iterative

Visualize Metrics & Plots
with DVC and Studio

Lesson 5

DVC tools for Data Scientists & Analysts

2021

Course lessons

Lesson 1. Course Introduction

Lesson 2. Practices and Tools for Efficient Collaboration in ML projects

Lesson 3. Pipelines Automation and Configuration Management

Lesson 4. Versioning Data and Models

Lesson 5. Visualize Metrics & Compare Experiments with DVC and Studio

Lesson 6. Experiments Management and Collaboration

Lesson 7. Tools for Deep Learning Scenarios

Lesson 8. Review Advanced Topics and Use Cases

Lesson Outline

- ML experiment issues & requirements
- Metrics tracking with DVC
- Visualize metrics and plots with DVC
- Visualize metrics and compare experiments with Studio

ML Experiment Issues & Requirements

Experiment Management

- Provide reproducibility
- Pipeline automation
- Metrics & artifacts tracking
- Create, run & organize experiments
  - Configuration management
  - Search & compare
  - Saving & sharing views of experiment dashboard
  - Documentation & reports
- Model lifecycle management
- Team collaboration

Metrics Tracking with DVC

dvc.yaml: setup metrics and plots

```
stages:
 train:
   cmd: python train.py --config=params.yaml
   deps:
   - features.csv
   outs:
   - model.pkl
   metrics:
   - metrics.json:
       cache: false
   plots:
   - auc.json:
       cache: false
```

Add metrics/plots files in dvc.yaml

dvc.yaml: setup metrics and plots

```
stages:
 train:
   cmd: python train.py --config=params.yaml
   deps:
   - features.csv
   outs:
   - model.pkl
   metrics:
   - metrics.json:
         cache: false
   plots:
   - auc.json:
         cache: false
```

Set cache to False to keep file in Git history

Compare experiments: dvc metrics show

run command

```
$ dvc metrics show
```

output

metrics.json:

```
{
        "f1_score": 0.7861833464670345,
        "confusion_matrix": {
                "classes": ["setosa", "versicolor", "virginica"],
                "matrix":
                        [[23, 0, 0],
                         [0, 8, 0],
                         [0, 11, 18]]
                }
}
```

A metrics file specified in dvc.yaml

Compare experiments: dvc metrics diff

run command

```
$ dvc metrics diff
```

output

```
Path          Metric  Value   Change
metrics.json  AUC     0.9671  0.0028
metrics.json  TP      531     4
```

Show changes in metrics

Live code example

Metrics Tracking with DVC

Plots and graphics with DVC

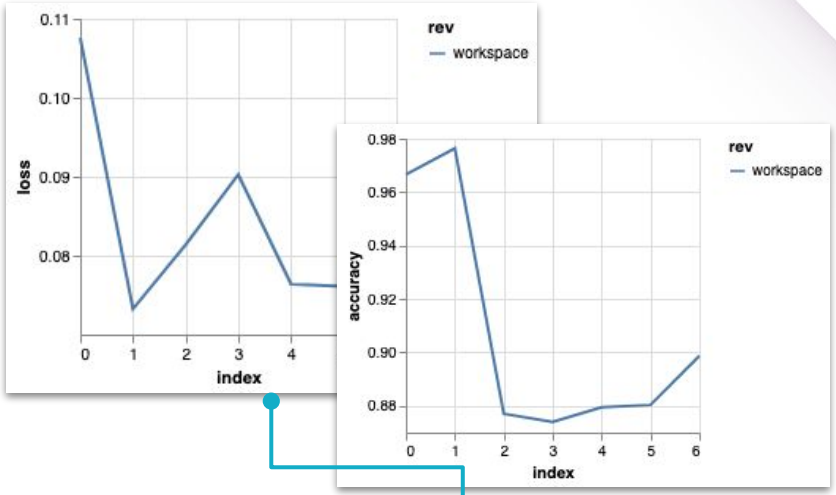
Compare experiments: dvc plots show

output

```
epoch,accuracy,loss,val_accuracy,val_loss
0,0.9418667,0.19958884770199656,0.9679,0.10217399864746257
1,0.9763333,0.07896138601688048,0.9768,0.07310650711813942
2,0.98375,0.0524111119088716 8,0.9788,0.06665669009438716
3,0.98801666,0.03681169906261687,0.9781,0.06697812260198989
4,0.99111664,0.027362171787042946,0.978,0.07385754839298315
5,0.9932333,0.02069501801203781,0.9771,0.08009233058886166
6,0.9945,0.017702101902437668,0.9803,0.07830339228538505
7,0.9954,0.01396906608727198,0.9802,0.07247738889862157
```

