Azure Machine Learning Service

*This document outlines the steps required to set up the python development to run Azure Machine Learning Service Experiments on your local machine. Theoretically, these steps are unneeded if the experiments are targeted on an Azure Virtual Machine, though this has not been tested. To avoid incurring unnecessary costs, it is suggested that these experiments are first run in a local environment to ensure that the experiment is set up properly (in case the developer wishes to expand on this already-working experiment). This tutorial uses the Anaconda distribution, and certain required packages (i.e. pandas) that are included in Anaconda are not included in the basic distribution.*

*The files in the MachineLearning folder define an AutoML experiment. An AutoML experiment differs from a regular ML experiment in that the AutoML experiment runs and trains multiple models automatically, whereas the regular ML experiment runs and trains only one model. Using the AutoML approach, we are able to automate the trial and error process in discovering which model can be trained to provide the greatest predictive accuracy.*

*Furthermore, the models defined in the MachineLearning folder train use both climatic data and elevation data as Y.*

Prerequisites:

1. Visual Studio Code is installed on your machine, along with the Azure Machine Learning Extension
2. You have a subscription to Azure, and the “Microsoft.MachineLearning” and “Microsoft.MachineLearningService” resource providers are registered.
3. The “MachineLearning” folder is installed on your local machine.
4. The .csv file containing the data must have the relevant columns labeled exactly as follows:
5. Column1, for the gridbox id.
6. NormalizedYield, for normalized yield.
7. MEAN\_Yld\_V , for mean yield volume.
8. Crop-Type , for crop type.
9. V.A.T(F) , for average temperature during vegetative stage.
10. R.A.T(F) , for average temperature during vegetative stage.
11. M.A.T(F) , for average temperature during maturity stage.
12. V.PET(inch) , for potential evapotranspiration during vegetative stage.
13. R.PET(inch) , for potential evapotranspiration during reproductive stage.
14. M.PET(inch) , for potential evapotranspiration during maturity stage.
15. V.T.R(inch), for total rainfall during vegetative stage.
16. R.T.R(inch), for total rainfall during reproductive stage.
17. M.T.R(inch), for total rainfall during maturity stage.

A screenshot of a cell phone

Description automatically generated

1. Files depicting the scenario for which you want to predict yield need to be mocked up in .csv format. These files must contain a unique row for every single grid square of the field. The mock climate data will be the same for every row. Only the “Column1” field will have unique values, and possibly Mean\_eleva if you wish to train on this.

Step 1: Install Anaconda 2019.03 from <https://www.anaconda.com/distribution/#download-section>

Step 2: After the installation process, search for “Anaconda Prompt”, and click on the result.

A screenshot of a cell phone

Description automatically generated

Step 3: Next, we will create the environment. Run the following prompt in the terminal:

“conda create -n AzureExperimentEnvironment python=3.7 anaconda”

Step 4: Once the environment is created, we must now install two Azure SDKs. Run the following prompt to enter AzureExperimentEnvironment:

conda activate AzureExperimentEnvironment

After entering this environment, we will install the first package, using the following prompt:

pip install –upgrade azureml-sdk[automl]

Once this install is completed, run the following prompt:

pip install –upgrade azureml-dataprep

Step 4: Within Azure Portal, go to the “Machine Learning service workspaces” menu. Click the “Add” button at the top of the page, and create a new workspace.

A screenshot of a social media post

Description automatically generated

Step 5: Click on the name of your new workspace to view its settings. At the top of the settings page, click on the “Download config.json” button. Save this file within your copy of the “MachineLearning” folder. This congif.json file is used to connect to your Azure workspace.

A screenshot of a cell phone

Description automatically generated

Step 6: Open the “MachineLearning” folder in Visual Studio Code. In this folder, there are three python files named “train1.py”, “dataPrep.py”, and “main.py”. “main.py” contains the main code to run your AutoML experiment. “dataPrep.py” contains code to prepare your datasets for training. “train1.py” contains code to assist in configuring the experiment.

Locate line 10 of the “main.py” document. You must define the “dataset\_root” to be the absolute path of the dataset .csv file you are training with. Note that this file must be formatted according to the conditions laid out in the prerequisites.

A screenshot of a computer screen

Description automatically generated

Step 7: Locate line 12. Here, an array of path names is defined for predictive scenarios (Y-values), formatted as a .csv file. Add a dictionary entry for every scenario you like to predict on, with any key value you like (preferably something descriptive, since the predicted value files will have these names.)

A black sign with white text

Description automatically generated

Step 8: Locate line 16. Here, define “predicted\_values\_folder” as the path of the folder you would like the predictive results to be written to.

A screenshot of a cell phone

Description automatically generated

Step 9: Return to the Anaconda Prompt. In the terminal (and still in the AzureExperimentEnvironment), run the following prompt:

python $pathname

where $pathname is the absolute path to the “main.py” file.

The experiment will now run, displaying the results in the terminal. After the experiment finished running, it will grab the most accurate model (according to the primary metric), and use it to predict the yield for every file defined in “to\_predict\_array”. The predicted yield is then written to the “predicted\_values\_folder”.