

PROFINIT

NI-MLP MLOps

Samuel Fabo, Jan Palášek, Sergii Stamenov, Tomáš Duda

4.12.2023

Outline

PROFINIT

- › Motivation
- › DevOps
- › Version Control
- › Dependency Management
- › Experimenting
- › Mlflow
- › ML Pipelines

CRISP-DM

PROFINIT



Hidden Debts of ML Systems

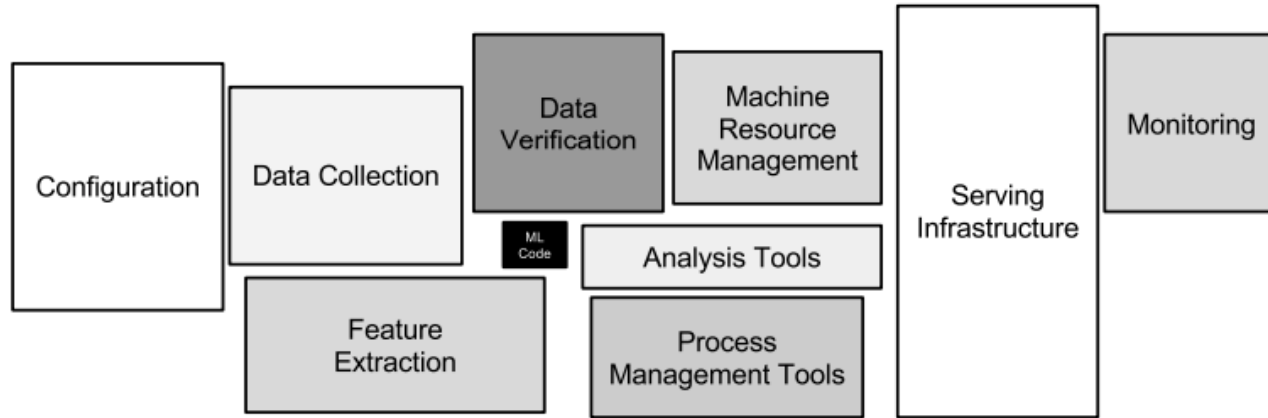


Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

1

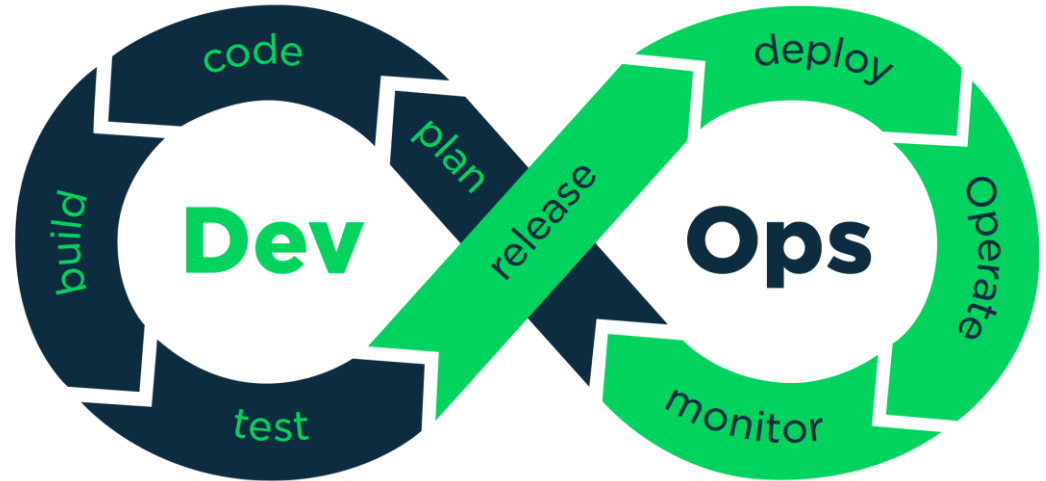
DevOps

DevOps – Development & Operations

PROFINIT

› Key Steps:

1. Plan
2. Develop
3. Test against real world data
4. Release
5. Deploy to cloud
6. Monitor – obtain feedback



DevOps

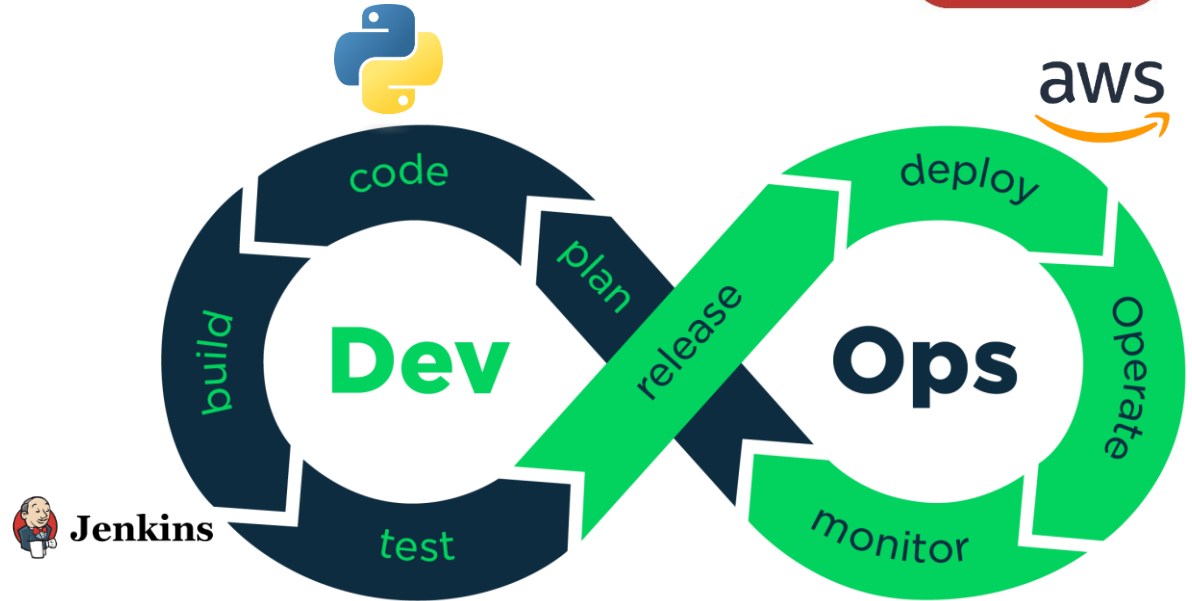
PROFITIT

> Key principles:

- shared ownership,
- workflow automation
- rapid feedback

> MLops

- Design
- Model Engineering
- Model Deployment and Monitoring



2

Version Control



Version control

- › Commenting out some old code and leave it 'just in case'

```
# model = RandomForest(n=100)
# model = RandomForest(n=1000)
# model.fit(X, y)

model = LogisticRergrression()
model.fit(X, y)
```

- › Using Subfolders

```
ls -l
total 0
drwxr-xr-x 1 sfabo titanic_v1/
drwxr-xr-x 1 sfabo titanic_v2/
drwxr-xr-x 1 sfabo titanic_v2.1/
```

Version control

› Branches in git

```
$ git branch --list  
  master  
  titanic_v1  
  titanic_v2  
* titanic_v2.1
```

