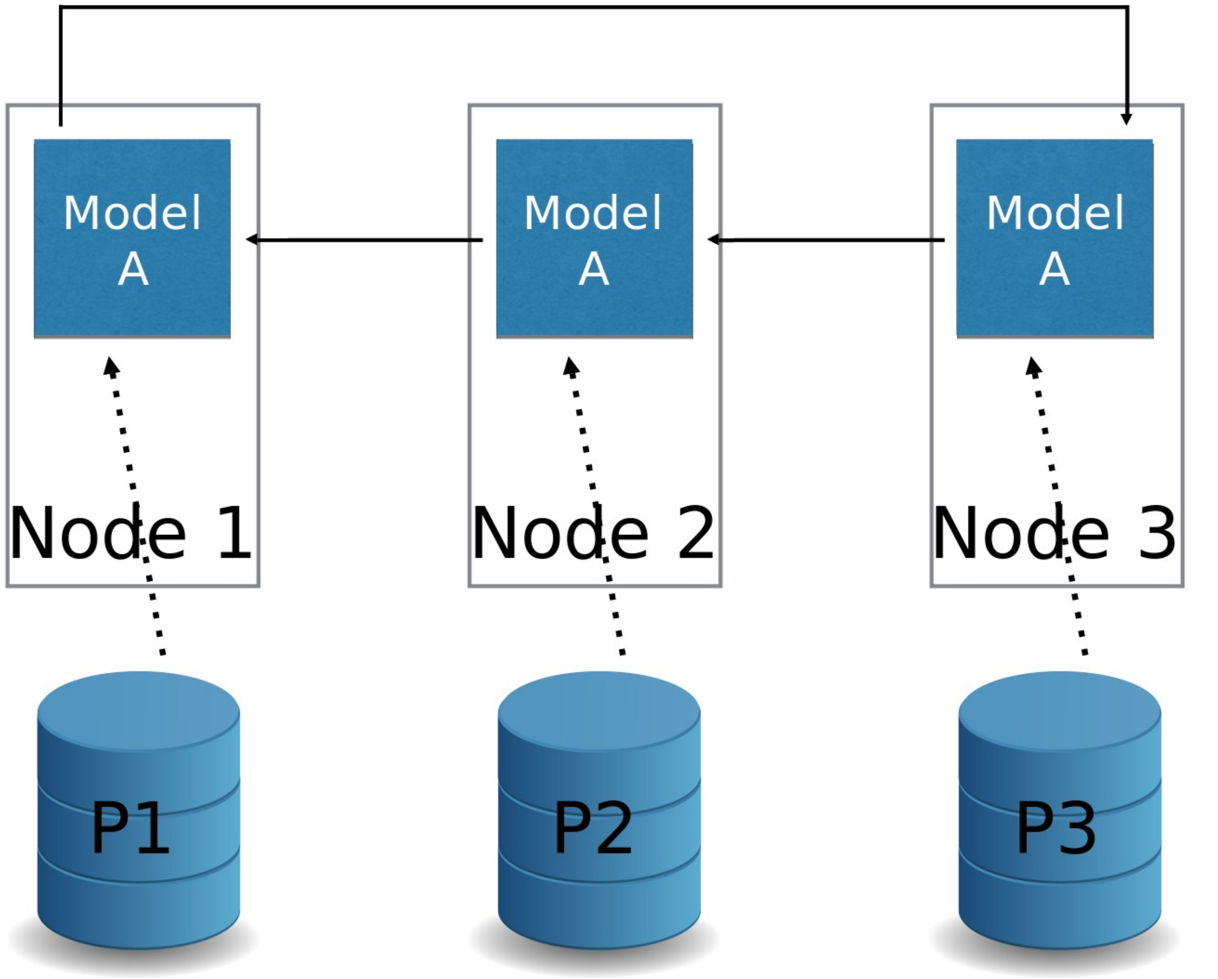
CLAIMED, a visual and scalable component library for Trusted AI[⋆]

Romeo Kienzler[1] and Ivan Nesic[2]

[1] IBM, Center for Open Source Data and AI Technologies (CODAIT)
[2] University Hospital of Basel, Department of Radiology and Nuclear Medicine

