Register Model on Azure ML Workspace

A screenshot of a computer

Description automatically generated

Upload the model file click create

A screenshot of a computer

Description automatically generated

Creating a custom runtime environment

Go to environment tab and click on create new, give environment name in select environment source select create new docker context.

A blue line on a white background

Description automatically generated

Click next and upload the following code

FROM mcr.microsoft.com/azureml/openmpi3.1.2-ubuntu18.04

ENV AZUREML\_CONDA\_ENVIRONMENT\_PATH /azureml-envs/tensorflow-2.14

# Create conda environment

RUN conda create -p $AZUREML\_CONDA\_ENVIRONMENT\_PATH \

    python=3.10 pip=23.1.2

# Prepend path to AzureML conda environment

ENV PATH $AZUREML\_CONDA\_ENVIRONMENT\_PATH/bin:$PATH

# Install pip dependencies

RUN pip install 'psutil~=5.8.0' \

                                          'Pillow' \

                                          'pandas==1.5.3' \

                                          'scipy==1.11.4' \

                                          'numpy==1.23.5' \

                                          'ipykernel~=6.0' \

                                          'azureml-core~=1.54.0' \

                                          'azureml-defaults~=1.54.0' \

                                          'azureml-mlflow==1.54.0' \

                                          'azureml-telemetry==1.54.0' \

                                          'tensorflow==2.14.0'

# This is needed for mpi to locate libpython

ENV LD\_LIBRARY\_PATH $AZUREML\_CONDA\_ENVIRONMENT\_PATH/lib:$LD\_LIBRARY\_PATH

Click next and click on create, you’ll be able to see your environment which you created

A screenshot of a computer

Description automatically generated

Creating a ML Endpoint using the registered model and custom environment

Go to the endpoint tab and click on create new endpoint

A screenshot of a computer

Description automatically generated

Select the model which you registered

A white screen with black text

Description automatically generated

Give an appropriate model name and configure the authentication (or you can keep them as it )

Continue next till the time you reach the “Code+Environment” section of the set up. Here you’ll upload your scoring script and select the custom environment which you just created. (The scoring script can be found in the GitHub repo)

A screenshot of a computer

Description automatically generated

Click next and in the next step select the appropriate compute resource needed to deploy the model, in our case we have gone ahead with “Standard\_D2\_V2”, which has 2vCPU and 7Gb RAM. And then go to the review stage and create the end.

After couple of minutes your endpoint should be provisioned

A screenshot of a computer

Description automatically generated

After than you can go in the consume section to get information on how you can connect to this end point.