DagsHub

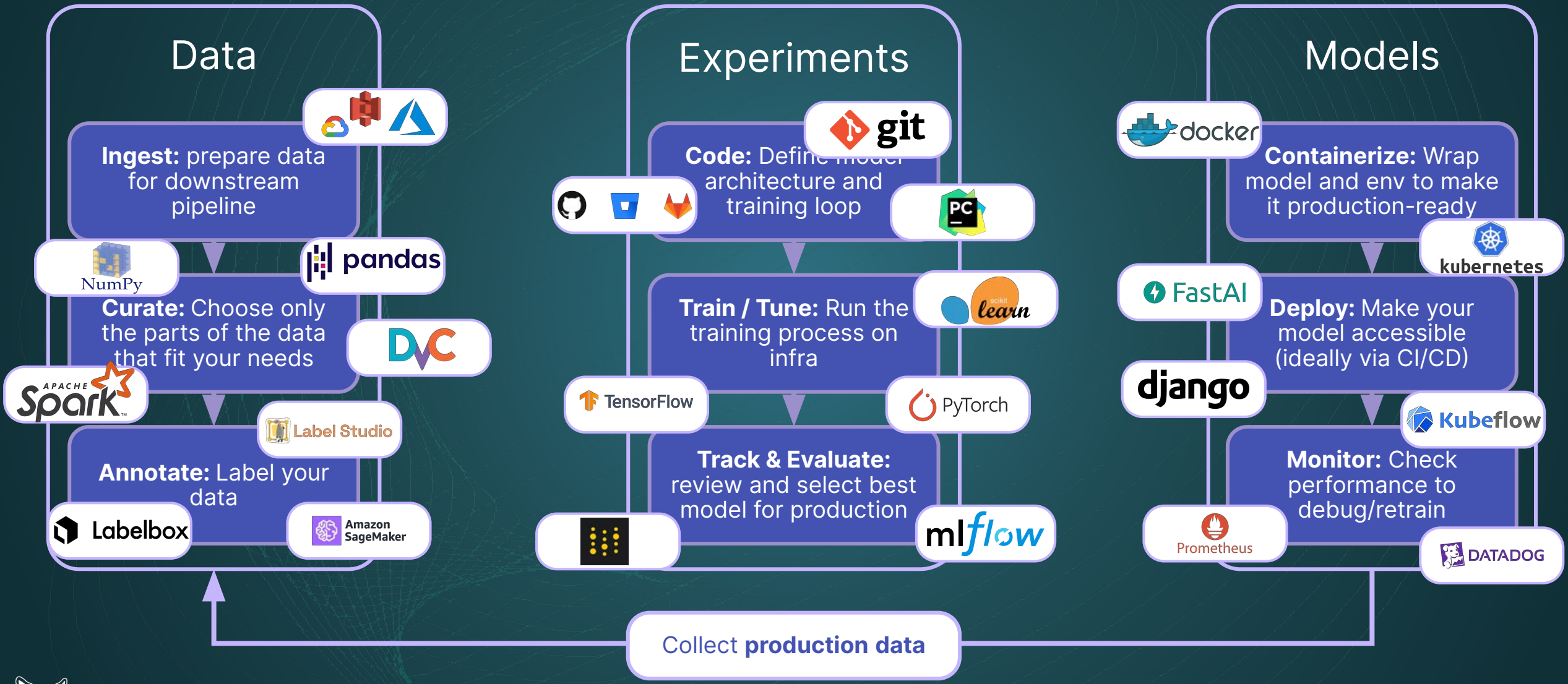
A Hitchhiker's Guide to the MLOps Experience

Noa Goldman
Lead Product Manager at Dagshub

ML engineers' lives can be challenging

- **Data quality and availability**
- **Model selection and tuning**
- **Model interpretability**
- **Work in a team**
- **Keeping up with new technologies**

The current solution



DagsHub

Who am I

- Lead Product Manager at Dagshub
- Ex software engineer
- Making simple dev-tools a passione
- Love crossfit & my puppy hippo

DagsHub

What is mlops

The **set of practices and technologies** that enable the **development, deployment**, and **management** of machine learning models throughout their lifecycle.

To ensure that they are **reliable, scalable**, and can be **updated and maintained** over time.

Mlops is important to ML engineers

Productivity

Scale

Reliability

Future-proof

The main components MLOps should cover

Data

Experiments

Models

DagsHub

– CONFIDENTIAL –

The main components MLOps should cover

Data

– CONFIDENTIAL –

Data challenges

- Keeping track, managing and organizing **constantly collected data** from various sources.
- Quickly doubling down with the **data points that matters**.

Data Versioning

- Treat your data like you treat your code:
  - Clear display of different versions
  - Reproducibility
- Easy collaboration:
  - Comments
  - Pull requests
  - Issues

Dean / SavtaDepth

forked from OperationSavta/SavtaDepth

Watch 1 | Star 0 | Fork 0

Files | Experiments 1 | Issues 0 | Pull Requests 0 | Reports new | Discussions 0 | Annotations

