**IRIS ML Model deployment architecture:**

Diagram

Description automatically generated

**Below are the steps performed to create iris-model and deployed in ECS container and created endpoint using API Gateway.**

1. **Train your model using a framework and save the trained model's weights to a file. Below is the train.py function that performs the training task and saves model file:**

**Text

Description automatically generated**

1. **Create a container image using Docker that includes the necessary dependencies, such as the framework and any additional libraries needed to run the model. Below is the code:**

**Text

Description automatically generated**

**Execution:**

[ec2-user@ip-10-0-1-74 ~]$ **docker build . -t iris\_image**

Sending build context to Docker daemon 615.3MB

Step 1/5 : FROM public.ecr.aws/lambda/python:3.8

---> 7bc9979f2c04

Step 2/5 : RUN pip install scikit-learn pandas pickle4 requests

---> Using cache

---> 6162d88f6d82

Step 3/5 : COPY iris\_model.sav .

---> cdaec2179fe4

Step 4/5 : COPY lambda\_function.py .

---> 87f21a36f7d8

Step 5/5 : CMD [ "lambda\_function.lambda\_handler" ]

---> Running in e15a1e909625

Removing intermediate container e15a1e909625

---> eeb95b10c3ce

Successfully built eeb95b10c3ce

Successfully tagged iris\_image:latest

1. **Use AWS Elastic Container Registry (ECR) to store the above container image. Below are the commands.**

aws ecr create-repository --repository-name iris-registry --region us-west-1

# Log in to ECR

aws ecr get-login-password --region us-west-1 | docker login --username AWS --password-stdin 149509065833.dkr.ecr.us-west-1.amazonaws.com

docker build -t iris-registry .

# Tag the image

docker tag iris\_image:latest 149509065833.dkr.ecr.us-west-1.amazonaws.com/iris\_image

# Push the image to ECR

docker tag iris-registry:latest 149509065833.dkr.ecr.us-west-1.amazonaws.com/iris-registry:latest

1. **Create an ECS Cluster. Below is the screenshot.**

Graphical user interface, text, application, email

Description automatically generated

**Created policy for ECS Role and assigned permissions:**

Graphical user interface, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

1. **Create a Lambda function that calls the API endpoint and processes the input and output of the model.**

**Text

Description automatically generated**

1. **Create an API Gateway that triggers the Lambda function, and expose the API Gateway's endpoint as the API for the model. I have performed these tasks using Amazon Management Console but we can also use Cloudformation template to automate the whole process. I have created template but not tested.**

Created Lambda function inside ECS container and Created API endpoint:

Graphical user interface, text, application, email

Description automatically generatedGraphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

**Testing :**

Rest API POST call for endpoint:

<https://kx87oq3ou5.execute-api.us-west-1.amazonaws.com/testing>/predict

Input:

{"values": [[5, 2, 4, 4]]}

Output:

[2]

website for sample test: <https://reqbin.com/>

Graphical user interface, text, application, email

Description automatically generated

1. **Installed required libraries:**

[ec2-user@ip-10-0-1-74 ~]$ pip install scikit-learn pandas pickle4 requests

DEPRECATION: Python 2.7 reached the end of its life on January 1st, 2020. Please upgrade your Python as Python 2.7 is no longer maintained. pip 21.0 will drop support for Python 2.7 in January 2021. More details about Python 2 support in pip can be found at https://pip.pypa.io/en/latest/development/release-process/#python-2-support

Defaulting to user installation because normal site-packages is not writeable

Collecting scikit-learn

Downloading scikit\_learn-0.20.4-cp27-cp27mu-manylinux1\_x86\_64.whl (5.5 MB)

|████████████████████████████████| 5.5 MB 36.7 MB/s

Collecting pandas

Downloading pandas-0.24.2-cp27-cp27mu-manylinux1\_x86\_64.whl (10.1 MB)

|████████████████████████████████| 10.1 MB 106.3 MB/s

Collecting pickle4

Downloading pickle4-0.0.1.tar.gz (19 kB)

Requirement already satisfied: requests in /usr/lib/python2.7/site-packages (2.6.0)

Collecting numpy>=1.8.2

Downloading numpy-1.16.6-cp27-cp27mu-manylinux1\_x86\_64.whl (17.0 MB)

|████████████████████████████████| 17.0 MB 122.6 MB/s

Collecting scipy>=0.13.3

Downloading scipy-1.2.3-cp27-cp27mu-manylinux1\_x86\_64.whl (24.8 MB)

|████████████████████████████████| 24.8 MB 103.2 MB/s

Requirement already satisfied: python-dateutil>=2.5.0 in ./.local/lib/python2.7/site-packages (from pandas) (2.8.2)

Collecting pytz>=2011k

Downloading pytz-2022.7.1-py2.py3-none-any.whl (499 kB)

|████████████████████████████████| 499 kB 119.8 MB/s

Requirement already satisfied: setuptools in /usr/lib/python2.7/site-packages (from pickle4) (41.2.0)

Requirement already satisfied: six>=1.5 in ./.local/lib/python2.7/site-packages (from python-dateutil>=2.5.0->pandas) (1.16.0)

Using legacy 'setup.py install' for pickle4, since package 'wheel' is not installed.

Installing collected packages: numpy, scipy, scikit-learn, pytz, pandas, pickle4

Running setup.py install for pickle4 ... done

Successfully installed numpy-1.16.6 pandas-0.24.2 pickle4-0.0.1 pytz-2022.7.1 scikit-learn-0.20.4 scipy-1.2.3

**Issues faced during the deployment process:**

1. I not created keypair and not assigned to ECS cluster while creating and I had a problem of not able to connect using SSH with putty. I have to delete the cluster and assign keypair and with the .ppk file downloaded I have connected to putty.
2. After the while training the model train.py failing by saying sklearn and other libraries not found. So I have to install the libraries.
3. Docker command for creating image is not working and I have to spend lot of time to change parameters and try.

Faced problem while creating the ECR registry and I must add policy and assign to ECS role. The error is:

error occurred (AccessDeniedException) when calling the CreateRepository operation: User:

1. After creating API using API gateway ,when I click on the invoke URL,I have faced below issue:

API Gateway REST API endpoint 403 "Missing Authentication Token"