run command

```
$ dvc plots show logs.csv
```

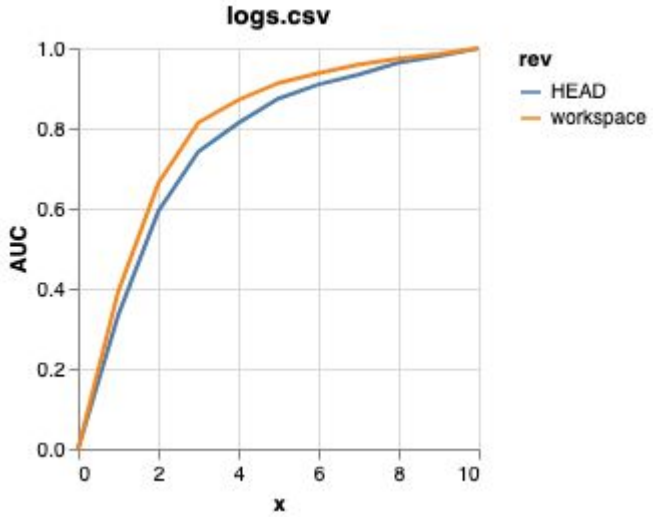


Use the -y option to change the field to plot

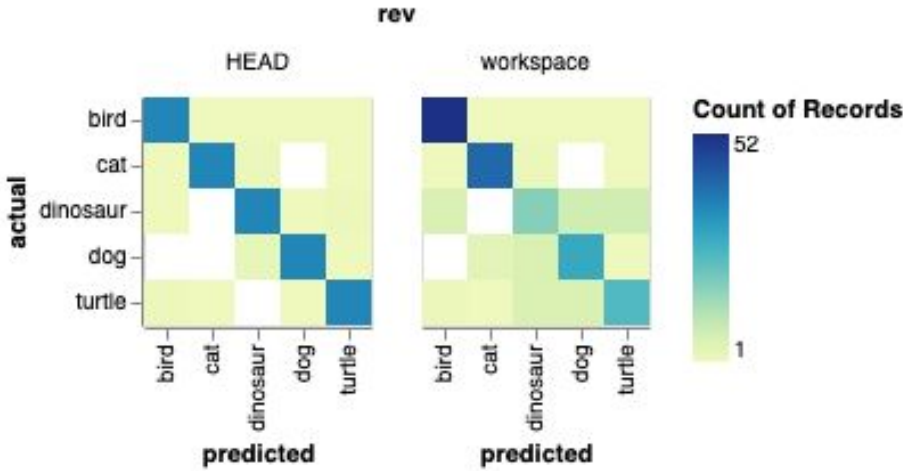
https://dvc.org/doc/command-reference/plots#example-tabular-data

Compare experiments: dvc plots diff

Linear plot diff example



Confusion matrix diff example



https://dvc.org/doc/command-reference/plots/diff

Add and modify plot templates

DVC project root folder

```
.dvc
  cache/
  plots/
    confusion.json
    default.json
    scatter.json
    smooth.json
  tmp/
...
```

```
{
    "$schema":
"https://vega.github.io/schema/vega-lite/v4.json",
    "data": {
        "values": "<DVC_METRIC_DATA>"
    },
    "title": "<DVC_METRIC_TITLE>",
    "mark": "rect",
    "width": 500,
    "height": 500,
    "encoding": {
        ....
    }
}
```

