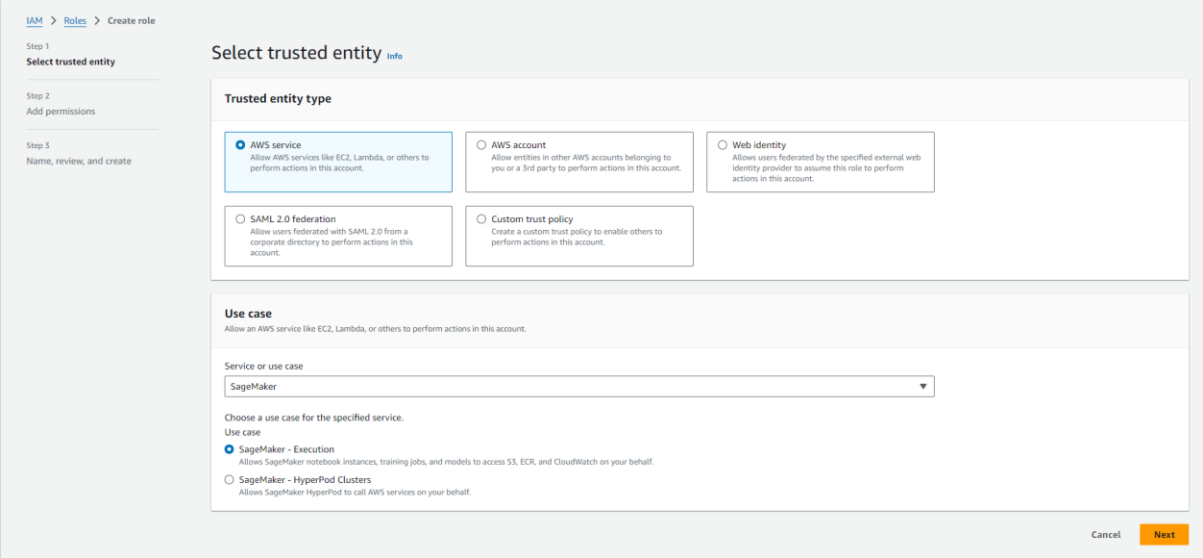
Name: Shreya Sameer Sane
SAP ID: 86062300009
Roll No: A054
M.Sc. Statistics & Data Science

