

Домашнее задание: MLflow & AirFlow

Задание

- **Тест по MLflow и AirFlow** (2 балла)
- **Интеграция модели в MLflow:** Добавление модели из первого ДЗ в MLflow и настройка автоматического вывода лучшей модели в production (4 балла)
- **Пайплайн AirFlow:** Построение и запуск DAG для автоматизации пайплайна (4 балла)
- **Отчёт:** Документирование всех этапов работы

1 Полезные команды и советы

1.1 MLflow

```
# start UI MLflow
mlflow ui --port 5000

# model registration in Model Registry using Bash
mlflow models register -m "runs:/<RUN_ID>/model" --name "my_model"
```

1.2 AirFlow

```
# airflow server start
airflow webserver --port 8080

# airflow db init
airflow db init

# airflow user create
airflow users create --username admin --password admin --firstname
Admin --lastname Admin --role Admin --email admin@example.com

# scheduler start
airflow scheduler

# DAG testing
airflow tasks test <DAG_ID> <TASK_ID> <DATE>
```

1.3 Kubernetes

```
#  
kubectl get pods -n airflow  
  
#  
kubectl logs <POD_NAME> -n airflow
```

1.4 Советы

- Используйте `mlflow.autolog()` для автоматического логирования параметров sklearn/pytorch.
- Для AirFlow всегда указывайте `dag=dag` в операторах.
- В Kubernetes проверяйте лимиты ресурсов для подов.
- После инициализации AirFlow (`airflow db init`) обязательно создайте пользователя (`airflow users create`) и запустите планировщик (`airflow scheduler`).

Требования к отчёту

- **Model Registry:** Скриншоты из MLflow UI с зарегистрированными версиями моделей.
- **Критерии вывода модели в production:** Пример кода для сравнения моделей:

```
if new_metric > current_best_metric:  
    mlflow.register_model("runs:/<RUN_ID>/model", "prod_model")
```

- **AirFlow DAG:** Скриншот из AirFlow UI с выполненными задачами (зеленые статусы).
- **Выводы:** Описание проблем и решений при настройке пайплайнов.

Форма сдачи

- **Репозиторий:** Укажите ссылку на Git-репозиторий с кодом и отчётом в ГУГЛ ФОРМЕ.
- **Форма сдачи:** Обязательно отправьте репозиторий через Google Форму с тестом. Репозитории, отправленные через Telegram, проверяться будут значительно дольше.
- По всем вопросам касательно MLFlow, AirFlow писать в тг @artemovma