Live code example

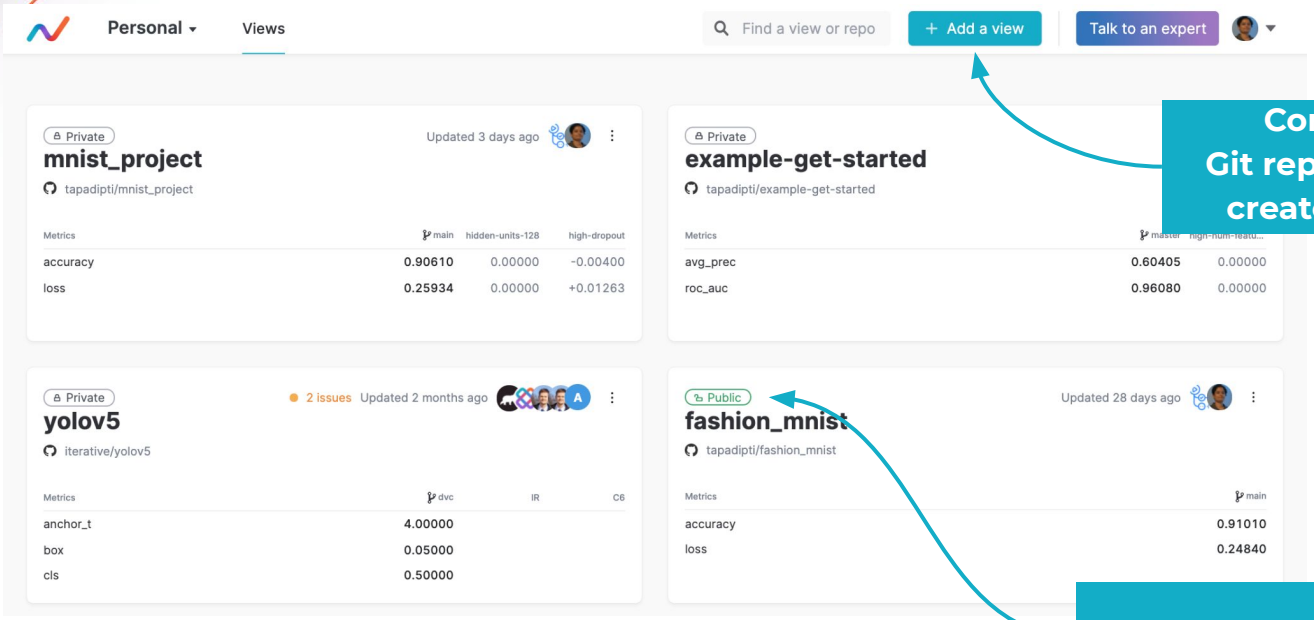
Plots and graphics with DVC

Metrics Tracking and Plots with Studio

Metrics need to be shareable, accessible

