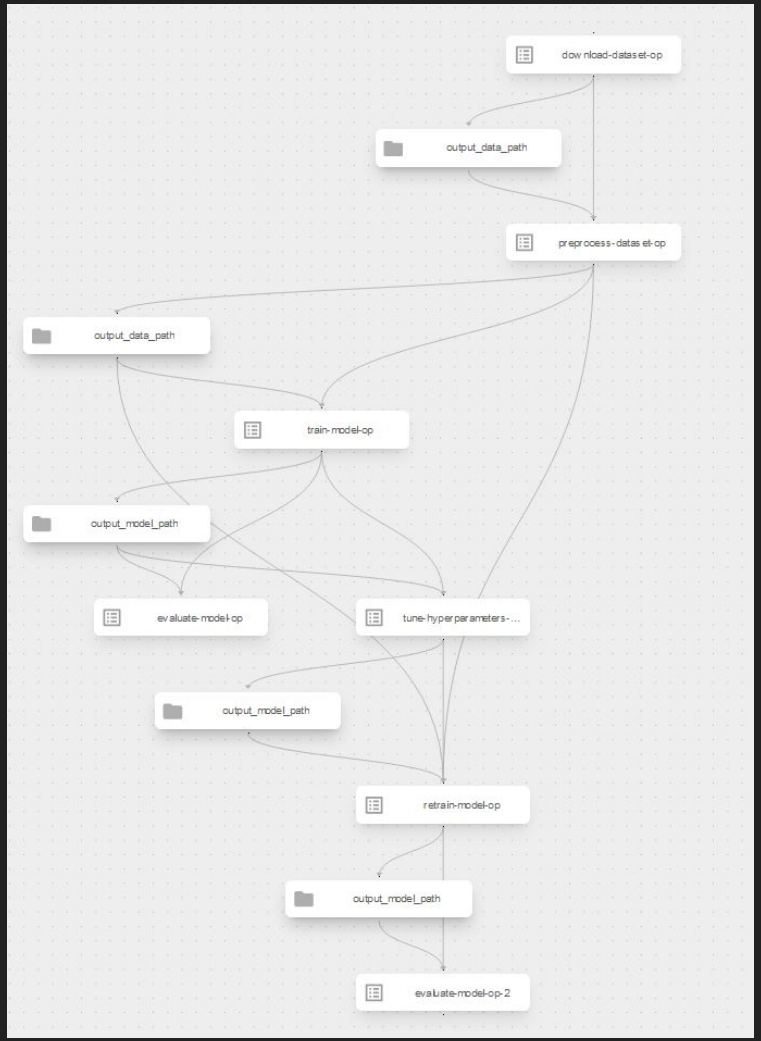
MLOps Pipeline using Kubeflow, Katib, and SHAP

Presented by: Mike Segal

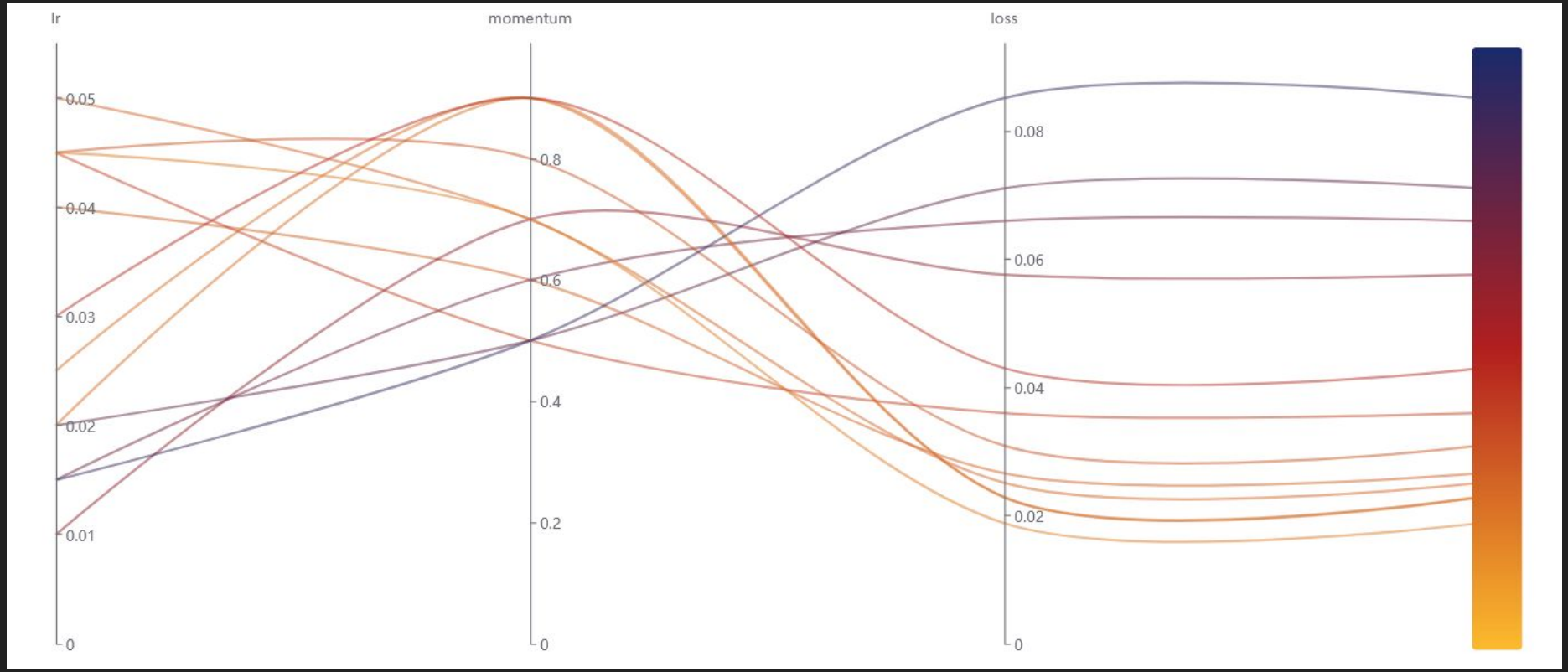
MLOps with Kubeflow

1) What is MLOps?
   a) data management
   b) model development
   c) version control
   d) automation
   e) testing
   f) deployment
   g) monitoring

