Андреев Александр 6233

ЛР сделана вместе с Корноушкиным Ильей 6231

Первый pipeline

Был написан скрипт для первого pipeline.

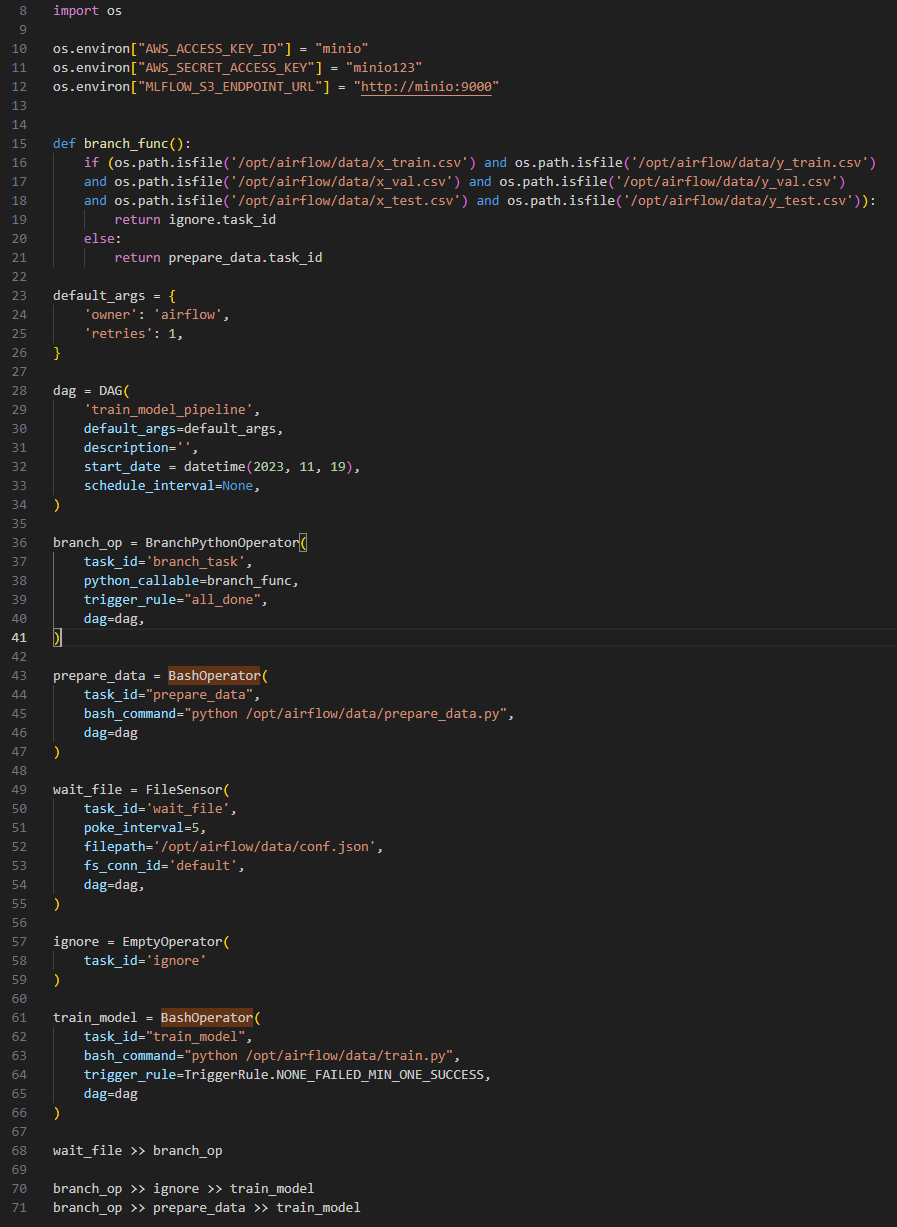


Рисунок 1 – Код первого dag файла

Также было выполнено дополнительное задание: чтобы скрипт prepare\_data выполнялся только один раз при запуске dag. Но на мой взгляд это не совсем логичное задание и нужно его переформулировать: чтобы скрипт prepare\_data выполнялся в том случае, если в рабочей директории отсутствуют один, несколько или все файлы необходимые для обучения и тестирования модели. Сделано это было с помощью BranchPythonOperator.

В один момент была идея «сделать интересно» с использованием bash команд, где одна таска командой bash определяла наличие файлов, а другой таска с помощью xcom\_pull перехватывала бы то сообщение и уже без функций из os python возвращала имя нужной таски.

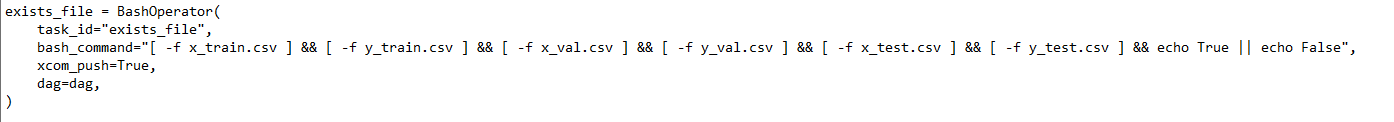


Рисунок 2 – Код bash запроса

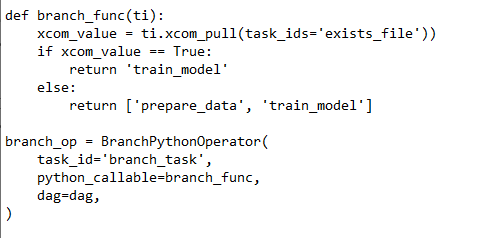


Рисунок 3 – Код метода branch\_func

Опыт был довольно интересный в bash среде и код работал, но был выбран в итоге более простой вариант.