Visualize, track, and share ML models across your team

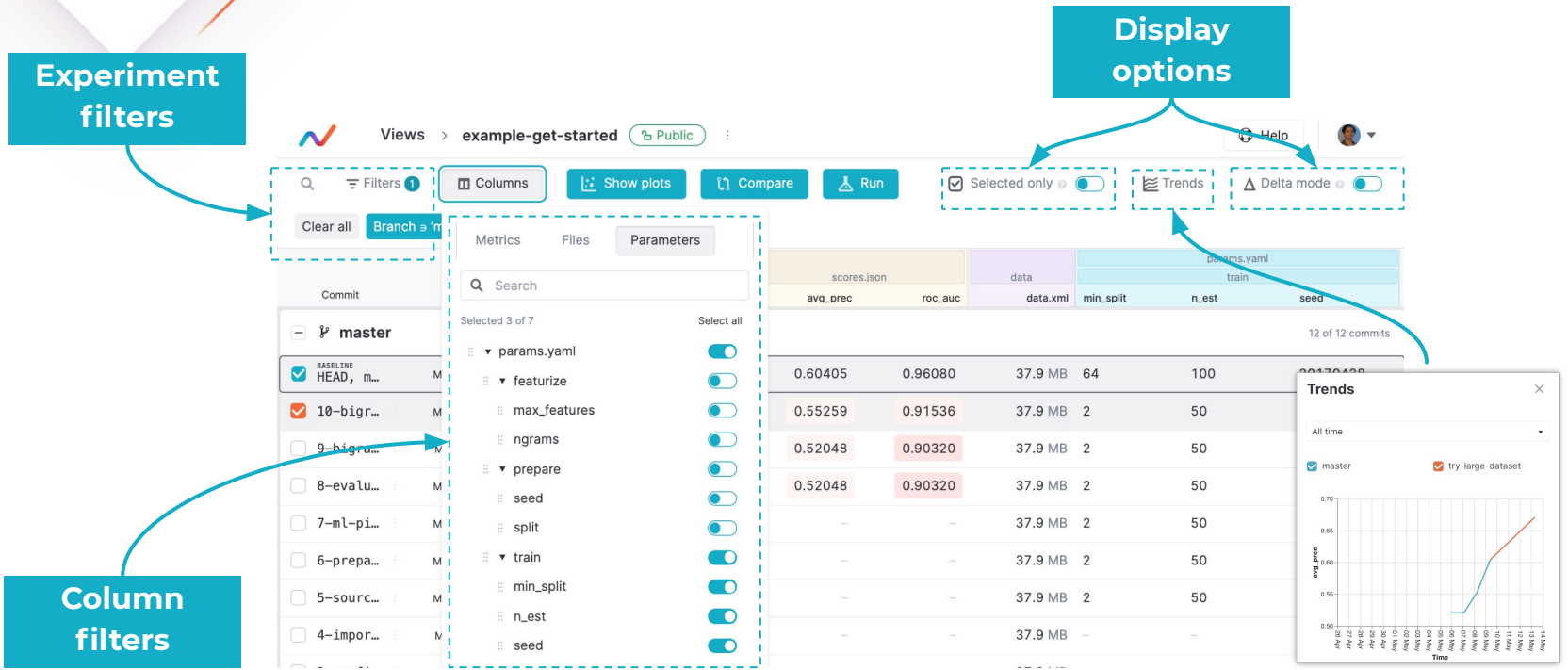
Studio: experiments dashboard



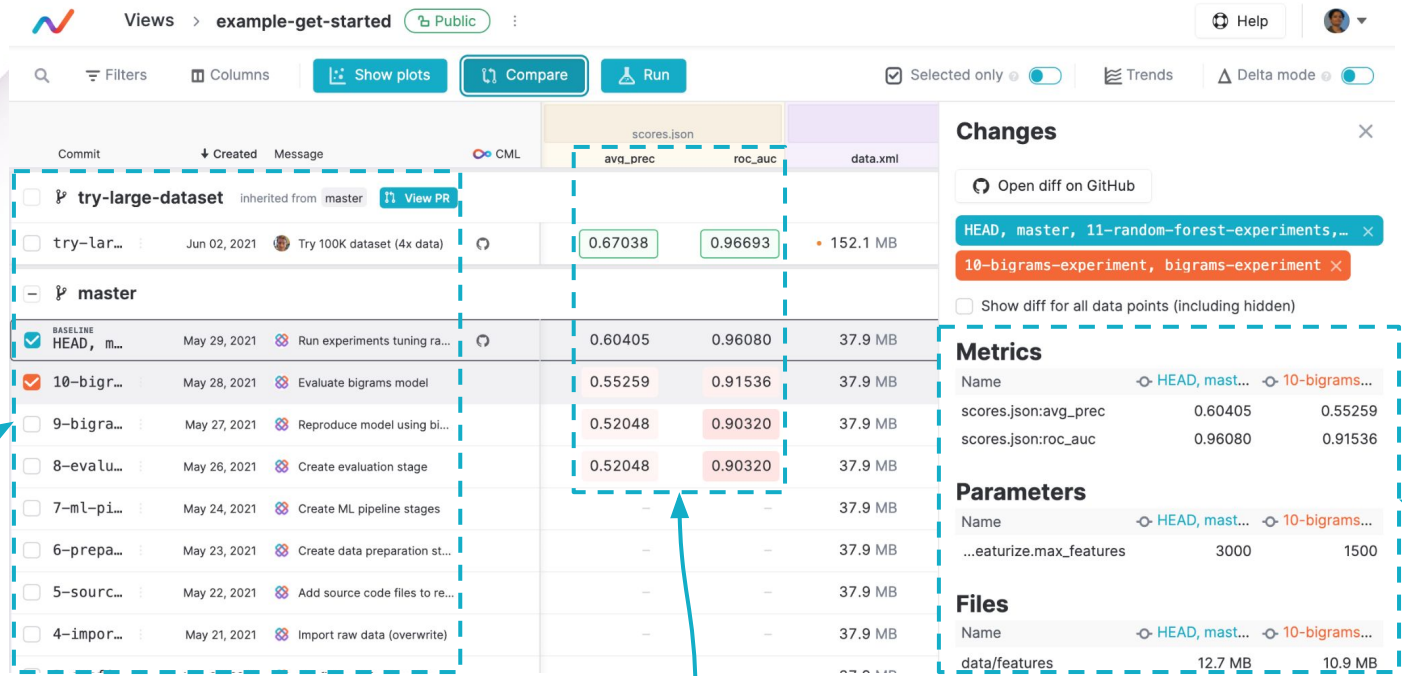
Connect to Git repository and create a "View"