Example Pipeline Components

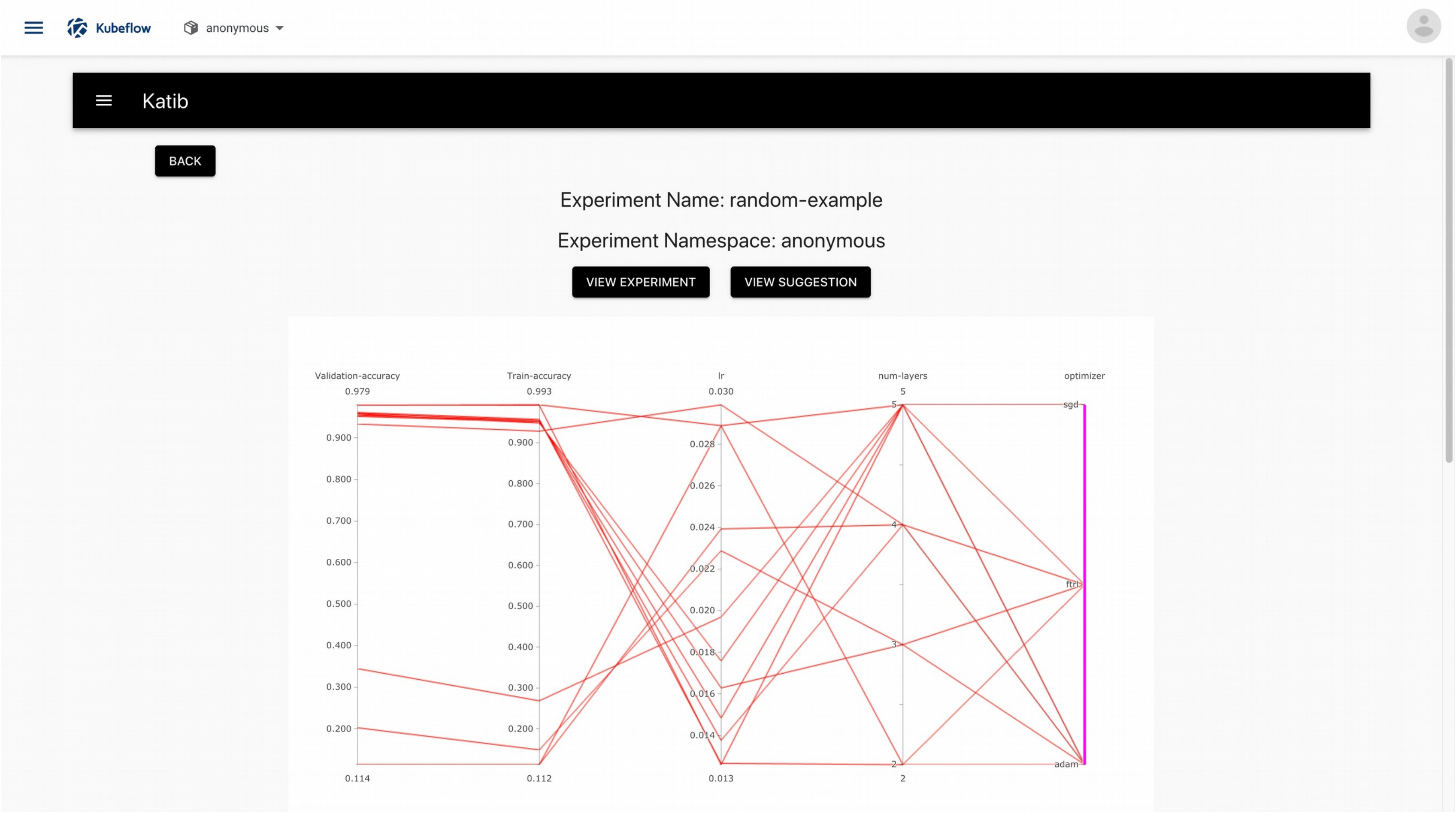
Category: Training **Group**: Distributed **Name**: TFJob