Cloud Computing

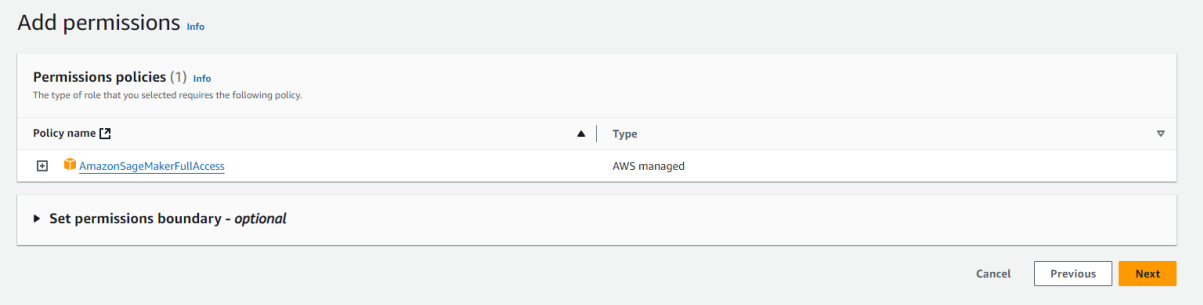
PRACTICAL 8

AMAZON SAGEMAKER

1)Creating IAM ROLE and assigning sagemaker permission

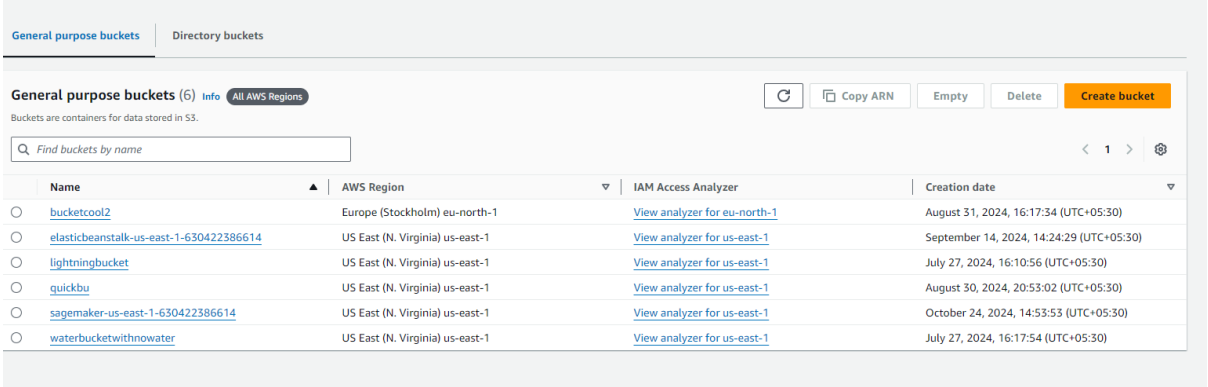


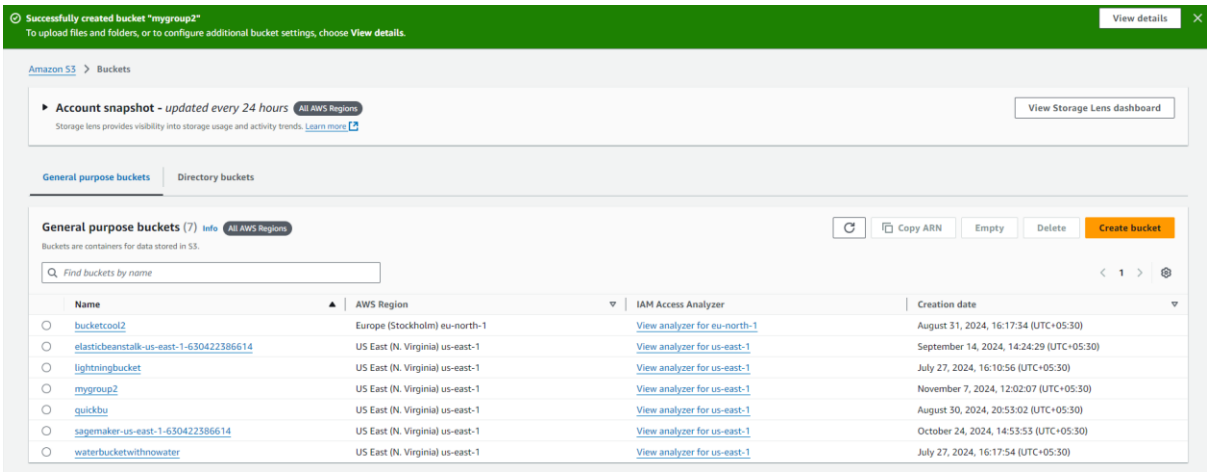
IAM Role is created.



2) creating s3 bucket named mygroup2

Name: Shreya Sameer Sane
SAP ID: 86062300009
Roll No: A054
M.Sc. Statistics & Data Science