› Advantages

- No code duplication
- Collaboration
- History

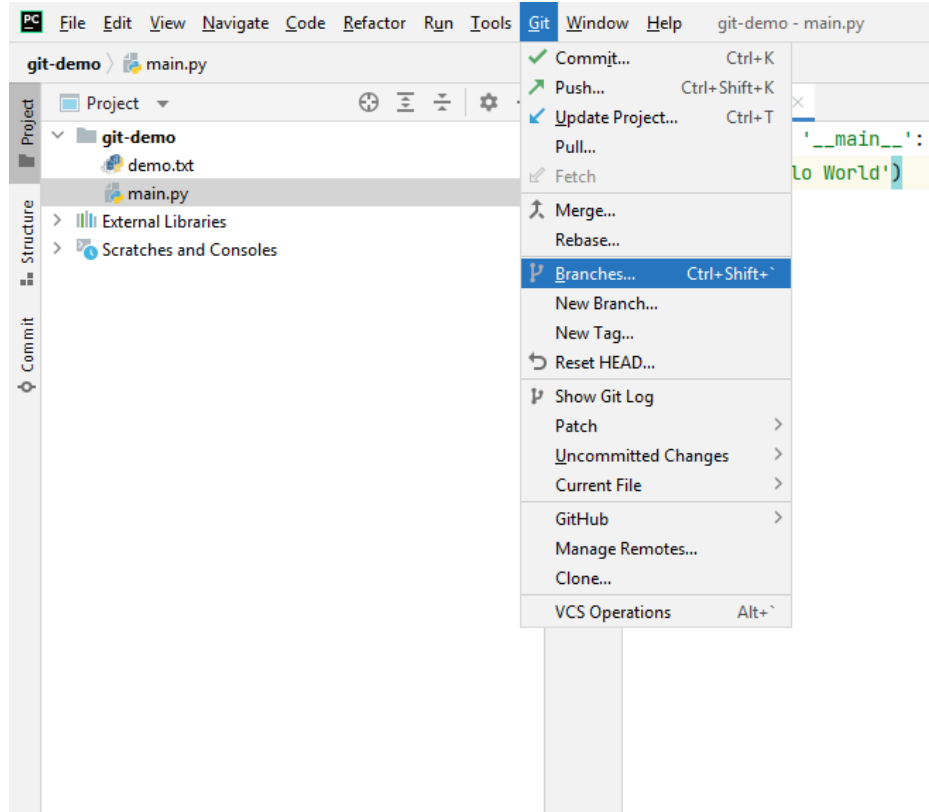
› Disadvantages

- ???

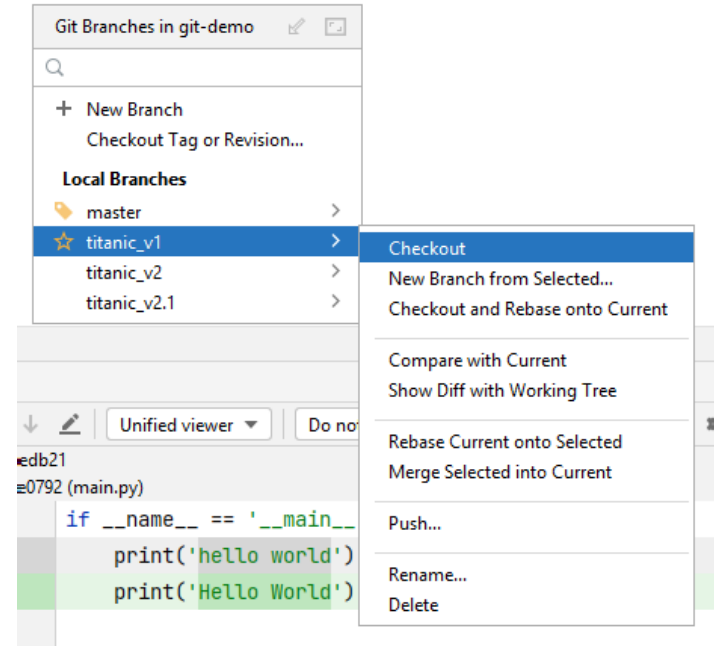
› Alternatives?

PyCharm git Integration

PROFINIT



A screenshot of the PyCharm interface showing the Git menu. The menu is open, displaying various options such as Commit..., Push..., Update Project..., Pull..., Fetch, Merge..., Rebase..., Branches..., New Branch..., New Tag..., Reset HEAD..., Show Git Log, Patch, Uncommitted Changes, Current File, GitHub, Manage Remotes..., Clone..., and VCS Operations. The 'Branches...' option is highlighted, and a sub-menu is visible showing the file content: `'__main__':` and `Lo World')`.



A screenshot of the 'Git Branches in git-demo' dialog box. The dialog shows a list of local branches: master, titanic_v1, titanic_v2, and titanic_v2.1. The 'titanic_v1' branch is selected, and a context menu is open over it, showing options like Checkout, New Branch from Selected..., Checkout and Rebase onto Current, Compare with Current, Show Diff with Working Tree, Rebase Current onto Selected, Merge Selected into Current, Push..., Rename..., and Delete.

VSCode git Extension

PROFINIT

SOURCE CONTROL

✓ SOURCE CONTROL

Message (Ctrl+Enter to commit on 'main')

✓ Commit

✓ Staged Changes 1

- ◆ .gitignore practice/week_9_mlops A

✓ Changes 11

- gender_baseline.csv U
- titanic_test.csv U
- titanic_train.csv U
- .python-version practice/week_3_titanic U
- NDBI048-data_understanding_solutions.ipyn... M
- NDBI048-mod1_titanic.ipynb practice/week_7... M
- NDBI048-regression_in_python_examples.ip... M
- NDBI048_mlops_titanic_solutions.ipynb prac... M
- NDBI048_mlops_titanic.ipynb practice/week_... M
- ① README.md practice/week_9_mlops U
- sandbox.ipynb practice/week_9_mlops U

Select a ref to checkout

- + Create new branch...
- + Create new branch from...
- ↻ Checkout detached...
- 🔗 main f01818d3
- ☁ origin/main Remote branch at f01818d3
- ☁ origin/HEAD Remote branch at f01818d3
- ☁ origin/master Remote branch at 65c82668