The TFJob operator supports parallel training on multiple nodes and GPUs

Example Pipeline Components

Category: Tunig **Group**: Hyperopt **Name**: Katib



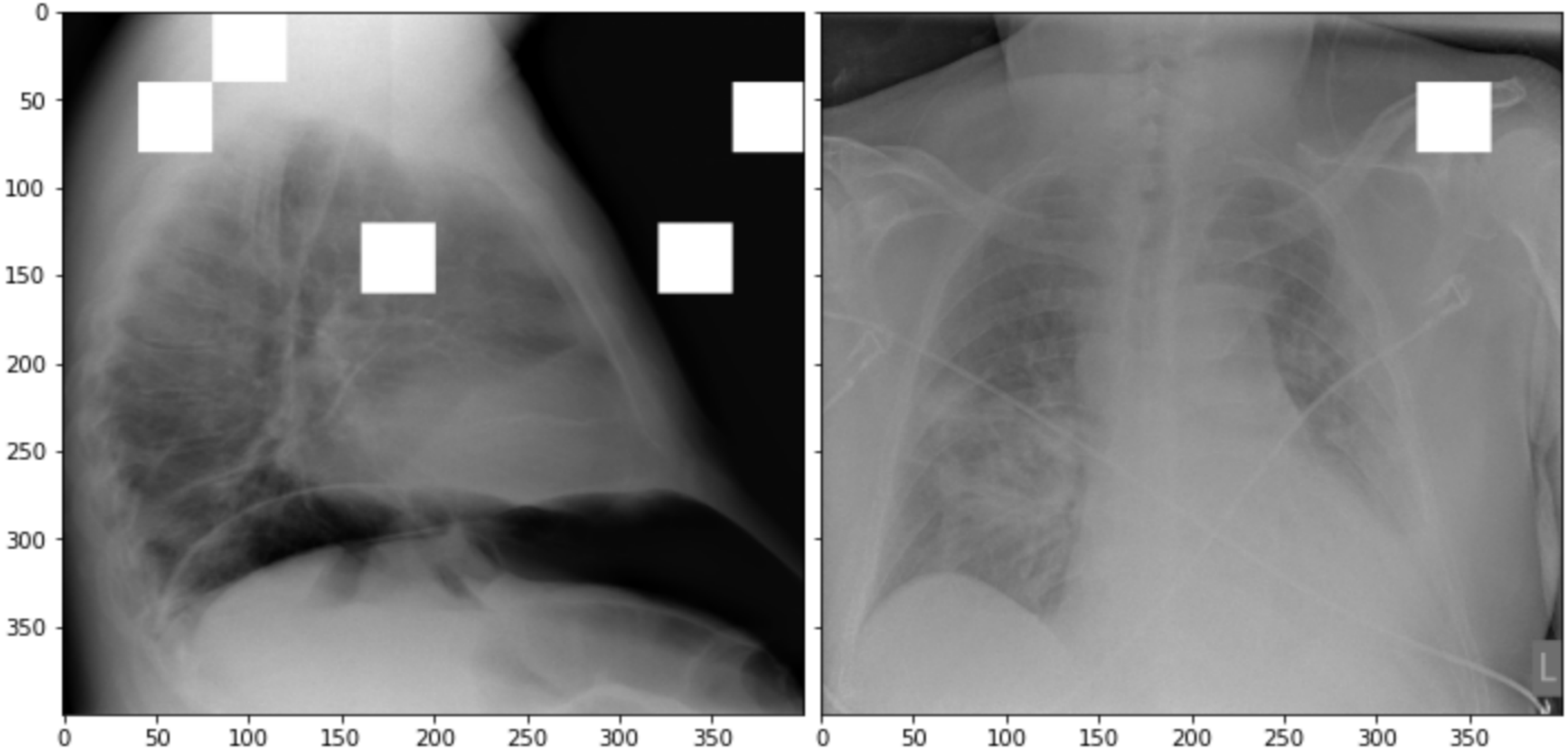
Experiment Name: random-example

