

Отчет по лабораторной работе №3

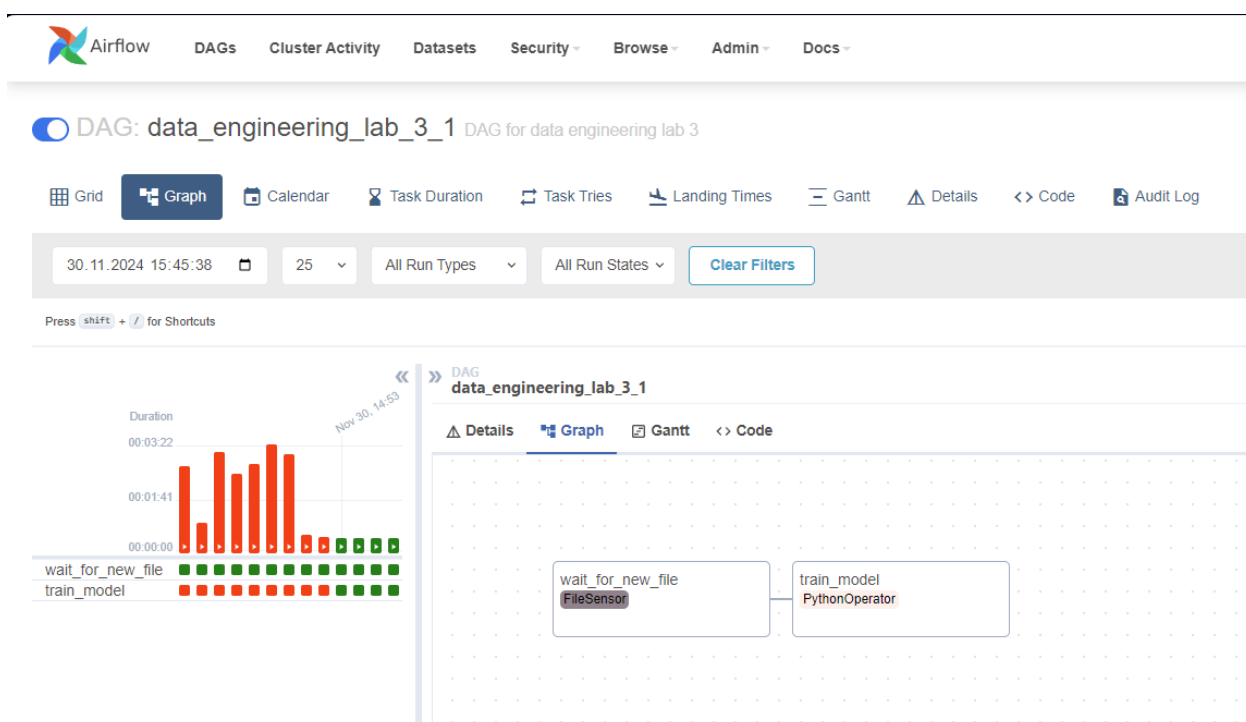
Выполнила Агафонова Елизавета, группа 6233

1 часть

Был создан пайплайн для обучения нескольких моделей классификации на основании конфигурационного файла.

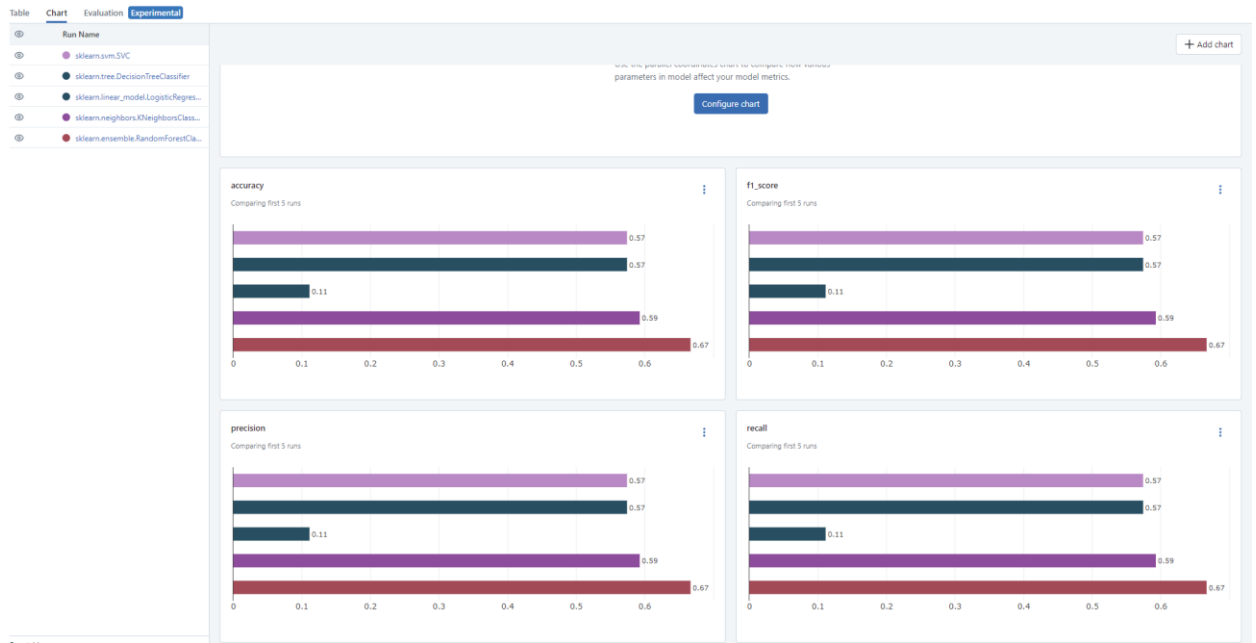
DAG состоит из:

- wait_for_new_file – ожидание файла. Для этого было создано connection с airflow (из 2 лр).
- train_model – обучение моделей
 1. считывание датасета
 2. кодирование категориальных данных
 3. разделение данных на выборки
 4. сохранение валидационных данных
 5. чтение конфигурационного файла
 6. настройка Mlflow для проведения эксперимента
 7. обучение моделей согласно конфигурационному файлу
 8. логирование результатов обучения



В начале возникала ошибка Failed to import Git, которая была решена при помощи добавления `os.environ["GIT_PYTHON_REFRESH"] = "quiet"`.

Получились следующие диаграммы

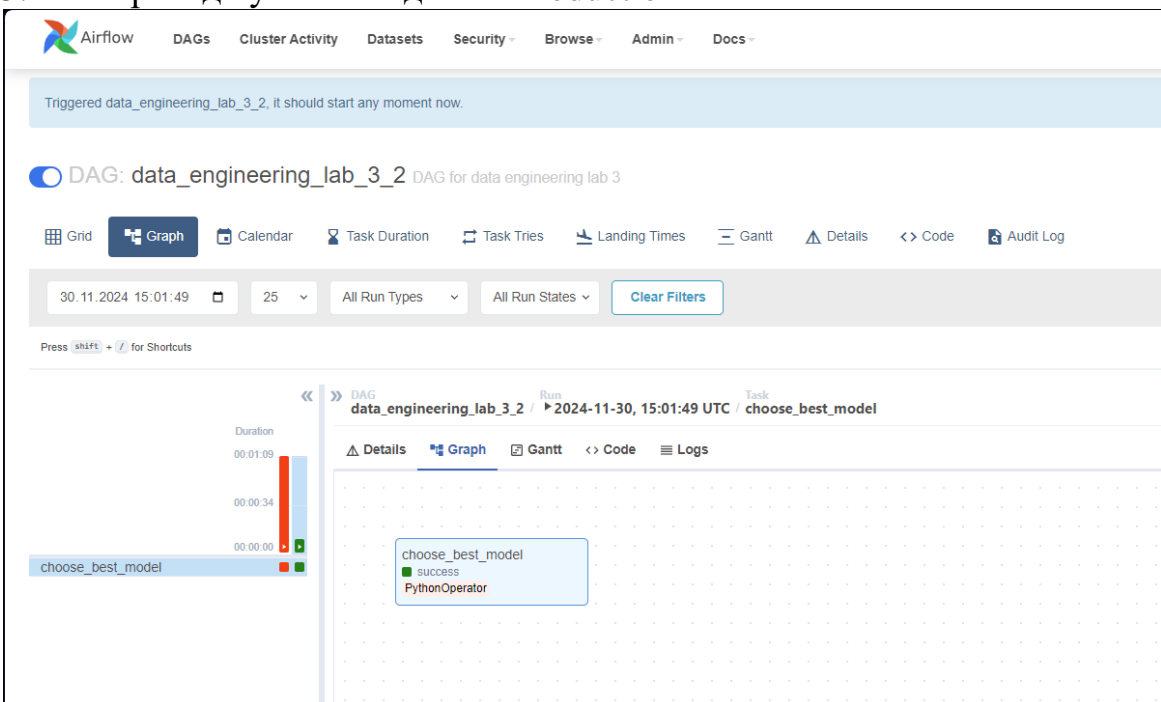


2 часть

Был создан пайплайн для валидации обученных моделей и перевод лучшей в Production.

DAG состоит из:

- validate – валидация обученных моделей
1. загрузка валидационных данных
 2. чтение конфигурационного файла
 3. валидация каждой модели
 4. выбор лучшей модели
 5. перевод лучшей модели в Production



Registered Models

Create Model

Filter registered models by name or tags

Name [?]	Latest version	Staging	Production	Created by	Last modified	Tags
sklearn.ensemble.RandomForestClassifier	Version 4	---	Version 4		2024-11-30 19:01:51	---
sklearn.linear_model.LogisticRegression	Version 4	---	---		2024-11-30 18:58:53	---
sklearn.neighbors.kneighborsClassifier	Version 4	---	---		2024-11-30 18:58:51	---
sklearn.svm.SVC	Version 4	---	---		2024-11-30 18:58:57	---
sklearn.tree.DecisionTreeClassifier	Version 4	---	---		2024-11-30 18:58:55	---