Using DVC pipelines

- › Data Version Control, or DVC:
 - *Data and ML experiment management tool that takes advantage of the existing engineering toolset that you're already familiar with (Git, CI/CD, etc.)*
- › Steps with deps. between stages
 - If stage doesn't change, skipping
- › `dvc repro`

```
# dvc.yaml
stages:
  prepare:
    cmd: python src/prepare.py data/data.xml
    deps:
      - src/prepare.py
      - data/data.xml
    outs:
      - data/prepared
  train:
    cmd: python src/train.py data/prepared
    deps:
      - data/prepared
    outs:
      - model.pkl
```

3

Project Dependencies

Dependencies

- › Let's build a model starting from a scratch.
 - `ModuleNotFoundError: No module named 'pandas'`
 - `pip install pandas`
- › Scripts don't work
- › Library's API has changed
- › WORSE: scripts do work but produce **different** results
- › Solution?

requirements.txt

- › Packages in one place

```
Matplotlib  
numpy  
pandas  
scikit-learn  
...
```

- › What can go wrong?

requirements.txt

- › Specific versions of the packages in one place

```
$ pip freeze
```

```
matplotlib==3.6.2
```

```
numpy==1.23.5
```

```
pandas==1.5.2
```

```
scikit-learn==1.1.3
```

setup.py

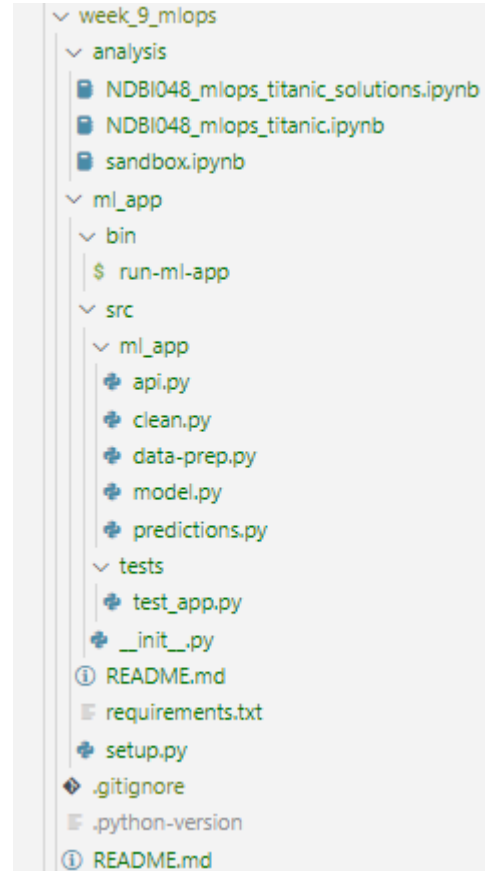
- Many good blogs with best practices out in the wild

```
from setuptools import find_packages, setup

setup(
    name='ml_app',
    version='0.0.1',
    python_requires='>=3.9',
    packages=find_packages(exclude=['tests']),
    scripts=['bin/run-ml-app'],
    include_package_data=True
)
```

Understandable source code

- › Always code as if the guy who ends up maintaining your code will be a *violent psychopath* who knows where you live.
- › Code is more often read than wrote
- › Explicit is better than implicit



```
week_9_mlops
├── analysis
│   ├── NDBI048_mlops_titanic_solutions.ipynb
│   ├── NDBI048_mlops_titanic.ipynb
│   └── sandbox.ipynb
├── ml_app
│   ├── bin
│   │   └── run-ml-app
│   └── src
│       ├── ml_app
│       │   ├── api.py
│       │   ├── clean.py
│       │   ├── data-prep.py
│       │   ├── model.py
│       │   └── predictions.py
│       └── tests
│           └── test_app.py
├── __init__.py
├── README.md
├── requirements.txt
├── setup.py
├── .gitignore
├── .python-version
└── README.md
```

4

Experimenting

Experiment Tracking

PROFINIT

- › How can we compare multiple models?