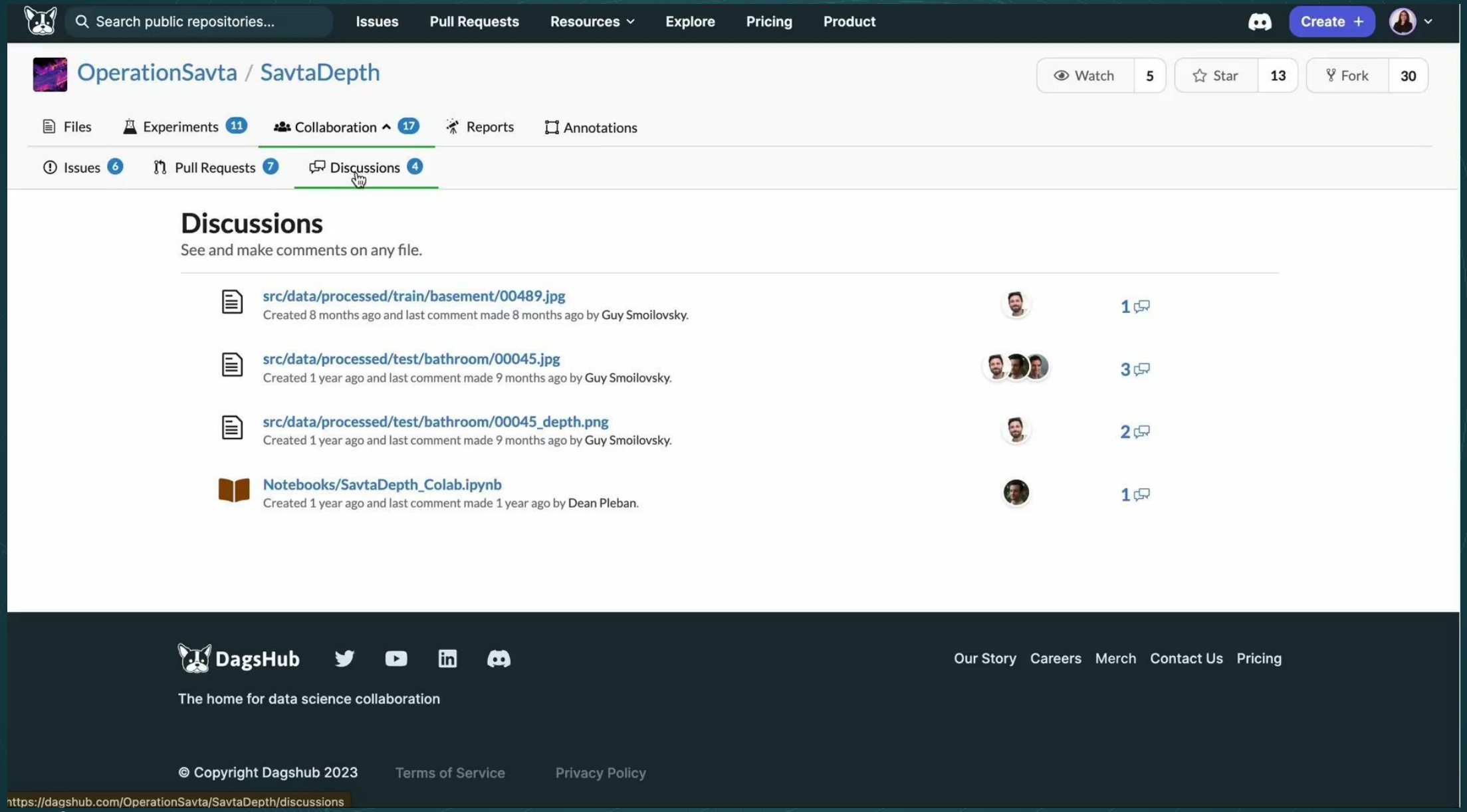
base: master ⇄ compare: epoch_testing

File Compare

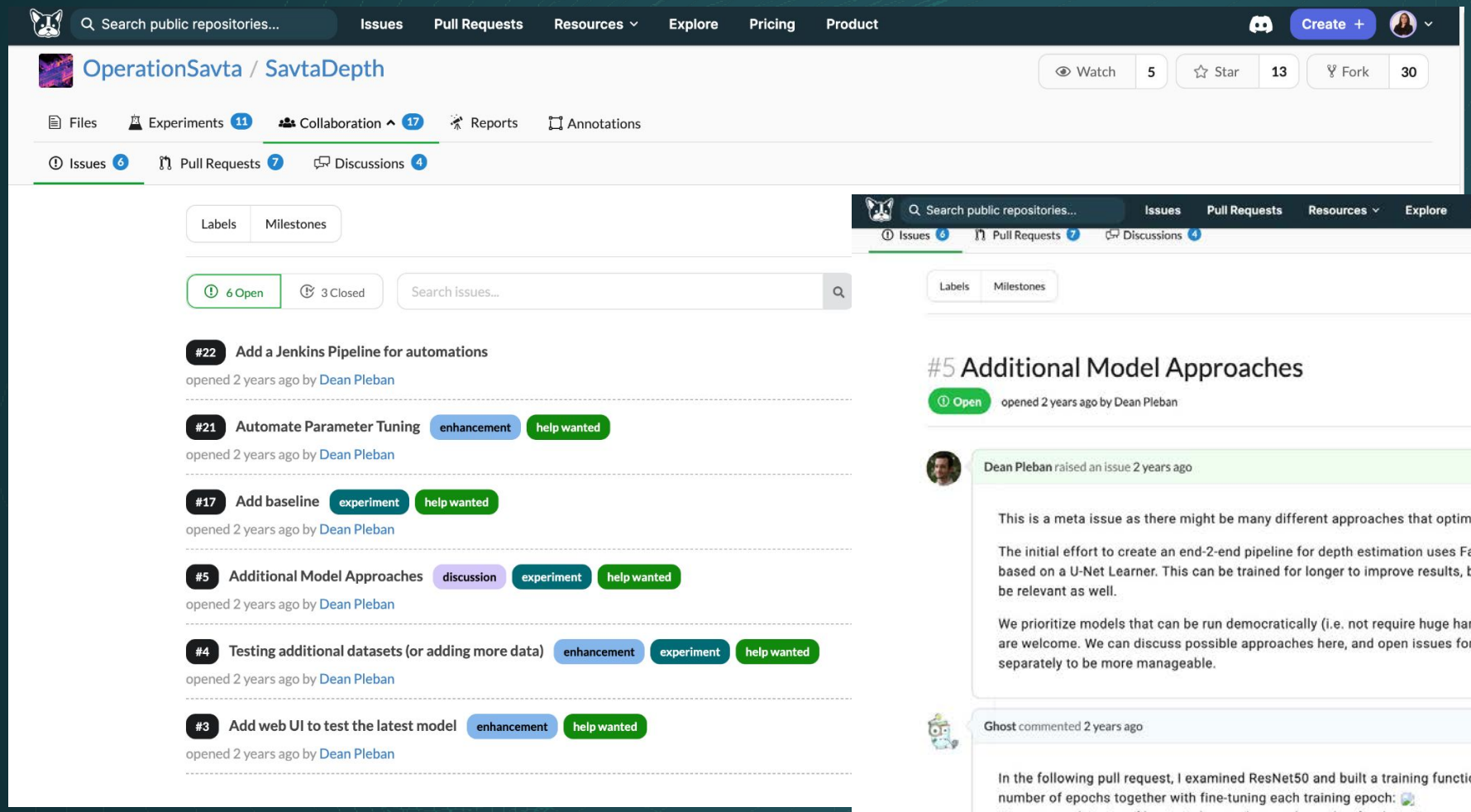
SavtaDepth / src / eval / **examples**

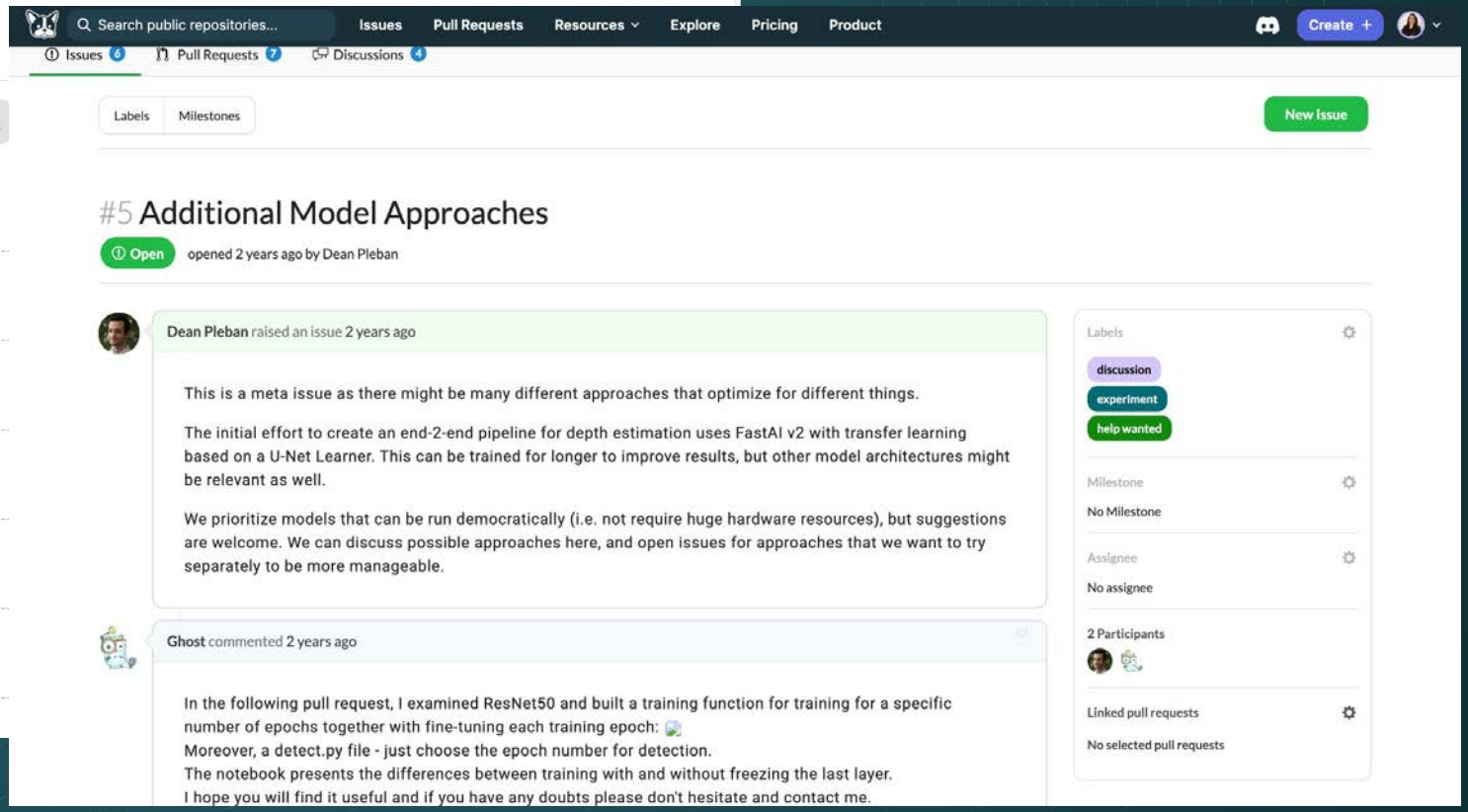
View left	View right
..	
00356_pred.png	00356_pred.png
00357_pred.png	00357_pred.png
00358_pred.png	00358_pred.png
00360_pred.png	00360_pred.png
00361_pred.png	00361_pred.png
00383_pred.png	00383_pred.png
00385_pred.png	00385_pred.png
00394_pred.png	00394_pred.png
00396_pred.png	00396_pred.png
00554_pred.png	00554_pred.png