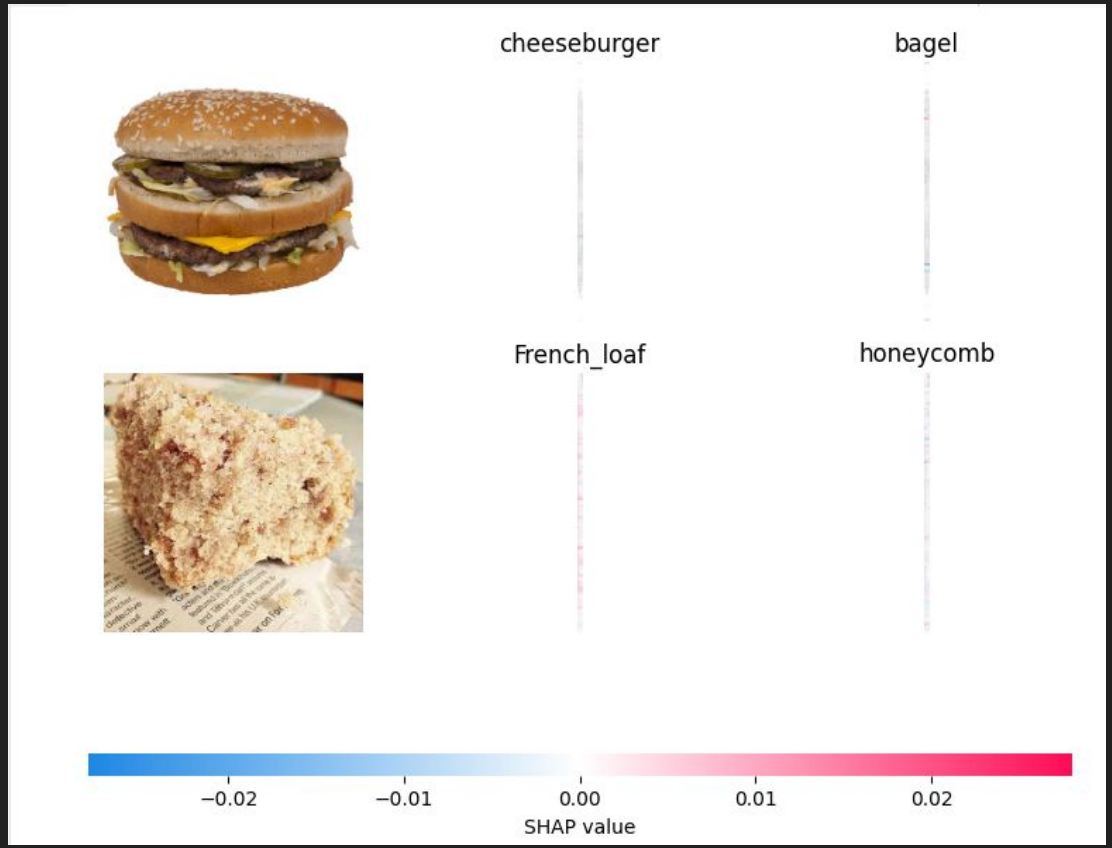
2) Why use it?
   a) Reproducibility
   b) Scalability



AutoML Integration with Katib - Grid



Model Interpretability with SHAP



Model Interpretability with SHAP

```
data = [
    ("Who was the founder of Constantinople?", "Constantinople was founded by Constantine."),
    ("What is the Haber process?", "The Haber process synthesizes ammonia."),
    ("What was Leonardo Da Vinci famous for?", "Leonardo Da Vinci was famous for his inventions.")
]

# Convert questions and contexts into input format
inputs = [q + " [SEP] " + c for q, c in data]

# Build an explainer using a token masker
explainer = shap.Explainer(predict, tokenizer)

# Generate SHAP values for each question-context pair
shap_values = explainer(inputs)

# Output SHAP values
for i, sv in enumerate(shap_values):
    print(f"SHAP Values for input {i+1}:")
    shap.plots.text(sv)
```

```
PartitionExplainer explainer:  33%|███      | 1/3 [00:00<?, ?it/s]

PartitionExplainer explainer: 100%|██████████| 3/3 [01:31<00:00, 25.06s/it]

PartitionExplainer explainer: 4it [02:21, 47.05s/it]SHAP Values for input 1:
```

inputs

who was the founder of constantinople ? [SEP] constantinople was founded by constantine .

SHAP Values for input 2:

inputs

what is the haber process ? [SEP] the haber process synthesizes ammonia .

SHAP Values for input 3:

inputs

what was leonardo da vinci famous for ? [SEP] leonardo da vinci was famous for his inventions .

Takeaways

1. Avoid third party platforms if possible - use cloud infrastructure directly
    a. Not tied to specific regions, can move infrastructure quickly
2. Otherwise, MLOps, AutoML, and SHAP are great!