Experiment Tracking

- › How can we compare multiple models?

```
In [ ]: model = LogisticRegression()
        model.fit(X, y)

        print(f'model AUC: {roc_auc_score(y, model.predict_proba(X)[:, 1])}')

```

```
In [ ]: model = LogisticRegression(C=999)
        model.fit(X, y)

        print(f'model AUC no reg: {roc_auc_score(y, model.predict_proba(X)[:, 1])}')

```

```
In [ ]: model = RandomForestClassifier(n_estimators=100)
        model.fit(X, y)

        print(f'model RF-100 AUC: {roc_auc_score(y, model.predict_proba(X)[:, 1])}')

```

```
In [ ]: model = RandomForestClassifier(n_estimators=1000)
        model.fit(X, y)

        print(f'model RF-1000 AUC: {roc_auc_score(y, model.predict_proba(X)[:, 1])}')

```

Tracking Values In Excel

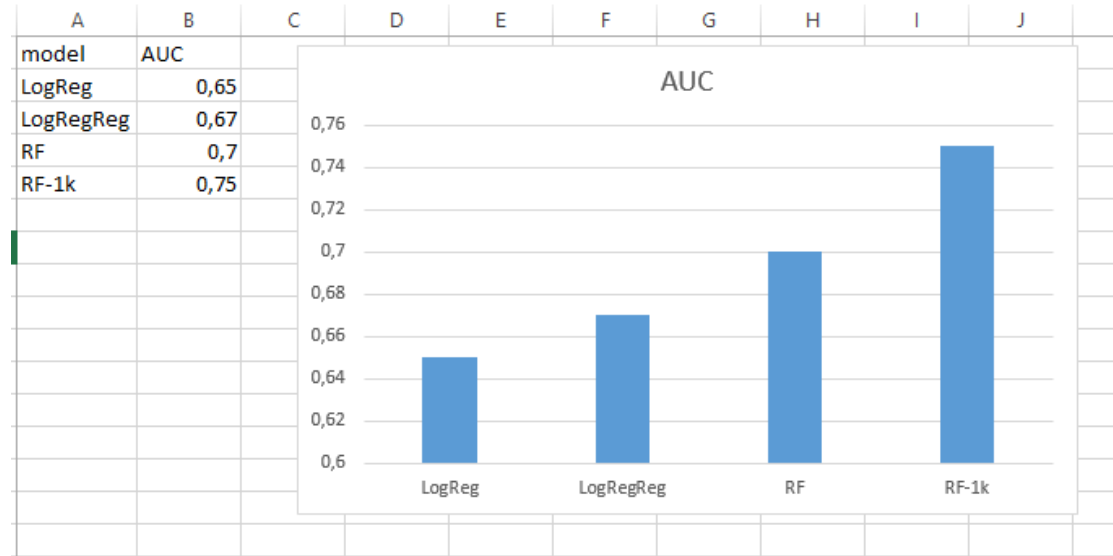
> Pros:

- Understandable
- Collaboration allowed
- Fast to create

> Cons:

- ???

> Alternatives?



MLFlow

- › Open Source platform managing end-to-end ML lifecycle.

```
import mlflow

mlflow.sklearn.autolog()
model = LogisticRegression()

with mlflow.start_run() as run:
    model.fit(X, y)
```


Experiments



Search Experiments

demo-experiment



demo-experiment

Experiment ID: 0

Notes

MLFlow demo on titanic dataset.

git: <https://github.com/profinit/MFF-DS>

Showing 3 matching runs

Refresh Compare Delete Download CSV Sort by All

Columns metrics.rmse < 1 and params.model = "tree" Search Filter Clear

	Start Time	Run Name	User	Source	Version	Models	training_accurac	training_f1_score	training_log_loss	Parameters
<input type="checkbox"/>	1 minute ago	-	sstamenov	C:\Program	-	sklearn	0.783	0.783	0.459	999
<input type="checkbox"/>	3 minutes ago	-	sstamenov	C:\Program	-	sklearn	0.697	0.683	0.6	999
<input type="checkbox"/>	5 minutes ago	-	sstamenov	C:\Program	-	sklearn	0.697	0.683	0.6	1.0

