MLOps – home assignment

Felix Platteau

Contents

[1 Opzetten 3](#_Toc94624517)

[1.1 Sourcecode 3](#_Toc94624518)

[1.2 Docker 3](#_Toc94624519)

[1.3 Database 3](#_Toc94624520)

[1.3.1 Influxdb 3](#_Toc94624521)

[1.3.2 Cosmosdb 3](#_Toc94624522)

[2 Installeren 4](#_Toc94624523)

[2.1 Backend 4](#_Toc94624524)

[2.1.1 Configuratie 4](#_Toc94624525)

[2.1.2 Build image 4](#_Toc94624526)

[2.1.3 Run container 4](#_Toc94624527)

[2.2 Frontend 4](#_Toc94624528)

[2.2.1 Configuratie 4](#_Toc94624529)

[2.2.2 Build image 4](#_Toc94624530)

[2.2.3 Run container 4](#_Toc94624531)

[3 Future-proof 5](#_Toc94624532)

# My project

## Fictional company

My fictional company is a rice factory that could use an application for r6ecognizing the different rice types.

## Dataset

I chose the [rice image dataset](https://www.kaggle.com/datasets/muratkokludataset/rice-image-dataset).

“Rice, which is among the most widely produced grain products worldwide, has many genetic varieties. These varieties are separated from each other due to some of their features. These are usually features such as texture, shape, and color. With these features that distinguish rice varieties, it is possible to classify and evaluate the quality of seeds.”

## AI model

I will create my own CNN model to classify the rice varieties.

# Prepare and train

## Azure

For my data preparing and training I will use Azure Machine Learning.

## Steps

1. Create resource group in Azure portal
2. Create Azure machine learning workspace in your just created resource group
   1. Create new container registry for this service -> **basic sku**

Graphical user interface, text, application, email

Description automatically generated

1. Access your newly created machine learning workspace

Graphical user interface

Description automatically generated with medium confidence

1. Create a new compute instance (choose cheapest option)

Graphical user interface, text, application, email

Description automatically generated

1. Upload dataset images (I started with 1000 of the 15000 per rice type, can add more in the future if needed)

First I tried with 5000 images per rice type but 32GB memory wasn’t enough to register the training and testing dataset

* 1. Data -> Data assets -> create
  2. Give it the name of 1 rice type
  3. Type = data asset types -> folder
  4. Choose from local storage
  5. Select blob storage
  6. Upload images
  7. Repeat for every rice type

Graphical user interface, text, application, email

Description automatically generated

1. Make dataset from uploaded images
   1. Data -> Data assets -> create
   2. Name it rice type + “\_v1”
   3. Type = **DATASET** types -> file
   4. Choose from Azure storage
   5. Select blob storage
   6. Choose correct folder where images are located (UI/*$timestamp/$rice\_type)*

Graphical user interface, application, email

Description automatically generated

1. Upload or make notebooks
   1. When running select “Python 3.8 - Pytorch and Tensorflow” as kernel
   2. Prepare data
   3. Train CNN

Table

Description automatically generated

1. When everything is done
   1. Download best model
   2. Download edited notebooks and files
   3. Delete cluster group

# Deployment

## Azure

1. Make a new resource group
2. Execute following command in the Azure CLI (change subID and resourceGroupName)

az ad sp create-for-rbac -n "MLOps-RBAC-FP" --role Contributor --scopes /subscriptions/9ae53766-1df1-4559-8ad6-262ddacf26b7/resourceGroups/mlops-project