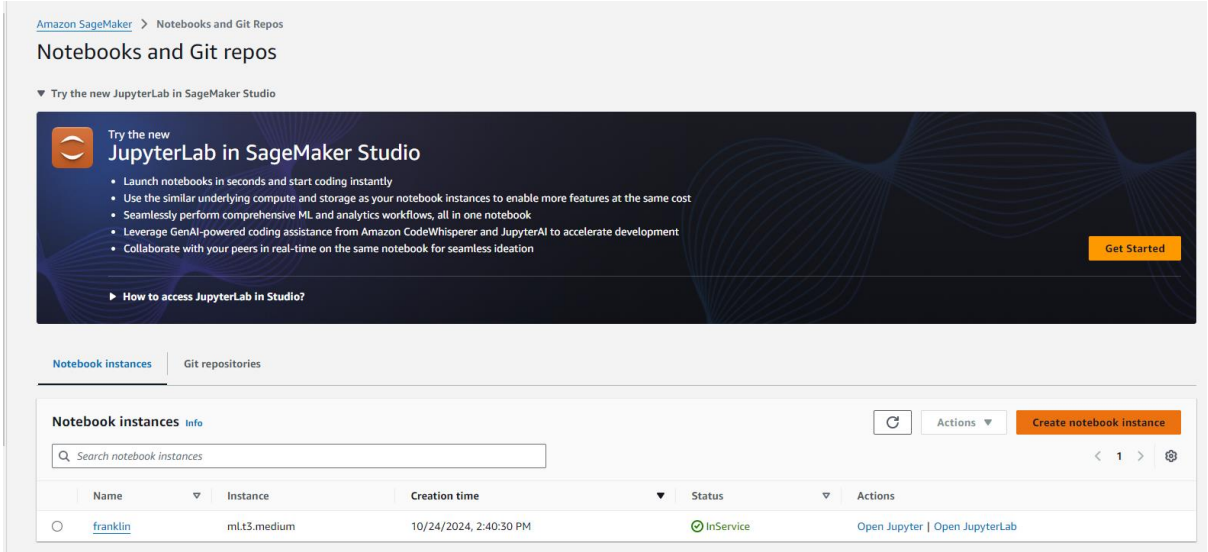




3)open Amazon SageMaker console

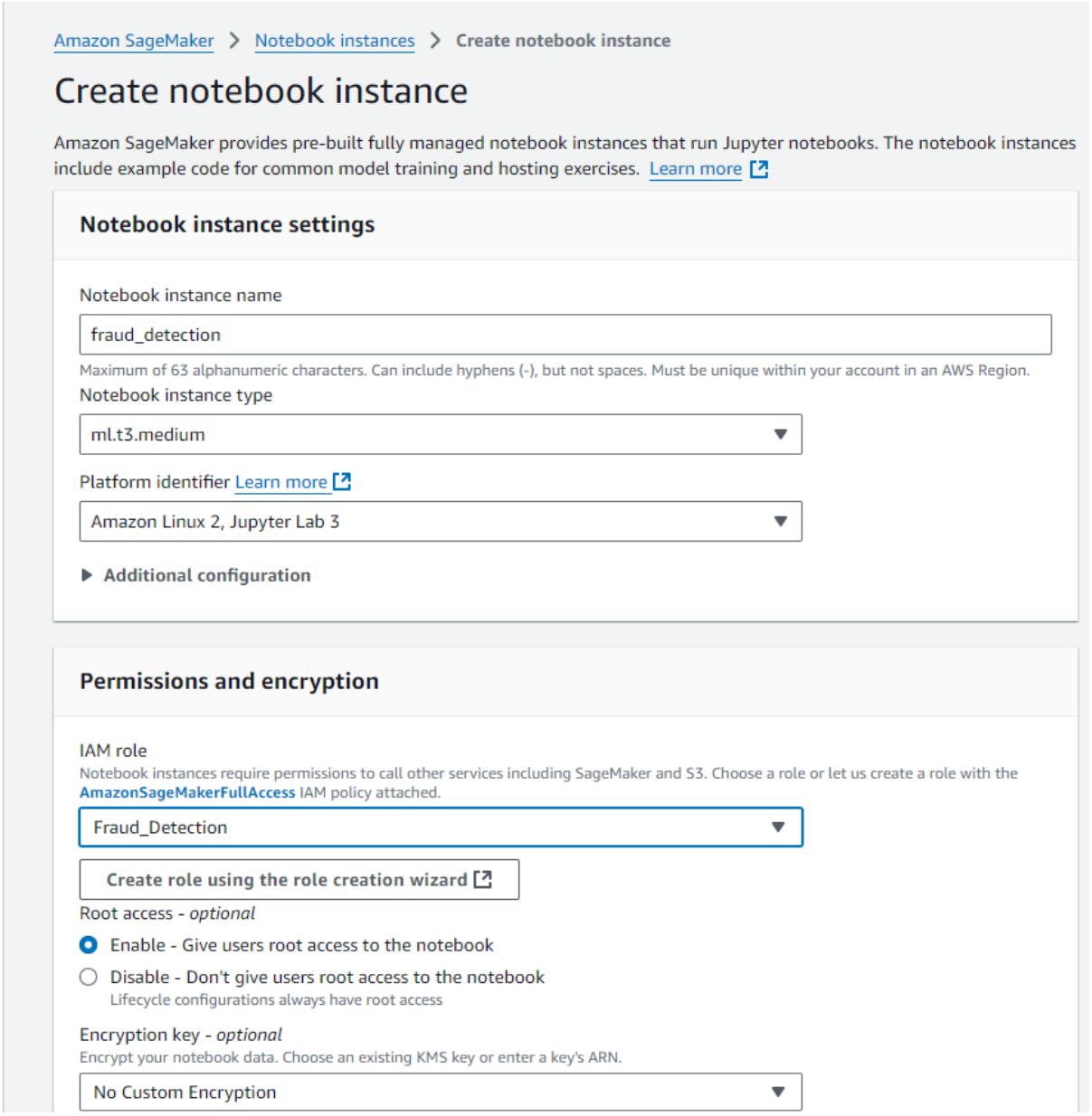
Select Notebook instances and click create notebook instances