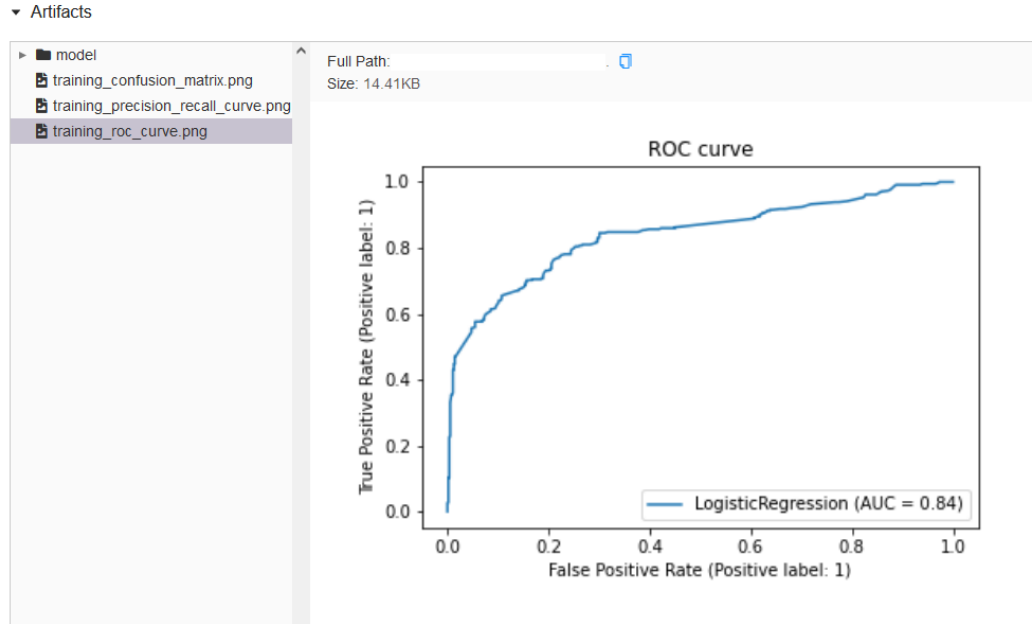
Load more

MLFlow Live Demo

MLFlow artifacts

PROFINIT

- › Additional outputs
- › Saving the model object

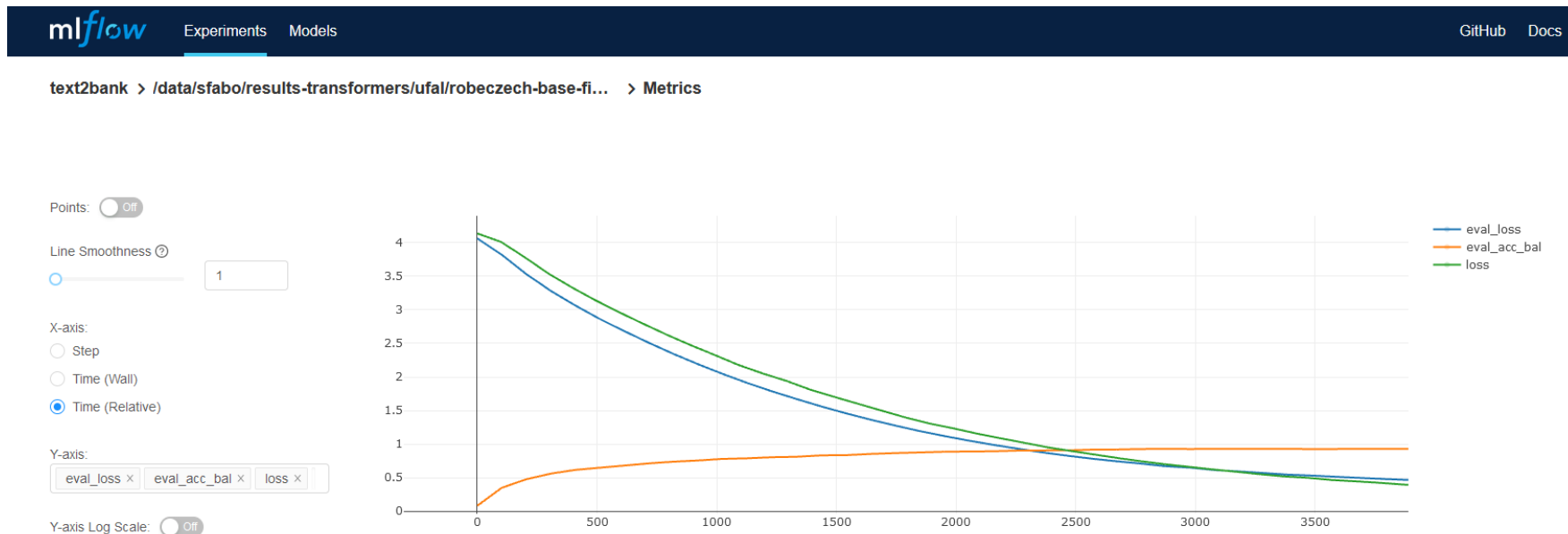


```
mlflow.sklearn.log_model(model_obj, "model")  
mlflow.log_artifact("auc_curve.png")
```