Comments


Search public repositories...
Issues
Pull Requests
Resources
Explore
Pricing
Product
Create +
OperationSavta / SavtaDepth
Watch 5
Star 13
Fork 30
Files
Experiments 11
Collaboration 17
Reports
Annotations
Issues 6
Pull Requests 7
Discussions 4
Discussions
See and make comments on any file.
src/data/processed/train/basement/00489.jpg
Created 8 months ago and last comment made 8 months ago by Guy Smoilovsky.
1
src/data/processed/test/bathroom/00045.jpg
Created 1 year ago and last comment made 9 months ago by Guy Smoilovsky.
3
src/data/processed/test/bathroom/00045_depth.png
Created 1 year ago and last comment made 9 months ago by Guy Smoilovsky.
2
Notebooks/SavtaDepth_Colab.ipynb
Created 1 year ago and last comment made 1 year ago by Dean Pleban.
1
DagsHub
Our Story
Careers
Merch
Contact Us
Pricing
The home for data science collaboration
© Copyright Dagshub 2023
Terms of Service
Privacy Policy
https://dagshub.com/OperationSavta/SavtaDepth/discussions

DagsHub

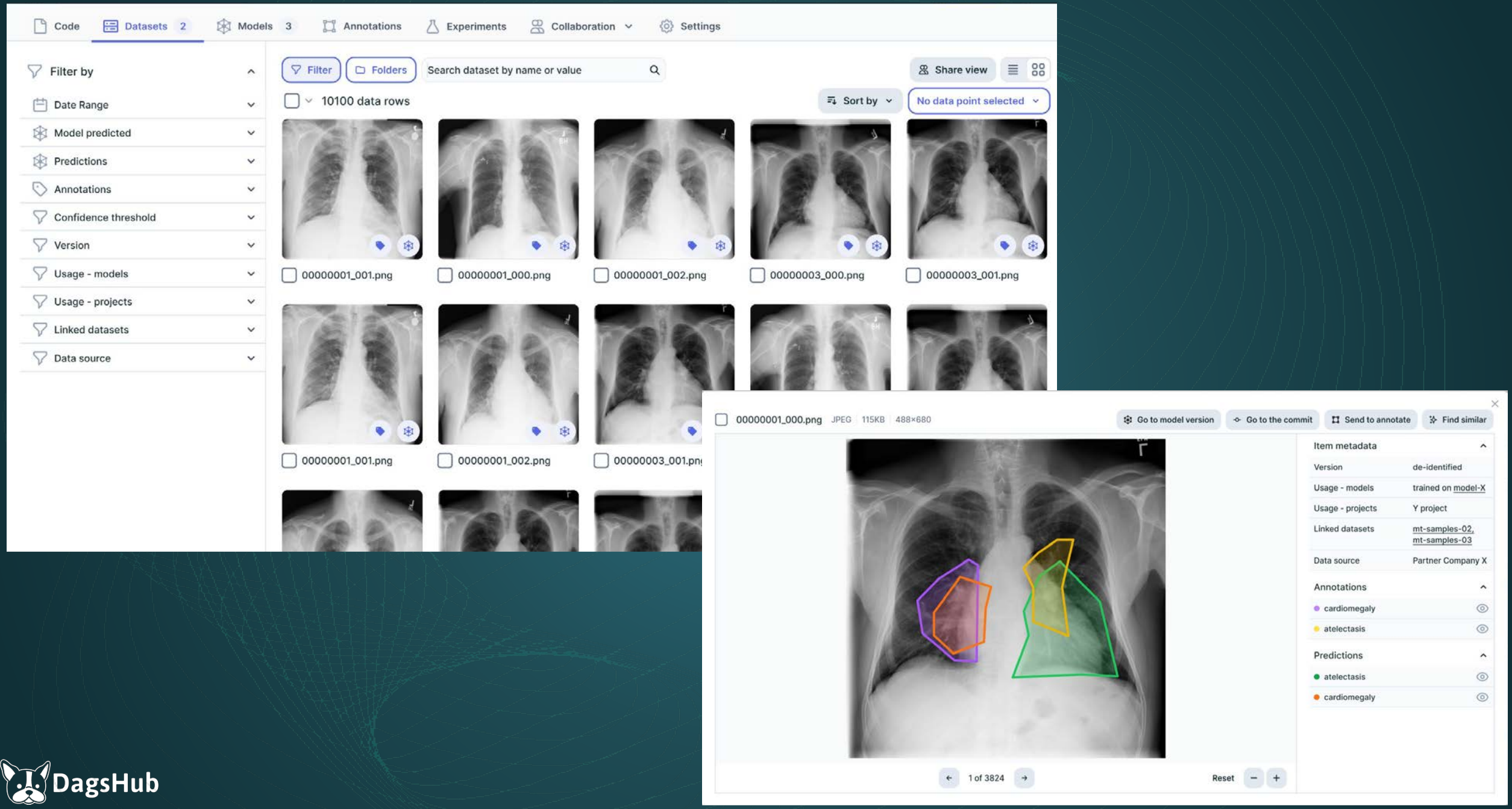

Search public repositories...
Issues
Pull Requests
Resources
Explore
Pricing
Product
Create +
OperationSavta / SavtaDepth
Watch 5
Star 13
Fork 30
Files
Experiments 11
Collaboration 17
Reports
Annotations
Issues 6
Pull Requests 7
Discussions 4
Labels
Milestones
6 Open
3 Closed
Search issues...
#22 Add a Jenkins Pipeline for automations
opened 2 years ago by Dean Pleban
#21 Automate Parameter Tuning enhancement help wanted
opened 2 years ago by Dean Pleban
#17 Add baseline experiment help wanted
opened 2 years ago by Dean Pleban
#5 Additional Model Approaches discussion experiment help wanted
opened 2 years ago by Dean Pleban
#4 Testing additional datasets (or adding more data) enhancement experiment help wanted
opened 2 years ago by Dean Pleban
#3 Add web UI to test the latest model enhancement help wanted
opened 2 years ago by Dean Pleban