Here we will assign the IAM role created earlier i.e fraud_detection

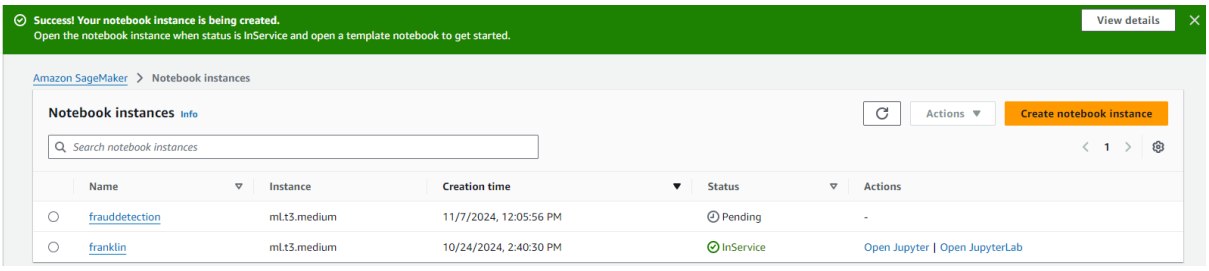


4) CREATE A JUPYTER NOTEBOOK

Name: Shreya Sameer Sane
SAP ID: 86062300009
Roll No: A054
M.Sc. Statistics & Data Science



