



Машинное обучение и  
высоконагруженные системы

Москва, осень 2023

# MLOps. Начало

Гаврилова Елизавета,  
Senior ML-engineer





## Неделя 6. Что будем обсуждать сегодня?

- MLFlow ликбез
- AirFlow: вспомним как обучить 3 модели










## MLFlow ликбез

[mlflow.log\\_model](#) – Сохранение артефактов в удаленное хранилище

[mlflow.save\\_model](#) – Сохранение артефактов в локальную файловую систему

Результат выполнения:

<input type="checkbox"/>	Имя	Размер
<input type="checkbox"/>	 MLmodel	705 Б
<input type="checkbox"/>	 conda.yaml	249 Б
<input type="checkbox"/>	 model.pkl	138.09 МБ
<input type="checkbox"/>	 python_env.yaml	114 Б
<input type="checkbox"/>	 requirements.txt	124 Б



[mlflow.models.infer\\_signature](#) – добавление описания данных в файл MLmodel

Результат выполнения:

▼ Artifacts

▼ hgb

MLmodel

conda.yaml

model.pkl

python\_env.yaml

requirements.txt

Full Path:s3://lizvladi-mlflow-artifacts/lizvladii/052bdf49492c4a79be85577f5c2385bf/artifacts/hgb/MLmodel

Size: 700B

```
artifact_path: hgb
flavors:
  python_function:
    env:
      conda: conda.yaml
      virtualenv: python_env.yaml
    loader_module: mlflow.sklearn
    model_path: model.pkl
    predict_fn: predict
    python_version: 3.11.5
  sklearn:
    code: null
    pickled_model: model.pkl
    serialization_format: cloudpickle
    sklearn_version: 1.3.0
mlflow_version: 2.7.1
model_uuid: 9b1ae890af23453685b0474802a1bb85
run_id: 052bdf49492c4a79be85577f5c2385bf
signature:
  inputs: '[{"type": "tensor", "tensor-spec": {"dtype": "float64", "shape": [-1, 8]}}]'
  outputs: '[{"type": "tensor", "tensor-spec": {"dtype": "float64", "shape": [-1]}}]'
  params: null
utc_time_created: '2023-10-13 09:54:12.461158'
```



## MLFlow ликбез

[mlflow.log\\_metric](#) – сохранение метрики модели

Результат выполнения:

› Description [Edit](#)

› Datasets

› Parameters (2)

▼ Metrics (3)

Name	Value
mae <a href="#">📈</a>	0.24
r2_score <a href="#">📈</a>	0.775
rmse <a href="#">📈</a>	0.543



## MLFlow ликбез

[mlflow.log\\_params](#) – сохранение параметров модели

Результат выполнения:

> Description [Edit](#)

> Datasets

▼ Parameters (2)

Name	Value
max_depth	10
n_estimators	100



## MLFlow ликбез

[\*\*mlflow.evaluate\*\*](#) — оценка модели одним из готовых [evaluators](#)

Результат выполнения:

▼ Metrics (10)	
Name	Value
example_count 	4128
max_error 	3.322
mean_absolute_error 	0.308
mean_absolute_percentage_error 	0.18
mean_on_target 	2.055
mean_squared_error 	0.214
r2_score 	0.836
root_mean_squared_error 	0.463
score 	0.836
sum_on_target 	8483.1



## MLFlow NestedRun

Позволяет создавать несколько **child run** в рамках одного **parent run**.

Когда может быть полезно?

- при обучении различных моделей в цикле
- при тестировании гиперпараметров в цикле.

Результаты выполнения:

Table	Chart	Evaluation	Experimental
🔍	Run Name		
🔍	📄	●	PARENT_RUN
🔍		●	hgb
🔍		●	rf
🔍		●	lr

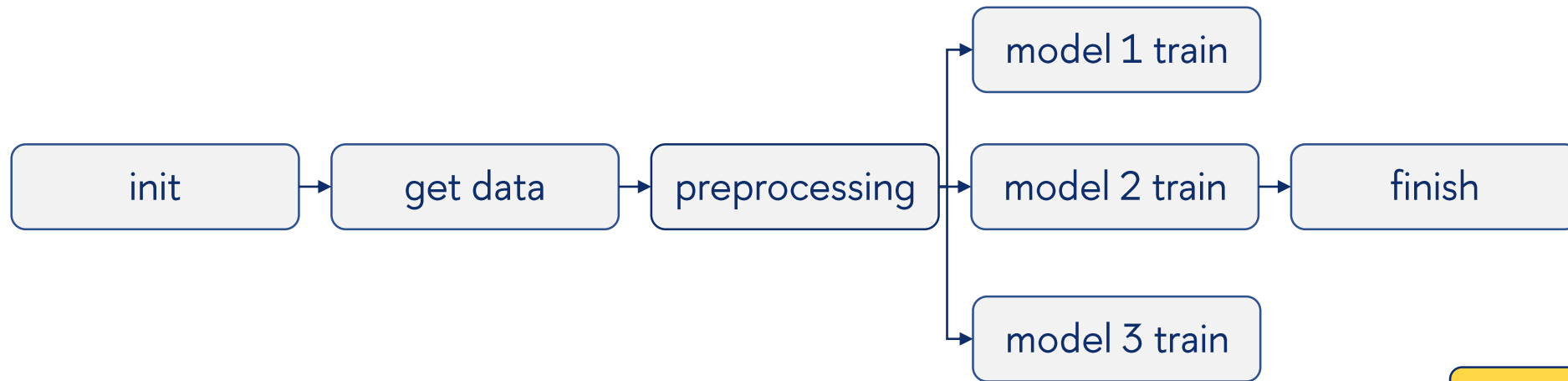
Table	Chart	Evaluation	Experimental
🔍	Run Name		
🔍	📄	●	HistGB
🔍		●	learning_rate_0.5_max_iter..
🔍		●	learning_rate_0.5_max_iter..
🔍		●	learning_rate_0.05_max_it...





## Повторяем AirFlow

Вариант 1



Вариант 2



## Разобрали сегодня

- Как устроен NestedRun
- Как обучить 3 модели в одном DAG

## ДЗ

- Попробовать реализовать NestedRun
- Посмотреть финальное задание и подготовить вопросы к вебинару