Search public repositories...
Issues
Pull Requests
Resources
Explore
Pricing
Product
Create +
Issues 6
Pull Requests 7
Discussions 4
Labels
Milestones
New Issue
#5 Additional Model Approaches
Open
opened 2 years ago by Dean Pleban
Dean Pleban raised an issue 2 years ago
This is a meta issue as there might be many different approaches that optimize for different things.
The initial effort to create an end-2-end pipeline for depth estimation uses FastAI v2 with transfer learning based on a U-Net Learner. This can be trained for longer to improve results, but other model architectures might be relevant as well.
We prioritize models that can be run democratically (i.e. not require huge hardware resources), but suggestions are welcome. We can discuss possible approaches here, and open issues for approaches that we want to try separately to be more manageable.
Ghost commented 2 years ago
In the following pull request, I examined ResNet50 and built a training function for training for a specific number of epochs together with fine-tuning each training epoch:
Moreover, a detect.py file - just choose the epoch number for detection.
The notebook presents the differences between training with and without freezing the last layer.
I hope you will find it useful and if you have any doubts please don't hesitate and contact me.
Labels
discussion
experiment
help wanted
Milestone
No Milestone
Assignee
No assignee
2 Participants
Linked pull requests
No selected pull requests

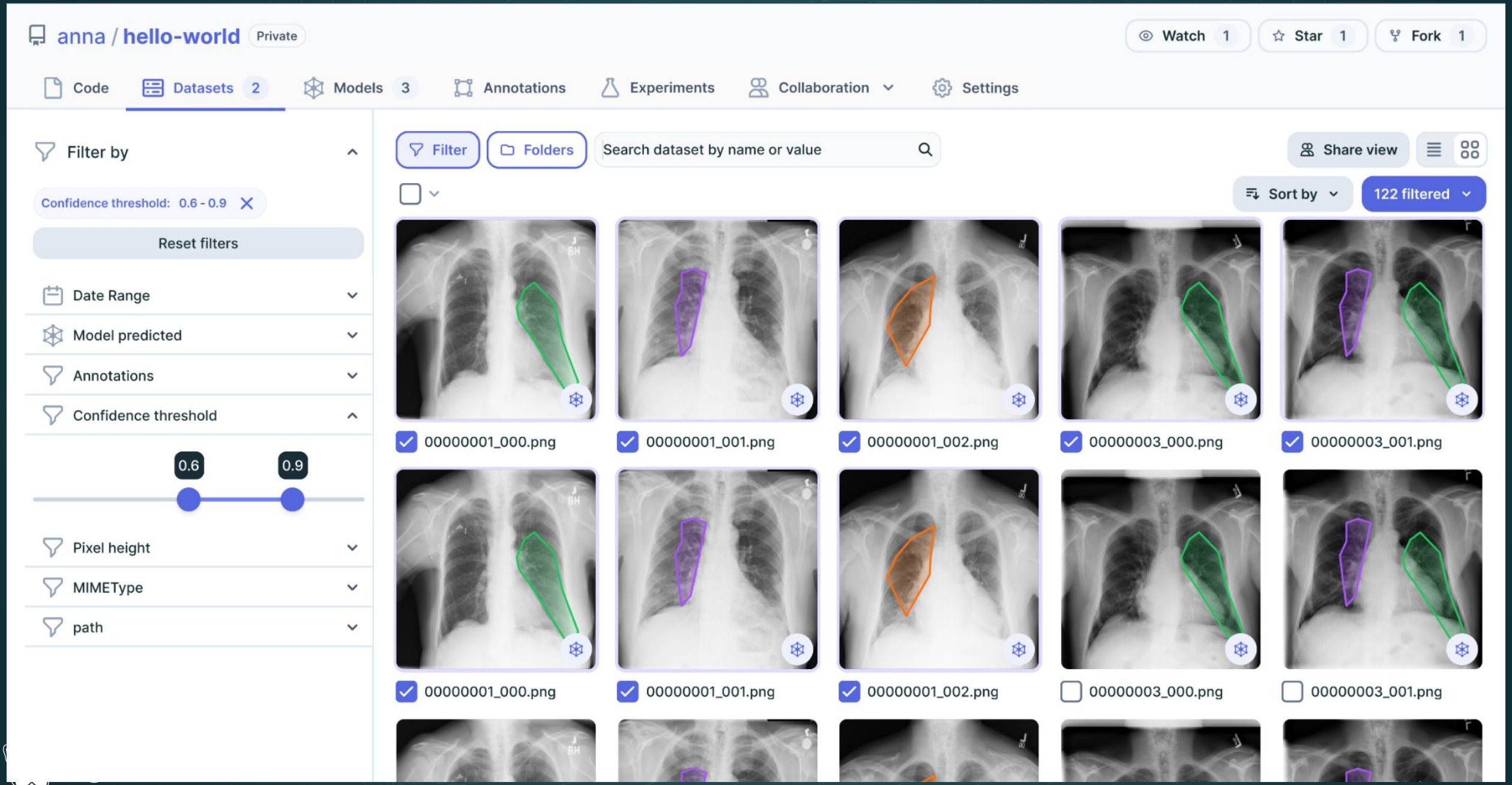
DagsHub

Data visualisations

- Clear display of your data
  - Include metadata, annotations, predictions
- Filtering and sorting abilities
- Simple flows to use subsets


Code
Datasets 2
Models 3
Annotations
Experiments
Collaboration
Settings
Filter by
Date Range
Model predicted
Predictions
Annotations
Confidence threshold
Version
Usage - models
Usage - projects
Linked datasets
Data source
Filter
Folders
Search dataset by name or value
Share view
10100 data rows
Sort by
No data point selected
00000001_001.png
00000001_000.png
00000001_002.png
00000003_000.png
00000003_001.png
00000001_001.png
00000001_002.png
00000003_001.png
00000001_000.png JPEG 115KB 488×680
Go to model version
Go to the commit
Send to annotate
Find similar
Item metadata
Version
de-identified
Usage - models
trained on model-X
Usage - projects
Y project
Linked datasets
mt-samples-02, mt-samples-03
Data source
Partner Company X
Annotations
cardiomegaly
atelectasis
Predictions
atelectasis
cardiomegaly
1 of 3824
Reset

DagsHub


anna / hello-world
Private
Watch 1
Star 1
Fork 1
Code
Datasets 2
Models 3
Annotations
Experiments
Collaboration
Settings
Filter by
Confidence threshold: 0.6 - 0.9
Reset filters
Date Range
Model predicted
Annotations
Confidence threshold
0.6
0.9
Pixel height
MIMEType
path
Filter
Folders
Search dataset by name or value
Share view
Sort by
122 filtered
00000001_000.png
00000001_001.png
00000001_002.png
00000003_000.png
00000003_001.png
00000001_000.png
00000001_001.png
00000001_002.png
00000003_000.png
00000003_001.png

anna / hello-world Private

Watch 1 | Star 1 | Fork 1

Code | Datasets 2 | Models 3 | Annotations | Experiments | Collaboration | Settings

Filter by

Confidence threshold: 0.6 - 0.9

Reset filters

Date Range

Model predicted

Annotations

Confidence threshold

0.6 0.9

Pixel height

MIMEType

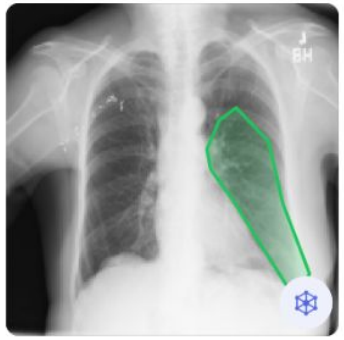
path

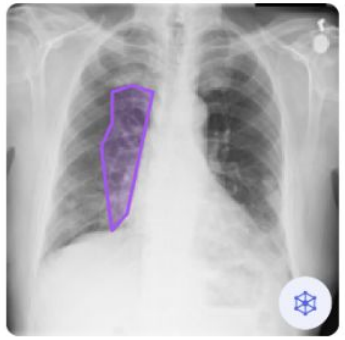
Filter | Folders | Search dataset by name or value