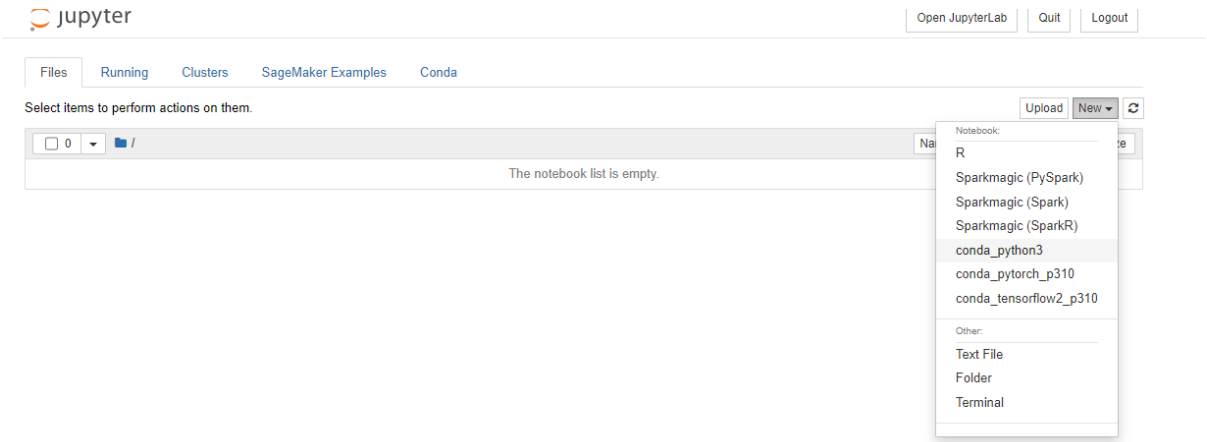
Notebook is created



1. Open Jupyter or JupyterLab according to the interface needed.
2. Go to File menu->Choose New-> Notebook.

Name: Shreya Sameer Sane
SAP ID: 86062300009
Roll No: A054
M.Sc. Statistics & Data Science

3. Select Kernel as ‘conda_python3’



Deploying the model (Here it is stored in s3 bucket that we had created)

In [1]:
```
import shap
X, y = shap.datasets.adult()
X_display, y_display = shap.datasets.adult(display=True)
feature_names = list(X.columns)
feature_names
```

```
Matplotlib is building the font cache; this may take a moment.
```

Out[1]:
```
['Age',
 'Workclass',
 'Education-Num',
 'Marital Status',
 'Occupation',
 'Relationship',
 'Race',
 'Sex',
 'Capital Gain',
 'Capital Loss',
 'Hours per week',
 'Country']
```

In [7]:
```
import sagemaker, boto3, os
bucket = sagemaker.Session().default_bucket()
prefix = "demo-sagemaker-xgboost-adult-income-prediction"

boto3.Session().resource('s3').Bucket(bucket).Object(
    os.path.join(prefix, 'data/train.csv')).upload_file('train.csv')
boto3.Session().resource('s3').Bucket(bucket).Object(
    os.path.join(prefix, 'data/validation.csv')).upload_file('validation.csv')
```

```
sagemaker.config INFO - Not applying SDK defaults from location: /etc/xdg/sagemaker/config.yaml
sagemaker.config INFO - Not applying SDK defaults from location: /home/ec2-user/.config/sagemaker/config.yaml
```

In [8]:
```
import sagemaker

region = sagemaker.Session().boto_region_name
print("AWS Region: {}".format(region))

role = sagemaker.get_execution_role()
print("RoleArn: {}".format(role))
```

```
AWS Region: us-east-1
RoleArn: arn:aws:iam::975050009706:role/lucifer007
```

