INF-0077 - T1 - Experiment tracking & Pipelines

Grupo:
Fabio Grassiotto
Wandemberg Santana Pharaoh Gibaut
Guilherme Ramirez

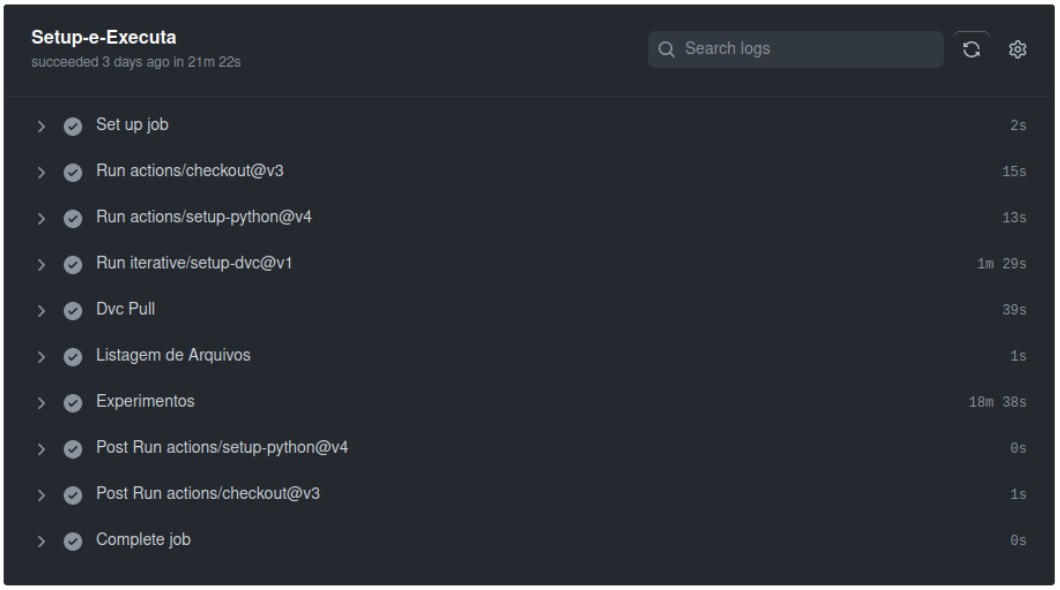
1. **Link para o Github**
   O projeto se encontra hospedado no link https://github.com/fabiograssiotto/INF-0077
2. **Decisões tomadas**
   Para este projeto utilizamos o Github Actions para a execução da pipeline devido à sua ampla adoção e quantidade de funcionalidades disponibilizadas. Para storage remoto o Google Drive foi utilizado devido à facilidade de uso.
3. **Comandos executados**
   O seguinte arquivo yml foi utilizado para configuração da pipeline no Github Actions:

```yaml
# Workflow para o treinamento do modelo.

name: Treinamento e Report
on:
 push:
 workflow_dispatch:
jobs:
 Setup-e-Executa:
   runs-on: windows-latest
   steps:
     - uses: actions/checkout@v3
     - uses: actions/setup-python@v4
       with:
         python-version: "3.9"
         cache: "pip"
     - uses: iterative/setup-dvc@v1
     - name: Dvc Pull
       env:
         GDRIVE_CREDENTIALS_DATA: "${{ secrets.GDRIVE_SECRET }}"
       run: |
         dvc pull --verbose
     - name: Listagem de Arquivos
       run: ls T1/data
     - name: Experimentos
       env:
         GDRIVE_CREDENTIALS_DATA: "${{ secrets.GDRIVE_SECRET }}"
       run: |
         python -m pip install --upgrade pip
         pip install -r requirements_win.txt --quiet
         dvc repro -f # Force without any changes
```

4. **Screenshots**

Abaixo encontra-se o resultado da execução da pipeline utilizando o Github Actions:



Já o registro do experimento resultante da pipeline é mostrado a seguir:

```
[venv:mlops] (git:main) $ python Untitled-1.py
2023-11-28 14:47:54.745855: I tensorflow/core/util/port.cc:113] oneDNN
custom operations are on. You may see slightly different numerical
results due to floating-point round-off errors from different
computation orders. To turn them off, set the environment variable
`TF_ENABLE_ONEDNN_OPTS=0`.
WARNING:tensorflow:From
C:\Users\fabio.grassiotto\Research\venv\mlops\Lib\site-packages\keras\sr
c\losses.py:2976: The name tf.losses.sparse_softmax_cross_entropy is
deprecated. Please use tf.compat.v1.losses.sparse_softmax_cross_entropy
instead.

WARNING:tensorflow:From
C:\Users\fabio.grassiotto\Research\venv\mlops\Lib\site-packages\keras\sr
c\backend.py:873: The name tf.get_default_graph is deprecated. Please
use tf.compat.v1.get_default_graph instead.

WARNING:tensorflow:From
C:\Users\fabio.grassiotto\Research\venv\mlops\Lib\site-packages\keras\sr
c\layers\pooling\max_pooling2d.py:161: The name tf.nn.max_pool is
deprecated. Please use tf.nn.max_pool2d instead.

2023-11-28 14:48:01.996558: I
tensorflow/core/platform/cpu_feature_guard.cc:182] This TensorFlow
binary is optimized to use available CPU instructions in
performance-critical
```

```
operations.
To enable the following instructions: SSE SSE2 SSE3 SSE4.1 SSE4.2 AVX2
FMA, in other operations, rebuild TensorFlow with the appropriate
compiler flags.
Found 8000 images belonging to 2 classes.
Found 2000 images belonging to 2 classes.
Epoch 1/10
WARNING:tensorflow:From
C:\Users\fabio.grassiotto\Research\venv\mlops\Lib\site-packages\keras\sr
c\utils\tf_utils.py:492: The name tf.ragged.RaggedTensorValue is
deprecated. Please use tf.compat.v1.ragged.RaggedTensorValue instead.

WARNING:tensorflow:From
C:\Users\fabio.grassiotto\Research\venv\mlops\Lib\site-packages\keras\sr
c\engine\base_layer_utils.py:384: The name
tf.executing_eagerly_outside_functions is deprecated. Please use
tf.compat.v1.executing_eagerly_outside_functions instead.

125/125 [==============================] - 229s 2s/step - loss: 0.7571 -
accuracy: 0.5063 - val_loss: 0.6925 - val_accuracy: 0.5060
Epoch 2/10
125/125 [==============================] - 156s 1s/step - loss: 0.6901 -
accuracy: 0.5336 - val_loss: 0.6792 - val_accuracy: 0.5645
Epoch 3/10
125/125 [==============================] - 211s 2s/step - loss: 0.6788 -
accuracy: 0.5723 - val_loss: 0.6595 - val_accuracy: 0.6326
Epoch 4/10
125/125 [==============================] - 210s 2s/step - loss: 0.6657 -
accuracy: 0.6020 - val_loss: 0.6588 - val_accuracy: 0.6084
Epoch 5/10
125/125 [==============================] - 154s 1s/step - loss: 0.6688 -
accuracy: 0.5971 - val_loss: 0.6429 - val_accuracy: 0.6557
Epoch 6/10
125/125 [==============================] - 157s 1s/step - loss: 0.6544 -
accuracy: 0.6211 - val_loss: 0.6421 - val_accuracy: 0.6411
Epoch 7/10
125/125 [==============================] - 215s 2s/step - loss: 0.6433 -
accuracy: 0.6332 - val_loss: 0.6190 - val_accuracy: 0.6623
Epoch 8/10
125/125 [==============================] - 245s 2s/step - loss: 0.6328 -
accuracy: 0.6491 - val_loss: 0.6053 - val_accuracy: 0.6764
Epoch 9/10
125/125 [==============================] - 306s 2s/step - loss: 0.6299 -
accuracy: 0.6442 - val_loss: 0.6308 - val_accuracy: 0.6361
Epoch 10/10
125/125 [==============================] - 259s 2s/step - loss: 0.6240 -
```

```
accuracy: 0.6495 - val_loss: 0.6175 - val_accuracy: 0.6497
[venv:mlops] (git:main) $
```

5. **Conclusão**
   Este projeto permitiu a exploração das capacidades do Github Actions para criar fluxos de execução de experimentos. Conseguimos automatizar o registro de dados e o armazenamento de informações cruciais, como métricas de perda e precisão dos experimentos. Essas ferramentas oferecem a possibilidade de equipes poderem implementar fluxos de trabalho de maneira eficiente e reprodutível, facilitando a colaboração entre os membros, reduzindo erros e agilizando o desenvolvimento, teste e deploy de modelos de machine learning.