Share view

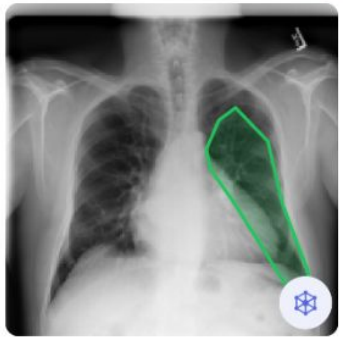
Sort by | 122 filtered

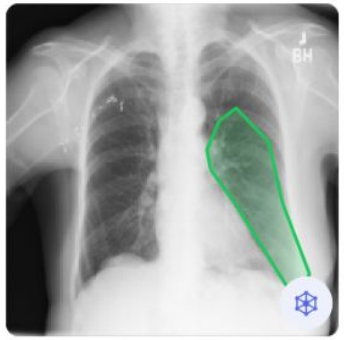
Download as .parquet
Download as .csv
Show python snippet
Find similar
Send for annotation
Add custom tag
Delete

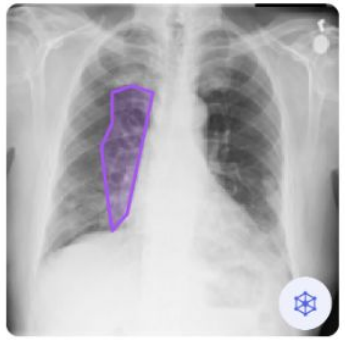

00000001_000.png


00000001_001.png

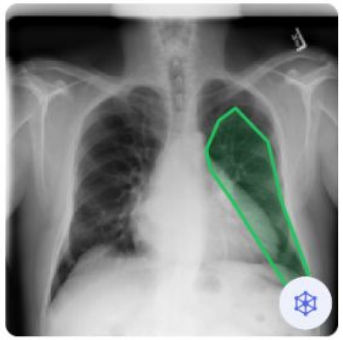
00000001_002.png

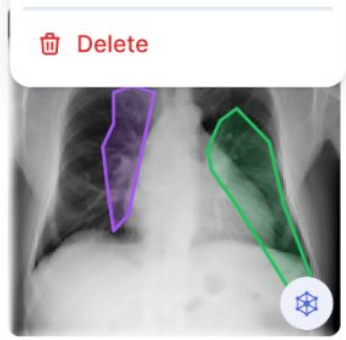

00000003_000.png

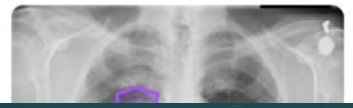

00000001_000.png


00000001_001.png

00000001_002.png


00000003_000.png


00000003_001.png



Data

Experiments

Models

– CONFIDENTIAL –

Experiments

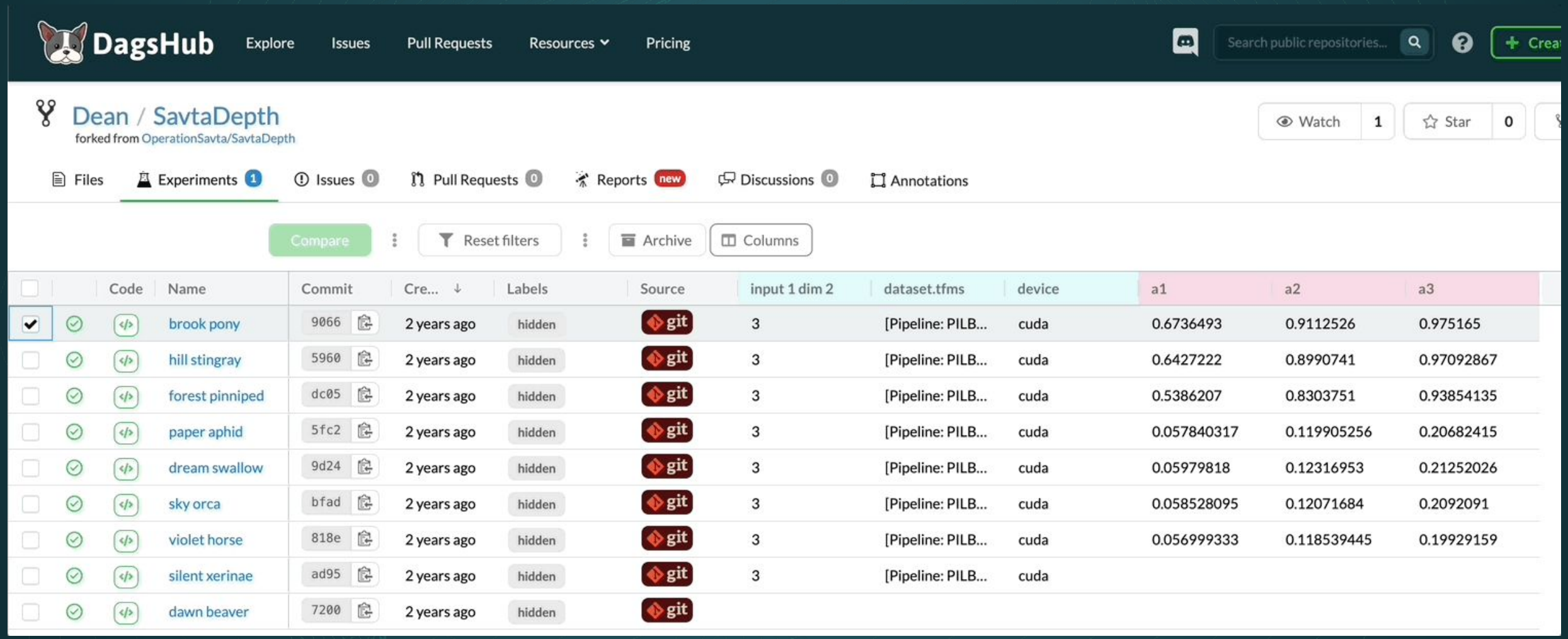
– CONFIDENTIAL –

Experiments challenges

- Experiment **fast** and keep track of **changes**.
- **Communicating** progress and results.

Experiment tracking

- Clear experiments display
  - Hyperparams
  - Metrics
  - Result
  -
- Results comparison
- Visual outputs for (not ML) stakeholders