MLFlow – Keras autolog

PROFINIT

> Transformers Training Curves (eval_loss, eval_acc_bal, loss)



MLFlow – Comparing Models

PROFINIT

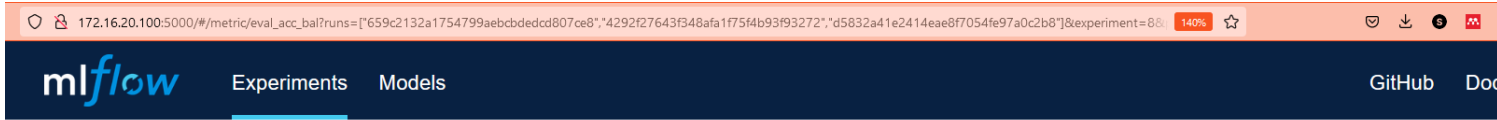
172.16.20.100:5000/#/compare-runs?runs=["659c2132a1754799aebcbdedcd807ce8","4292f27643f348afa1f75f4b93f93272","d5832a41e2414eae8f7054fe97a0c2b8"]

Metrics

epoch 	62.74	60.81	36
eval_acc 	0.924	0.926	0.914
eval_acc_bal 	0.926	0.931	0.912
eval_acc_top_3 	0.97	0.969	0.976

MLFlow – Comparing Models

PROFINIT



text2bank > Comparing 3 Runs > Metrics

Points: Off

Line Smoothness



1

X-axis:

Step

Time (Wall)

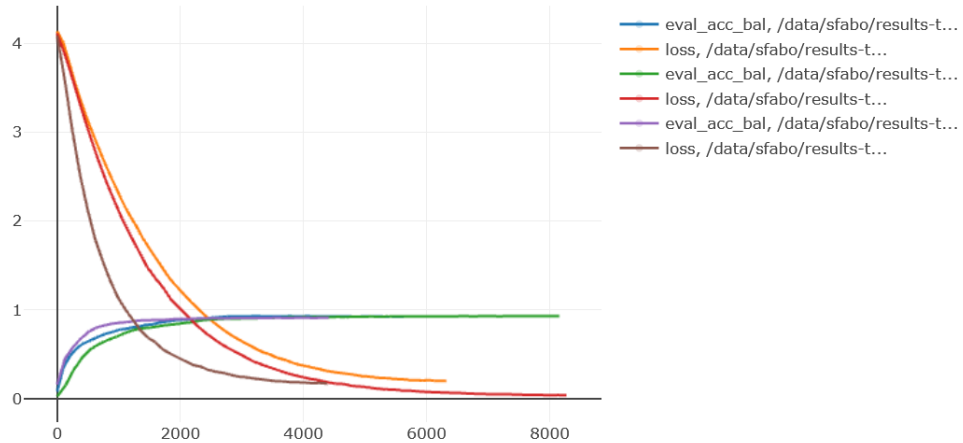
Time (Relative)

Y-axis:

eval_acc_bal ×

loss ×

Y-axis Log Scale: Off



MLFlow – Models Registry

PROFINIT

mlflow

Experiments

Models

GitHub

Docs

Registered Models

Create Model

Search by model name

Search

Filter

Clear

Name	Latest Version	Staging	Production	Last Modified	Tags
titanic	Version 1	-	-	2021-11-30 22:33:21	-

< Page 1 >

10 / page

MLFlow – Usage

- > `mlflow.set_tracking_uri("localhost:5000")`
 - # export MLFLOW_TRACKING_URI=localhost:5000

- > `mlflow.set_experiment("my-experiment")`

- `with mlflow.start_run() as run:`
 - `mlflow.sklearn.log_model(model, "model")`