Experiment Namespace: anonymous

Visualization of a hyper parameter optimization result

Example Pipeline Components

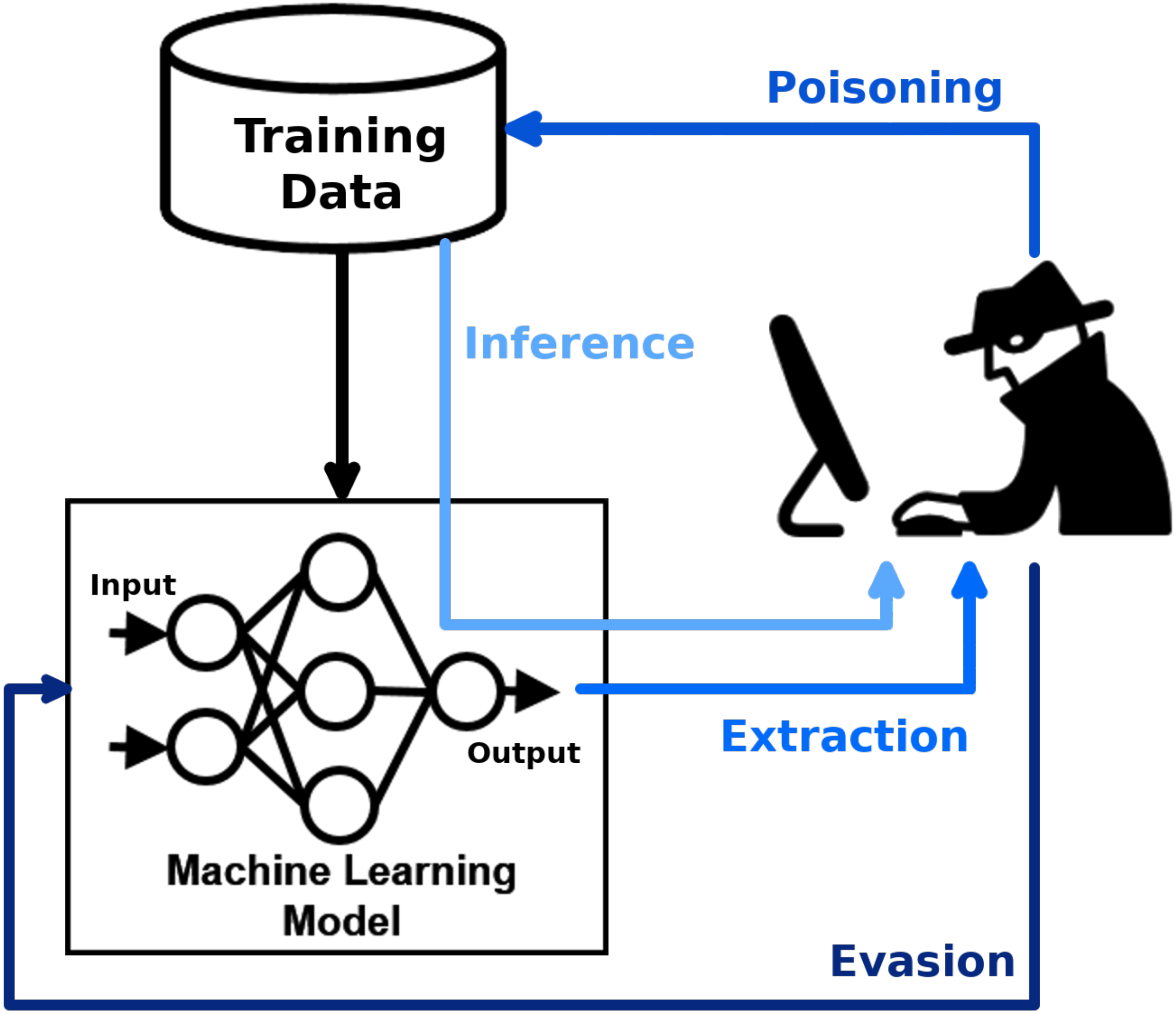
Category: Metric **Group**: Explainability **Name**: AIX360/LIME