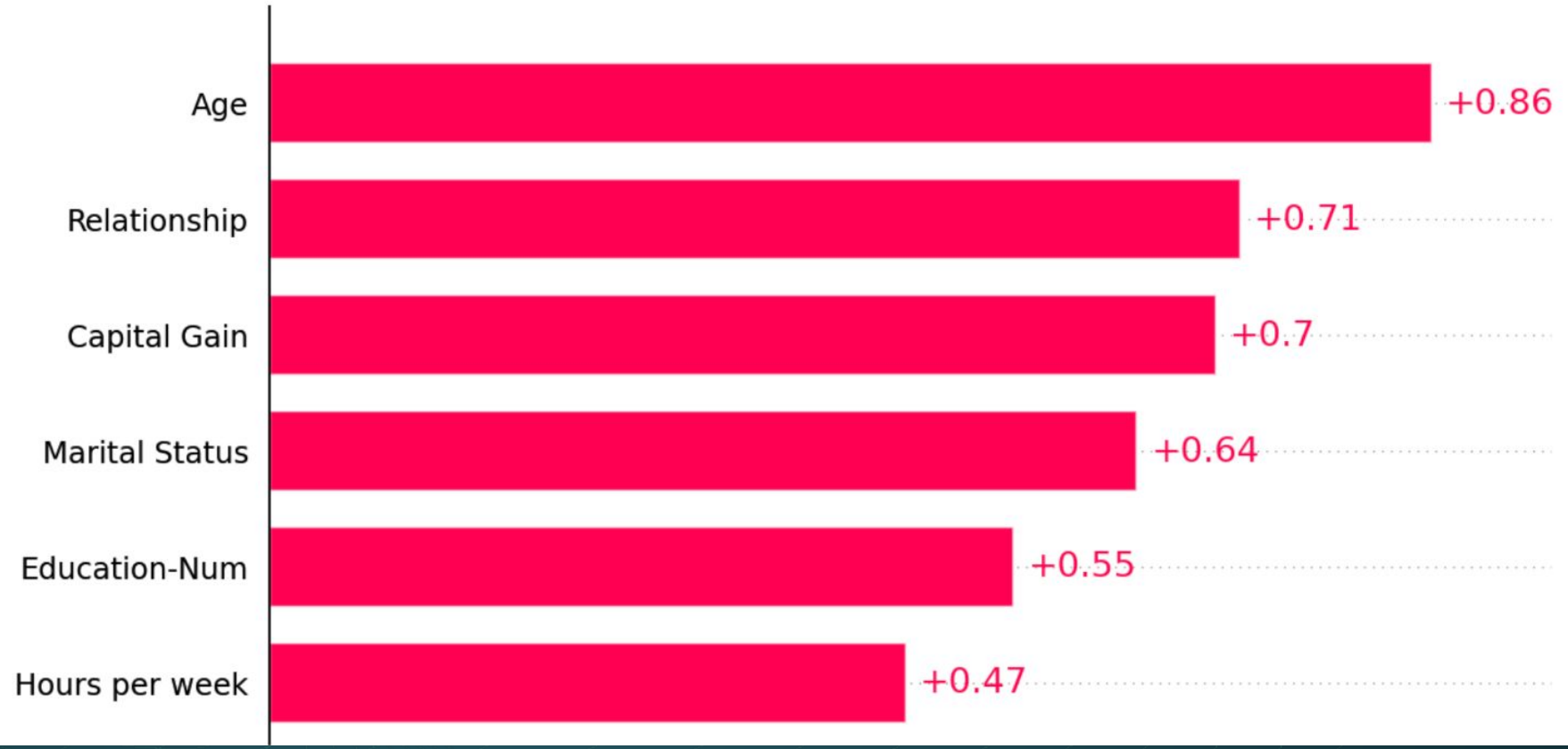

DagsHub
Explore
Issues
Pull Requests
Resources
Pricing
Search public repositories...
Dean / SavtaDepth
forked from OperationSavta/SavtaDepth
Watch 1
Star 0
Files
Experiments 1
Issues 0
Pull Requests 0
Reports new
Discussions 0
Annotations
Compare
Reset filters
Archive
Columns
Code
Name
Commit
Cre...
Labels
Source
input 1 dim 2
dataset.tfms
device
a1
a2
a3
brook pony 9066 2 years ago hidden git 3 [Pipeline: PILB... cuda 0.6736493 0.9112526 0.975165
hill stingray 5960 2 years ago hidden git 3 [Pipeline: PILB... cuda 0.6427222 0.8990741 0.97092867
forest pinniped dc05 2 years ago hidden git 3 [Pipeline: PILB... cuda 0.5386207 0.8303751 0.93854135
paper aphid 5fc2 2 years ago hidden git 3 [Pipeline: PILB... cuda 0.057840317 0.119905256 0.20682415
dream swallow 9d24 2 years ago hidden git 3 [Pipeline: PILB... cuda 0.05979818 0.12316953 0.21252026
sky orca bfad 2 years ago hidden git 3 [Pipeline: PILB... cuda 0.058528095 0.12071684 0.2092091
violet horse 818e 2 years ago hidden git 3 [Pipeline: PILB... cuda 0.056999333 0.118539445 0.19929159
silent xerinae ad95 2 years ago hidden git 3 [Pipeline: PILB... cuda
dawn beaver 7200 2 years ago hidden git

DagsHub

Artifacts

- model
  - MLmodel
  - conda.yaml
  - model.pkl
  - python_env.yaml
  - requirements.txt
- confusion_matrix.png
- lift_curve_plot.png
- precision_recall_curve_plot.png
- roc_curve_plot.png
- shap_beeswarm_plot.png
- shap_feature_importance_plot.png
- shap_summary_plot.png

Full Path:mlflow-artifacts:/0/6264d2944722415997b4b143567936a1/artifacts/shap_feature_importance_plot.png
Size: 76.81KB