Name: Shreya Sameer Sane
SAP ID: 86062300009
Roll No: A054
M.Sc. Statistics & Data Science

```
! aws s3 cp {rule_output_path} ./ --recursive

from IPython.display import FileLink, FileLinks
display("Click link below to view the XGBoost Training report", FileLink("CreateXgboostReport/xgboost_report.html"))
```

```
download: s3://sagemaker-us-east-1-975050009706/demo-sagemaker-xgboost-adult-income-prediction/xgboost_model/sagemaker-xgboost-2024-10-24-09-29-24-130/rule-output/CreateXgboostReport/xgboost-reports/EvaluationMetrics.json to CreateXgboostReport/xgboost-reports/EvaluationMetrics.json
download: s3://sagemaker-us-east-1-975050009706/demo-sagemaker-xgboost-adult-income-prediction/xgboost_model/sagemaker-xgboost-2024-10-24-09-29-24-130/rule-output/CreateXgboostReport/xgboost-reports/FeatureImportance.json to CreateXgboostReport/xgboost-reports/FeatureImportance.json
download: s3://sagemaker-us-east-1-975050009706/demo-sagemaker-xgboost-adult-income-prediction/xgboost_model/sagemaker-xgboost-2024-10-24-09-29-24-130/rule-output/ProfilerReport/profiler-output/profiler-report.ipynb to ProfilerReport/profiler-output/profiler-report.ipynb
download: s3://sagemaker-us-east-1-975050009706/demo-sagemaker-xgboost-adult-income-prediction/xgboost_model/sagemaker-xgboost-2024-10-24-09-29-24-130/rule-output/CreateXgboostReport/xgboost-reports/ConfusionMatrix.json to CreateXgboostReport/xgboost-reports/ConfusionMatrix.json
```

```
from sagemaker.debugger import Rule, ProfilerRule, rule_configs
from sagemaker.session import TrainingInput

s3_output_location='s3://{}/{}/{}'.format(bucket, prefix, 'xgboost_model')

container=sagemaker.image_uris.retrieve("xgboost", region, "1.2-1")
print(container)

xgb_model=sagemaker.estimator.Estimator(
    image_uri=container,
    role=role,
    instance_count=1,
    instance_type='ml.m4.xlarge',
    volume_size=5,
    output_path=s3_output_location,
    sagemaker_session=sagemaker.Session(),
    rules=[
        Rule.sagemaker(rule_configs.create_xgboost_report()),
        ProfilerRule.sagemaker(rule_configs.ProfilerReport())
    ]
)
```

Name: Shreya Sameer Sane
SAP ID: 86062300009
Roll No: A054
M.Sc. Statistics & Data Science

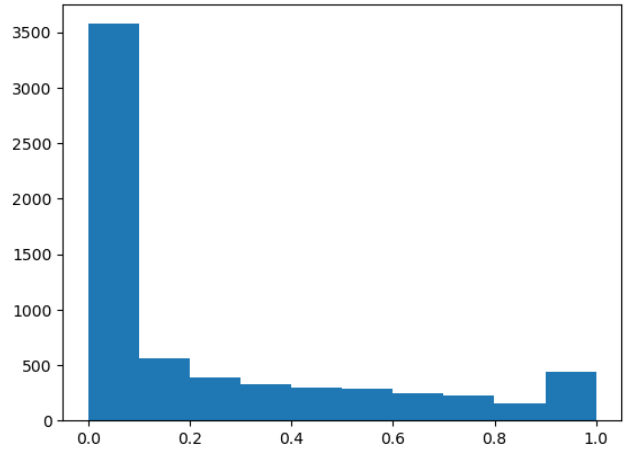
In [18]:
```
xgb_predictor.endpoint_name
```

Out[18]:
```
'sagemaker-xgboost-2024-10-24-09-34-02-816'
```

In [19]:
```
import numpy as np
def predict(data, rows=1000):
    split_array = np.array_split(data, int(data.shape[0] / float(rows) + 1))
    predictions = ''
    for array in split_array:
        predictions = ','.join([predictions, xgb_predictor.predict(array).decode('utf-8')])
    return np.fromstring(predictions[1:], sep=',')
```

In [20]:
```
import matplotlib.pyplot as plt

predictions=predict(test.to_numpy()[:,1:])
plt.hist(predictions)
plt.show()
```



In [21]:
```
import sklearn

cutoff=0.5
print(sklearn.metrics.confusion_matrix(test.iloc[:, 0], np.where(predictions > cutoff, 1, 0)))
print(sklearn.metrics.classification_report(test.iloc[:, 0], np.where(predictions > cutoff, 1, 0)))
```

```
[[4670  356]
 [ 480 1007]]
              precision    recall  f1-score   support

           0       0.91      0.93      0.92      5026
           1       0.74      0.68      0.71      1487

    accuracy                           0.87      6513
   macro avg       0.82      0.80      0.81      6513
weighted avg       0.87      0.87      0.87      6513
```

Name: Shreya Sameer Sane
SAP ID: 86062300009
Roll No: A054
M.Sc. Statistics & Data Science