5

ML Pipelines

ML Pipelines

- › Typical models starts with a single `train.ipynb`

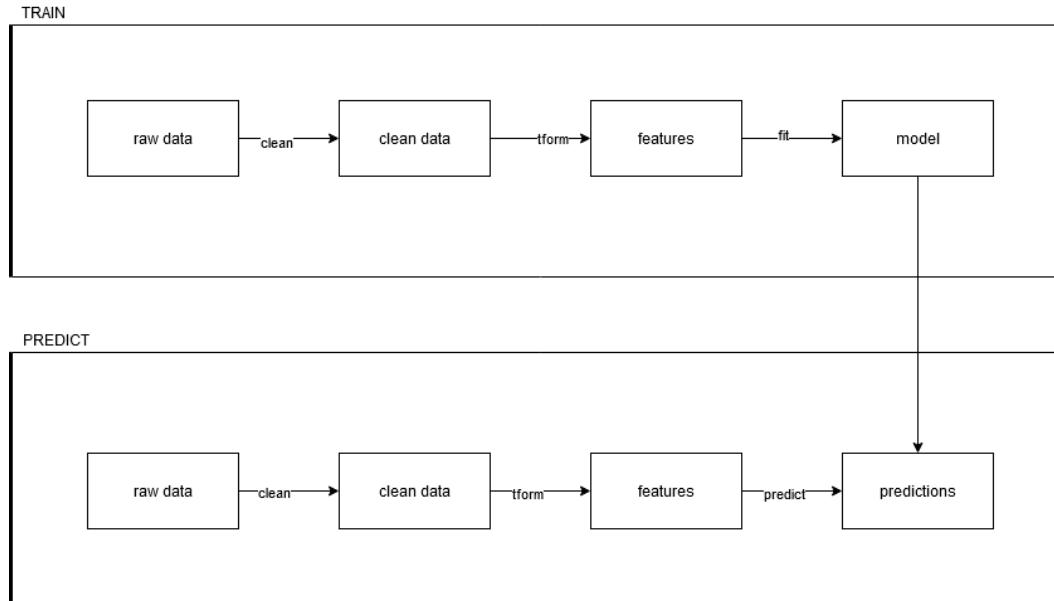
```
# read data pd.read_csv()
# transform data
# create a model

model = LogisticRegression()
model.fit(X, y)

predictions = model.predict(test)[: , 1]
predictions.to_csv('test_predictions.csv')
```

ML Pipelines

- › In practice, we are interested in predictions on new data
 - Beware of possible duplications (Example from projects)



Better ML Pipeline

PROFINIT

```
# clean data  
python clean.py
```

```
# transform data  
python features.py
```

```
# fit model  
python model.py
```

```
# generate predictions  
python predictions.py
```

Better ML Pipeline

PROFINIT

```
# clean_data
```

```
py
```

```
#
```

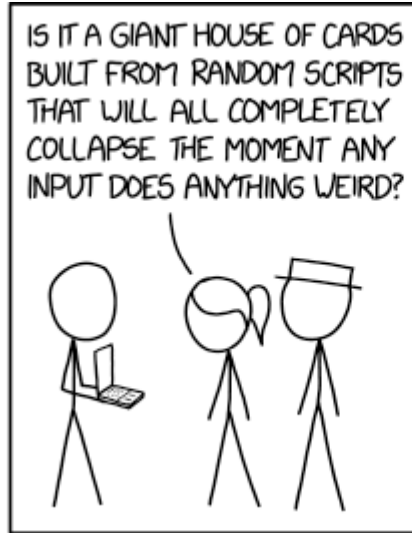
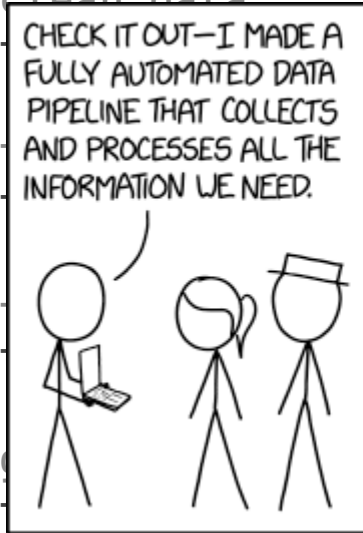
```
py
```

```
#
```

```
py
```

```
#
```

```
py
```





5

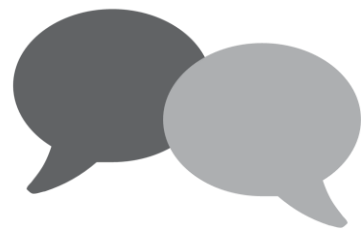
Recapitulation & Conclusion

Recapitulation

› Devops → MLOps



Bad 	Good 
folders / data_v123.csv	git & DVC
Dependency hell	requirements.txt & setup.py
Chaos in notebooks	readable & understandable code
Excel tables	MLFlow



Diskuze

Děkujeme za pozornost

PROFINIT

Profinit EU, s.r.o., Tychonova 2, 160 00 Praha 6
Tel.: + 420 224 316 016, web: www.profinit.eu



LinkedIn
linkedin.com/company/profinit



Twitter
twitter.com/Profinit_EU



Facebook
facebook.com/Profinit.EU



Youtube
Profinit EU