DagsHub

The main components MLOps should cover

Data

Experiments

Models

– CONFIDENTIAL –

The main components MLOps should cover

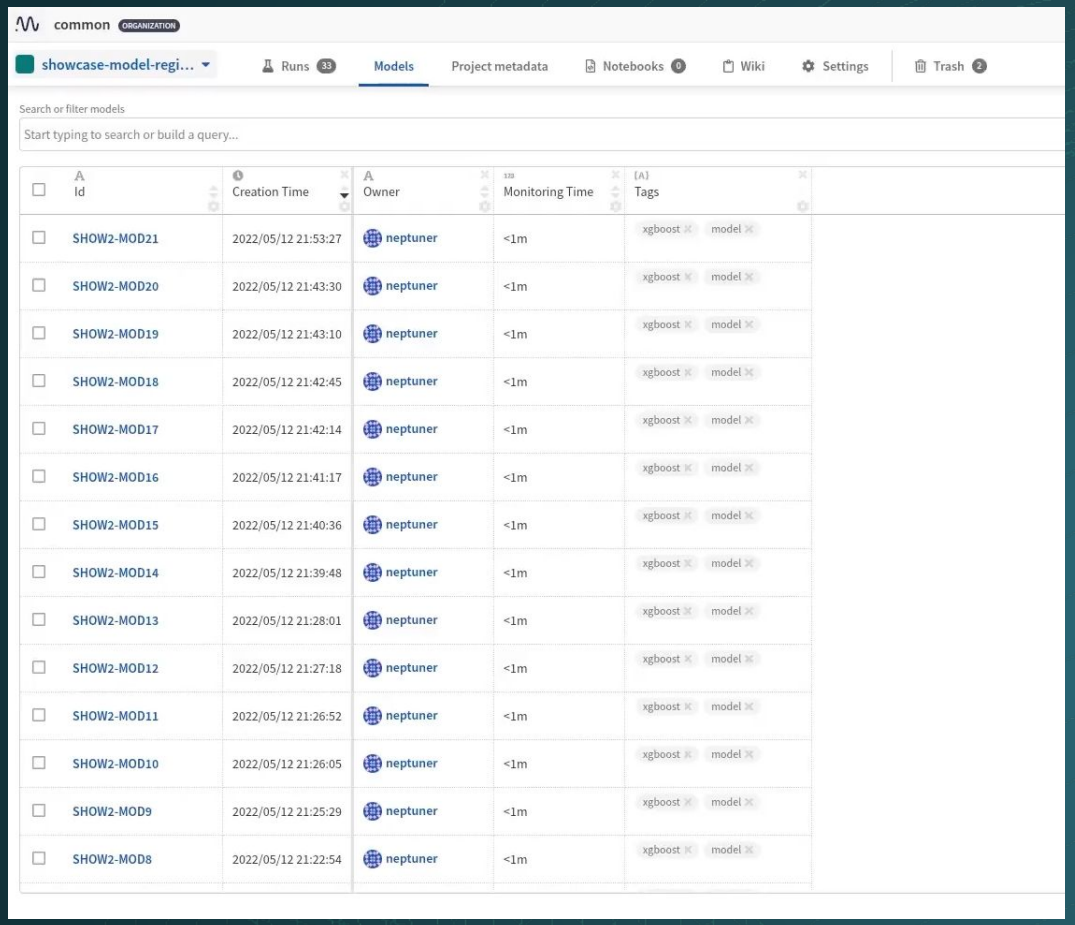
Models

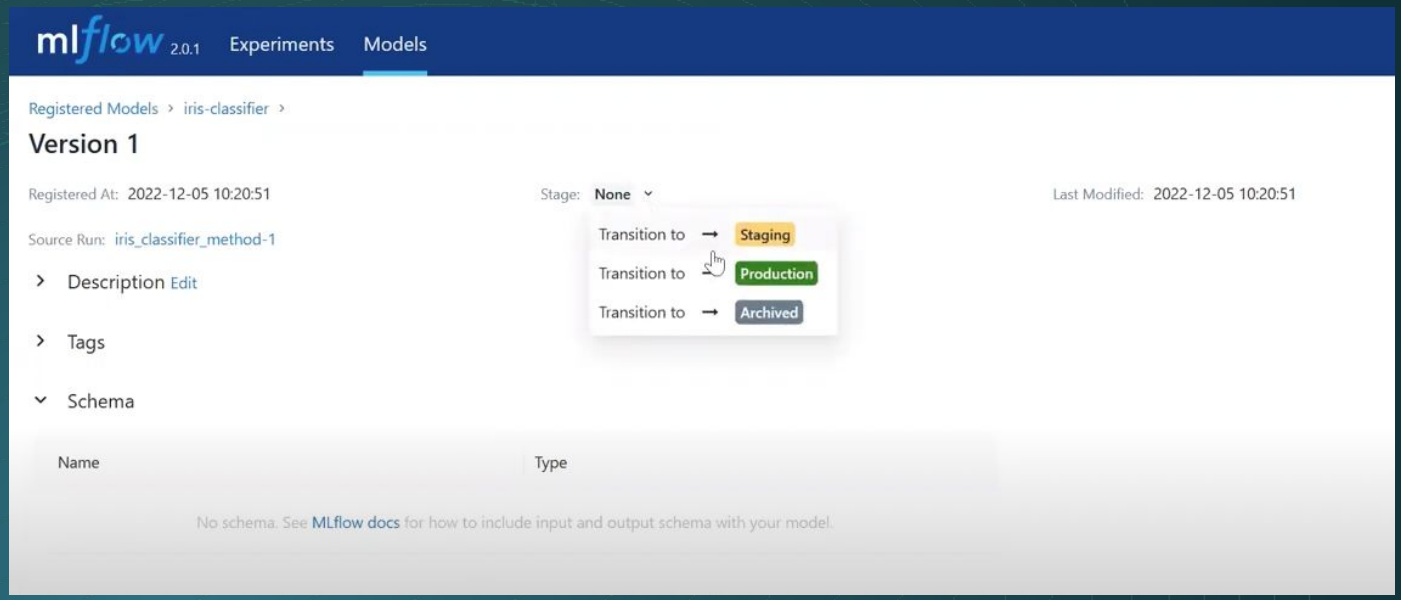
Models challenges

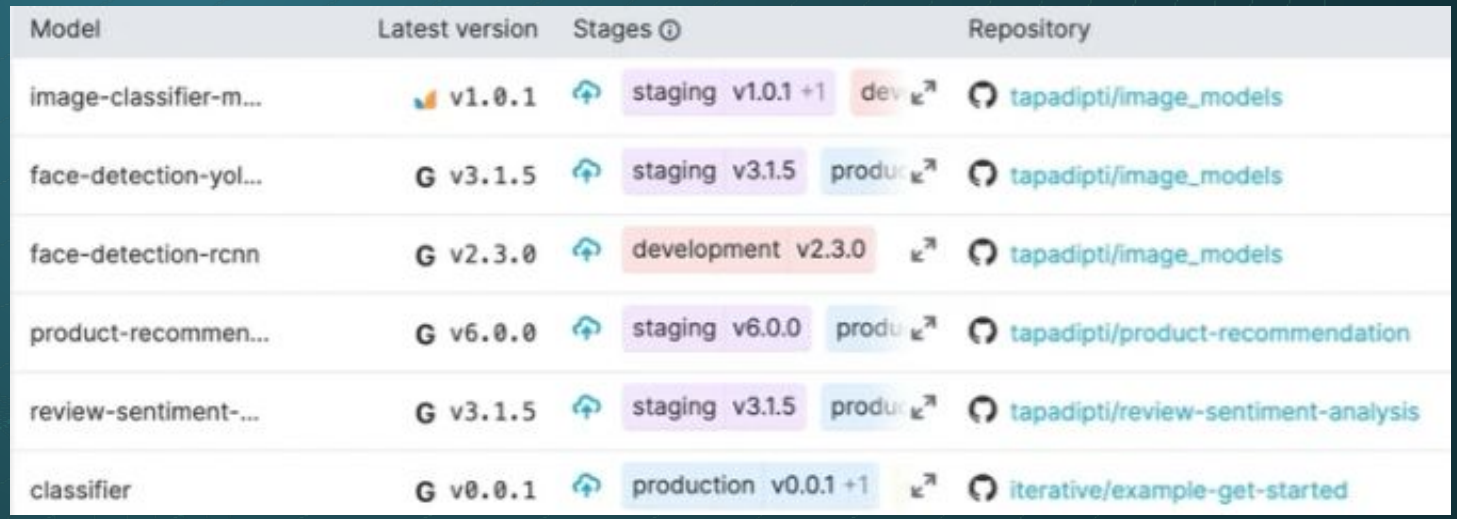
- **Keep track** of your models
  - Access and compare different versions at scale
  - Collaborate
- Reproduce (production)

Model registry

- Collaborate effectively
- Central location to manage and version models at scale
- Easy and intuitive way to deploy production models


common ORGANIZATION
showcase-model-regi...
Runs 33
Models
Project metadata
Notebooks 0
Wiki
Settings
Trash 2
Search or filter models
Start typing to search or build a query...
Id
Creation Time
Owner
Monitoring Time
Tags
SHOW2-MOD21 2022/05/12 21:53:27 neptuner <1m xgboost model
SHOW2-MOD20 2022/05/12 21:43:30 neptuner <1m xgboost model
SHOW2-MOD19 2022/05/12 21:43:10 neptuner <1m xgboost model
SHOW2-MOD18 2022/05/12 21:42:45 neptuner <1m xgboost model
SHOW2-MOD17 2022/05/12 21:42:14 neptuner <1m xgboost model
SHOW2-MOD16 2022/05/12 21:41:17 neptuner <1m xgboost model
SHOW2-MOD15 2022/05/12 21:40:36 neptuner <1m xgboost model
SHOW2-MOD14 2022/05/12 21:39:48 neptuner <1m xgboost model
SHOW2-MOD13 2022/05/12 21:28:01 neptuner <1m xgboost model
SHOW2-MOD12 2022/05/12 21:27:18 neptuner <1m xgboost model
SHOW2-MOD11 2022/05/12 21:26:52 neptuner <1m xgboost model
SHOW2-MOD10 2022/05/12 21:26:05 neptuner <1m xgboost model
SHOW2-MOD9 2022/05/12 21:25:29 neptuner <1m xgboost model
SHOW2-MOD8 2022/05/12 21:22:54 neptuner <1m xgboost model