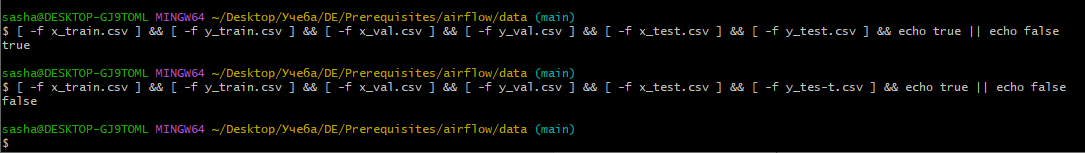


Рисунок 4 – Запросы в bash среде

По итогу работы были получены модели в mlflow.

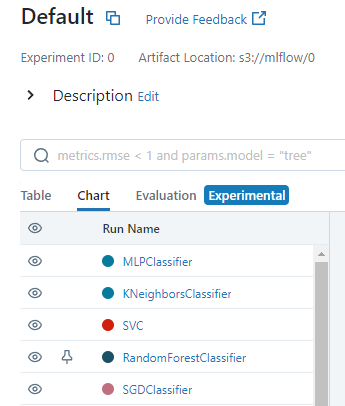


Рисунок 5 – Запущенные модели

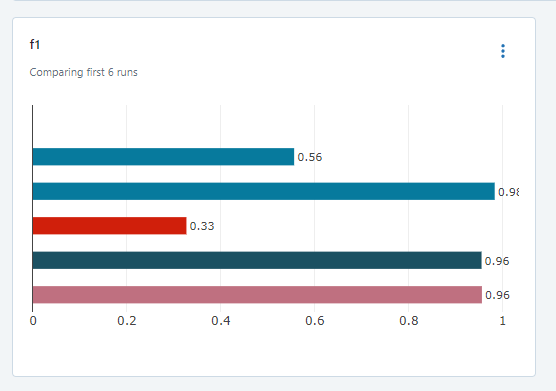


Рисунок 6 – Результаты обучения моделей по метрике f1

Второй pipeline

Аналогично был написан второй dag файл, в котором определялась лучшая модель и выводилась в Production.

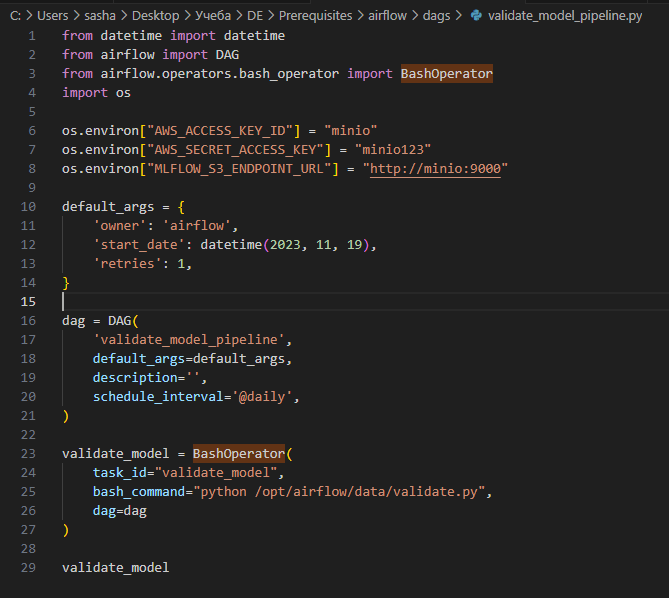


Рисунок 7 – Код второго dag файла

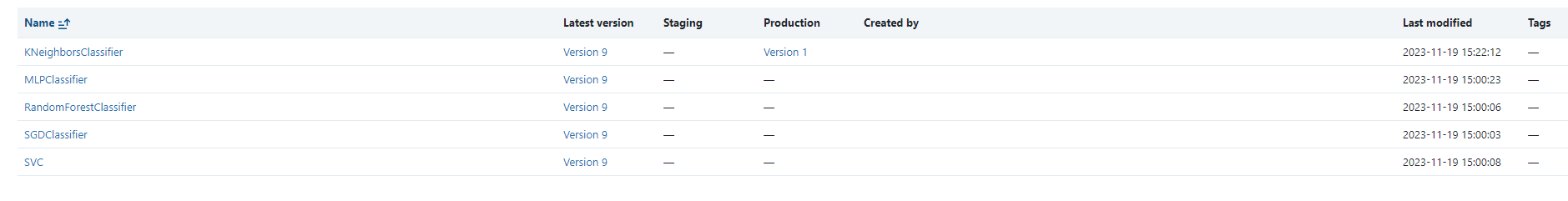


Рисунок 8 – Запущенная модель