1. This will give you the following output: (paste to save)

*{*

*"appId": "8a679773-f887-4b53-8b4f-dd3450d60fde",*

*"displayName": "MLOps-RBAC-FP",*

*"password": "**lW.8Q~vYvxjqw1lrSZ1zgDGhKKkoSGN5oijJ4bIs",*

*"tenant": "4ded4bb1-6bff-42b3-aed7-6a36a503bf7a"*

*}*

1. To go on we just need a few other details:
   1. Subscription ID: Navigate to your Subscription List, and copy the ID of the right Subscription **9ae53766-1df1-4559-8ad6-262ddacf26b7**
   2. Tenant ID: Click on your Subscription and find the Parent Management Group **4ded4bb1-6bff-42b3-aed7-6a36a503bf7a**
   3. Service Principal name: The name you entered earlier: **MLOps-RBAC-FP**
   4. Service Principal App ID: The app ID from the result of the CLI / The Azure Portal **8a679773-f887-4b53-8b4f-dd3450d60fde**
   5. Service Principal Secret: The password from the result of the CLI / The Azure Portal **lW.8Q~vYvxjqw1lrSZ1zgDGhKKkoSGN5oijJ4bIs**

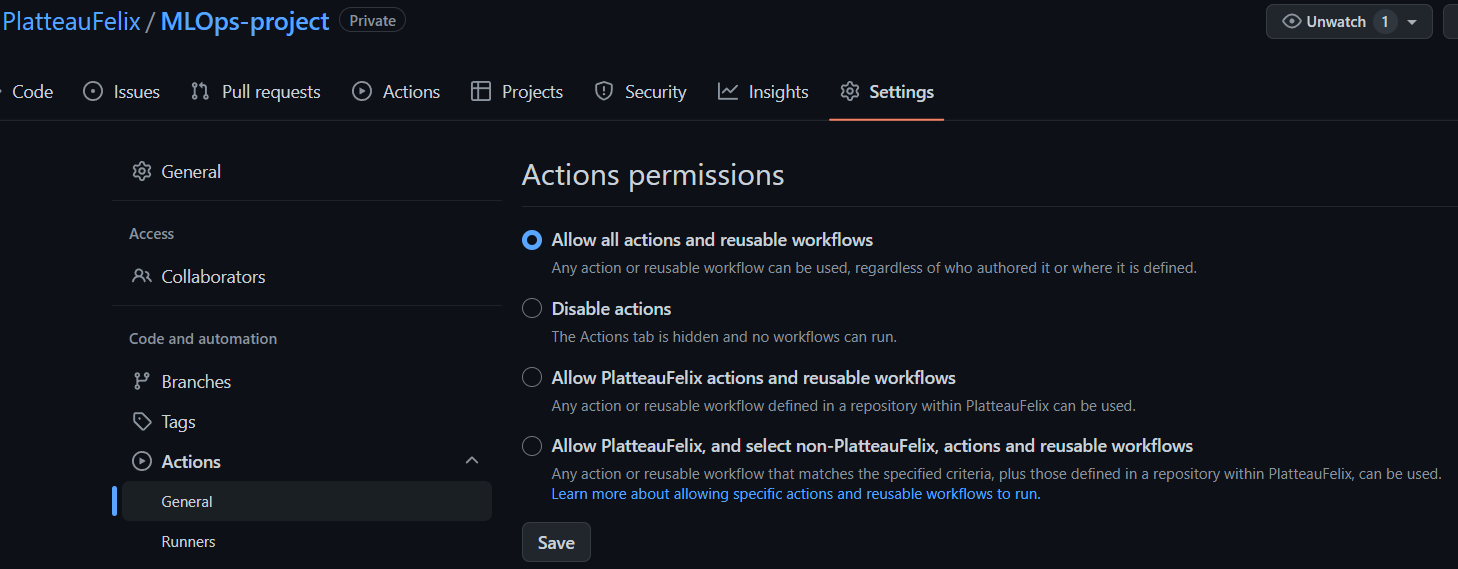
## Github

1. Clone repository: <https://github.com/nathansegers/05_AzureMLOps-2022>
2. Add principal secret into Github secret actions as “CLIENT\_SECRET”

A screenshot of a computer

Description automatically generated with medium confidence

1. Make new .env file from the example file and fill in all values
2. Change values in .github/workflows/azure-ai.yml
3. Push/publish to own repo
4. Be sure to allow access to the actions



# Future-proof