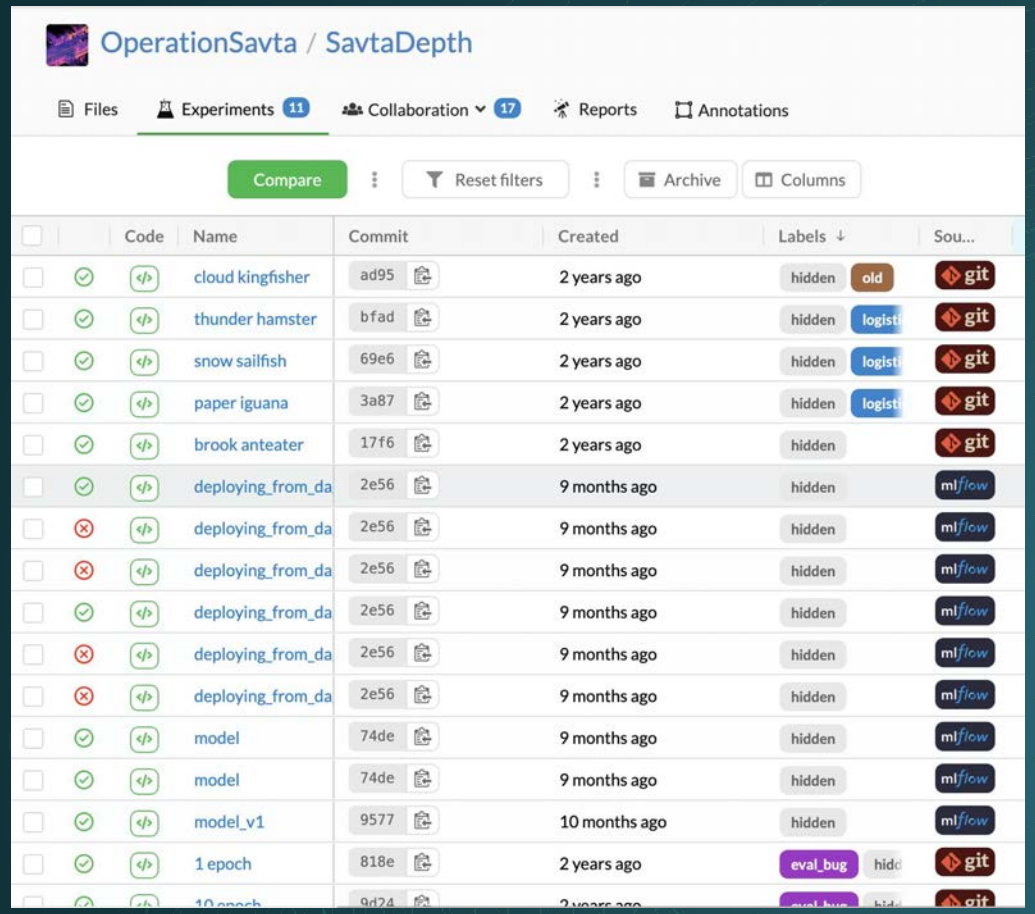

mlflow 2.0.1
Experiments
Models
Registered Models › iris-classifier ›
Version 1
Registered At: 2022-12-05 10:20:51
Stage: None
Last Modified: 2022-12-05 10:20:51
Source Run: iris_classifier_method-1
Transition to → Staging
Transition to → Production
Transition to → Archived
Description Edit
Tags
Schema
Name
Type
No schema. See MLflow docs for how to include input and output schema with your model.


Model
Latest version
Stages
Repository
image-classifier-m... v1.0.1 staging v1.0.1 +1 tapadipti/image_models
face-detection-yol... v3.1.5 staging v3.1.5 tapadipti/image_models
face-detection-rcnn v2.3.0 development v2.3.0 tapadipti/image_models
product-recommen... v6.0.0 staging v6.0.0 tapadipti/product-recommendation
review-sentiment-... v3.1.5 staging v3.1.5 tapadipti/review-sentiment-analysis
classifier v0.0.1 production v0.0.1 +1 iterative/example-get-started

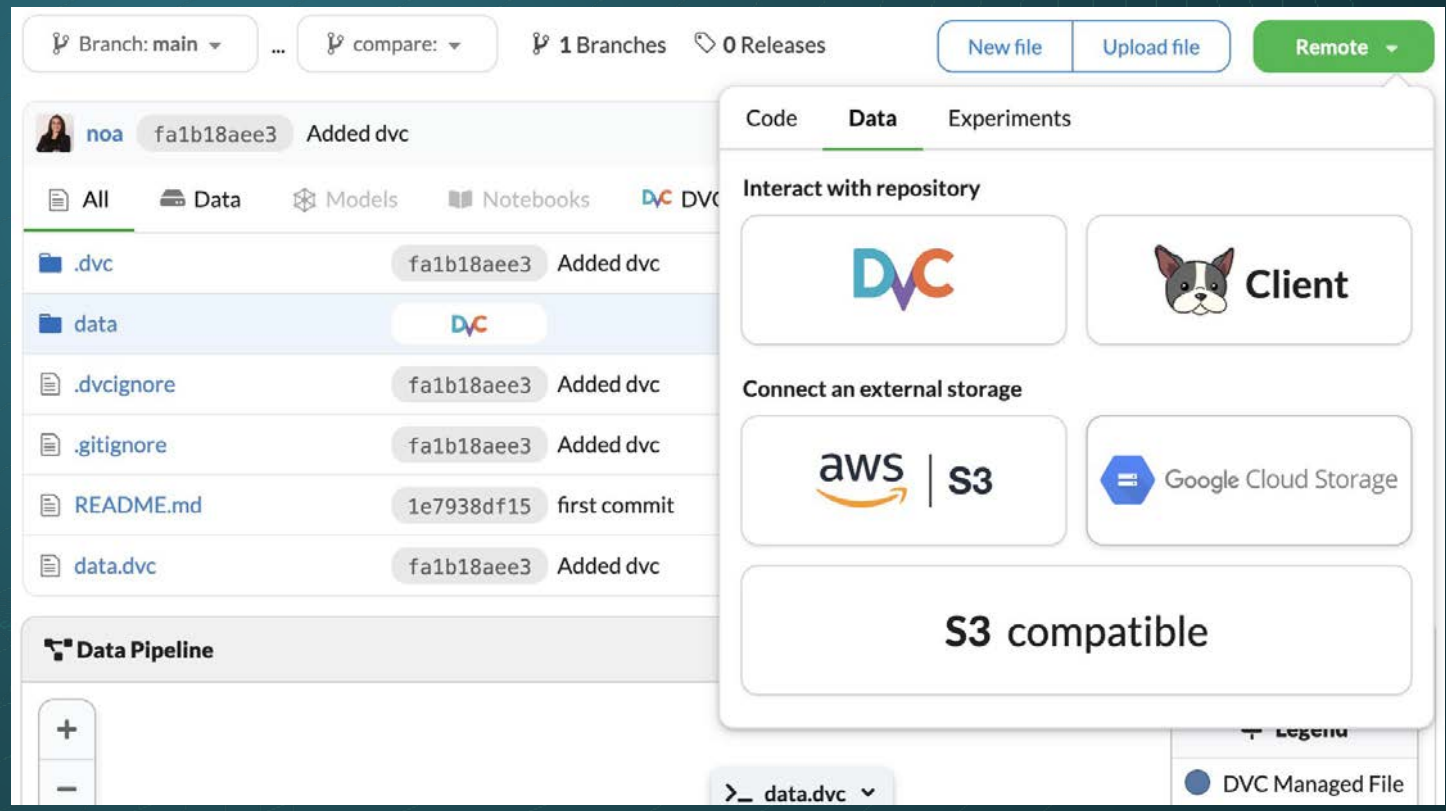
DagsHub

General UX tips

- Simplification
  - Keep things clean
  - Don’t add unused abilities
- Integrate tools that ML engineers love
- Think about the flows first


OperationSavta / SavtaDepth
Files
Experiments 11
Collaboration 17
Reports
Annotations
Compare
Reset filters
Archive
Columns
Code
Name
Commit
Created
Labels
Sou...
cloud kingfisher
ad95
2 years ago
hidden
old
thunder hamster
bfad
2 years ago
hidden
snow sailfish
69e6
2 years ago
hidden
paper iguana
3a87
2 years ago
hidden
brook anteater
17f6
2 years ago
hidden
deploying_from_da
2e56
9 months ago
hidden
model
74de
9 months ago
hidden
model_v1
9577
10 months ago
hidden
1 epoch
818e
2 years ago
eval_bug

Prometheus
git
mlflow
DVC


Branch: main
compare:
1 Branches
0 Releases
New file
Upload file
Remote
noa
fa1b18aee3
Added dvc
All
Data
Models
Notebooks
.dvc
data
.dvcignore
.gitignore
README.md
1e7938df15
first commit
data.dvc
Data Pipeline
Code
Data
Experiments
Interact with repository
Client
Connect an external storage
aws S3
Google Cloud Storage
S3 compatible
data.dvc
DVC Managed File

DagsHub

What is your favorite MLOps experience? tell me!

Noa Goldman

noa@dagshub.com