Make views public

Studio: experiments table



Studio: experiments and metrics tracking



List of experiments

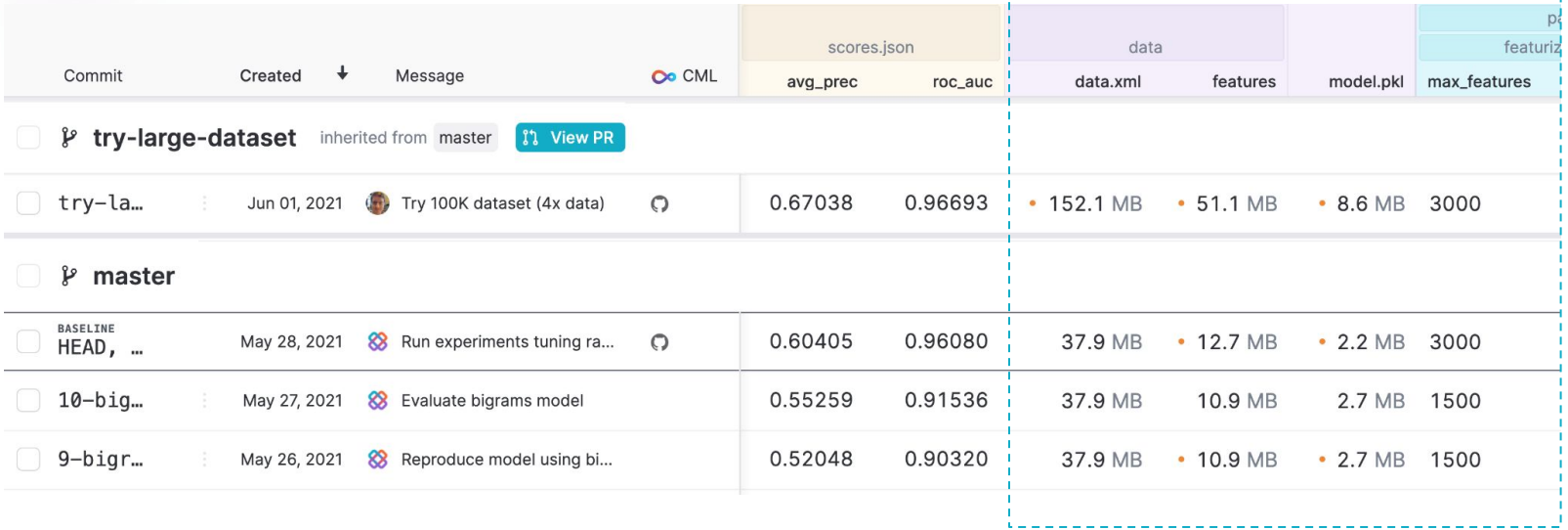
Set metric goals

Track changes

Studio: show plots

Show plots for metrics differences

Studio: track params and data changes



Commit	Created	Message	CML	scores.json		data			p / featuriz
				avg_prec	roc_auc	data.xml	features	model.pkl	max_features
try-large-dataset inherited from master View PR									
try-la…	Jun 01, 2021	Try 100K dataset (4x data)		0.67038	0.96693	152.1 MB	51.1 MB	8.6 MB	3000
master									
BASELINE HEAD, …	May 28, 2021	Run experiments tuning ra…		0.60405	0.96080	37.9 MB	12.7 MB	2.2 MB	3000
10-big…	May 27, 2021	Evaluate bigrams model		0.55259	0.91536	37.9 MB	10.9 MB	2.7 MB	1500
9-bigr…	May 26, 2021	Reproduce model using bi…		0.52048	0.90320	37.9 MB	10.9 MB	2.7 MB	1500