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

Example Pipeline Components

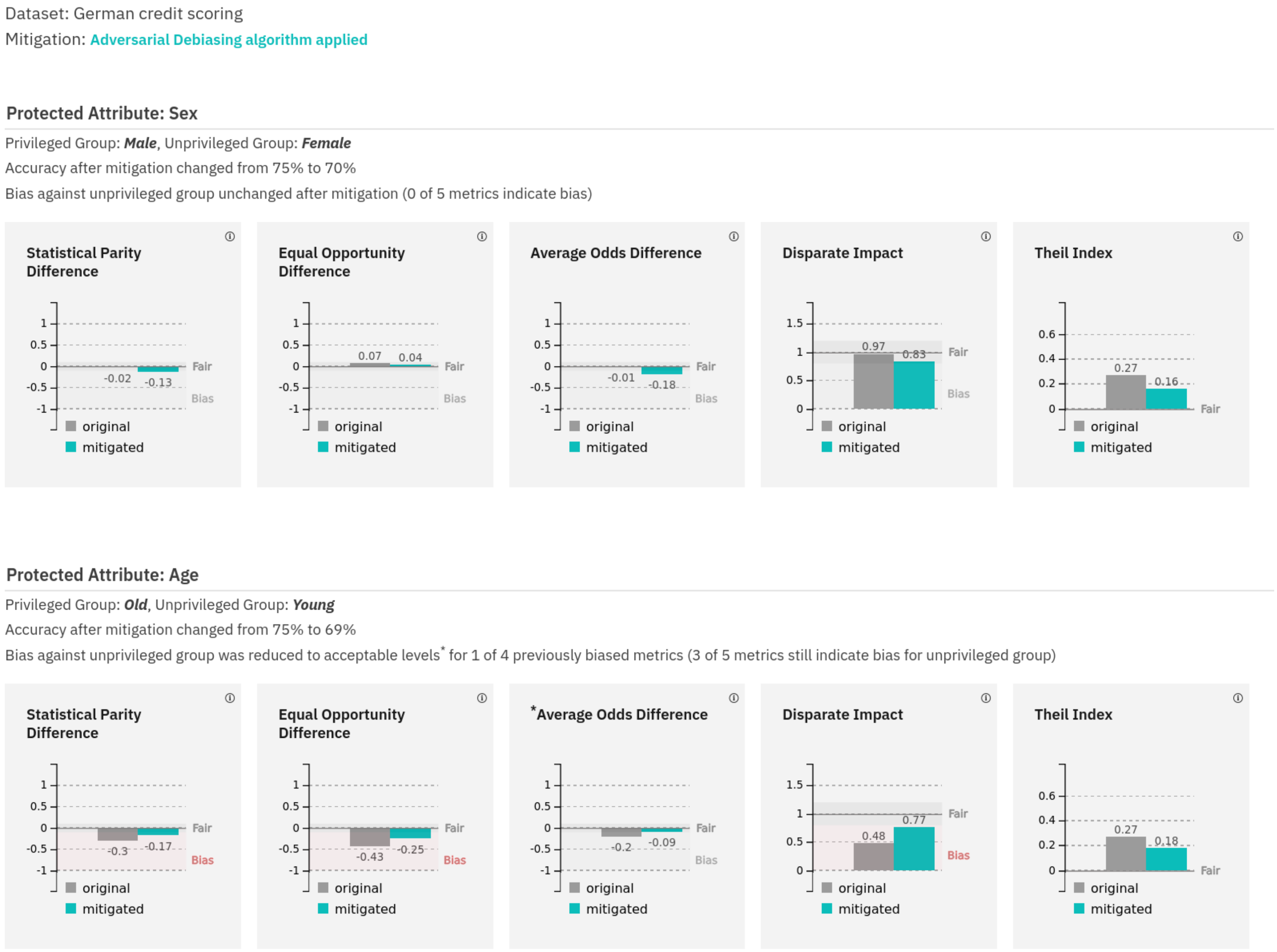
Category: Metric **Group**: Adversarial Robustness **Name**: ART



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 70%

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 69%

Bias against unprivileged group was reduced to acceptable levels* for 1 of 4 previously biased metrics (3 of 5 metrics still indicate bias for unprivileged group)

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

Links

- https://github.com/Trusted-AI/adversarial-robustness-toolbox
- https://github.com/Trusted-AI/AIF360
- https://github.com/Trusted-AI/AIX360
- https://github.com/kubeflow/kubeflow
- https://www.slideshare.net/AnimeshSingh/kfserving-serverless-model-inferencing-236725227
- https://www.tensorflow.org/api_docs/python/tf/keras/applications
- https://www.docker.com/
- https://github.com/kubernetes/kubernetes
- https://elyra.readthedocs.io/en/latest/