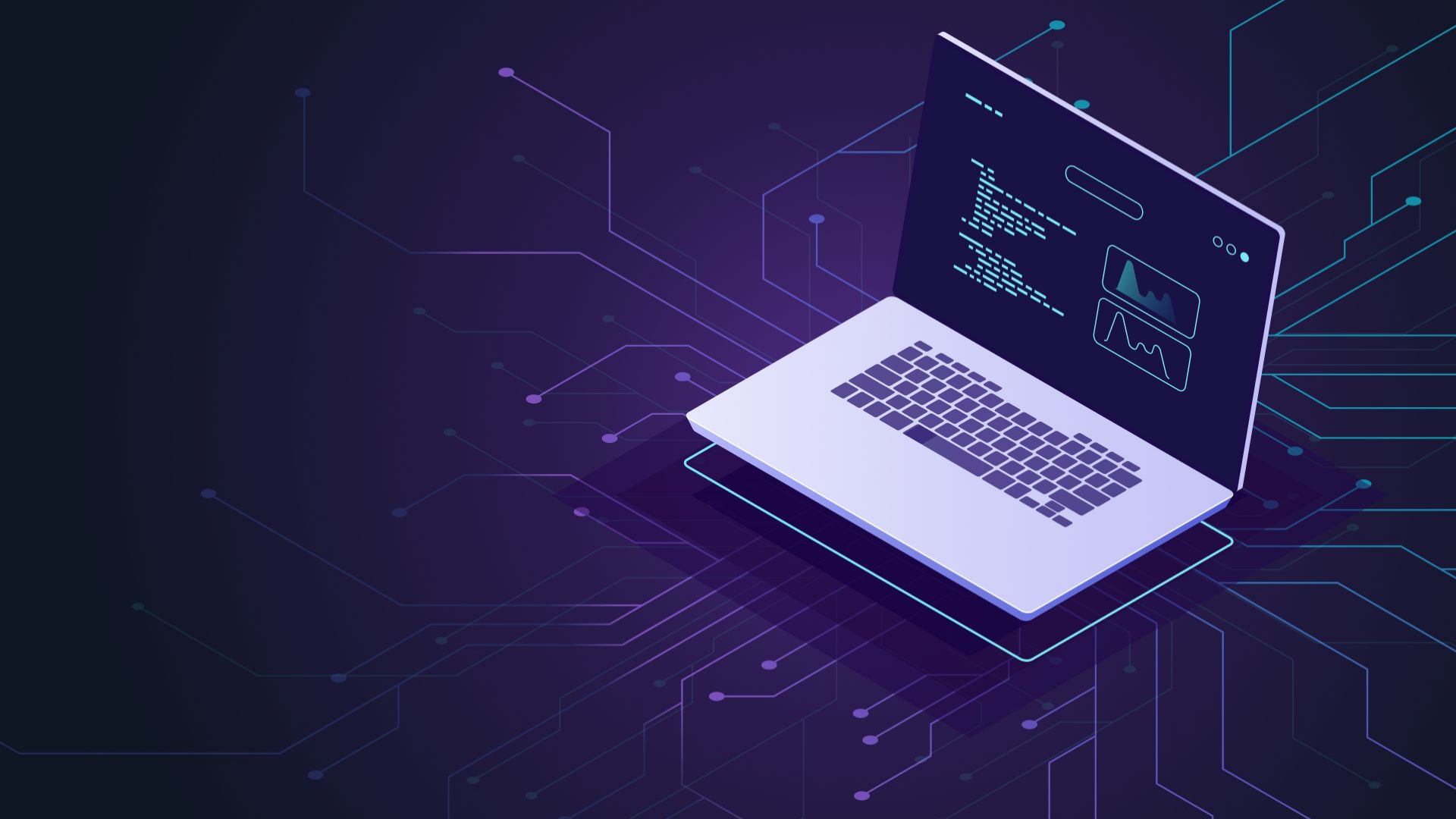
Live code example

Metrics Tracking and Plots with DVC Studio

What have we learned?

What have we learned?

1. Requirements for ML experiment management
2. How to setup metrics and plots with DVC
3. How to visualize metrics and plots with DVC and DVC Studio



Links

- Data Science blueprint
  https://data-science-blueprint.readthedocs.io/en/latest/presentation/schema.html