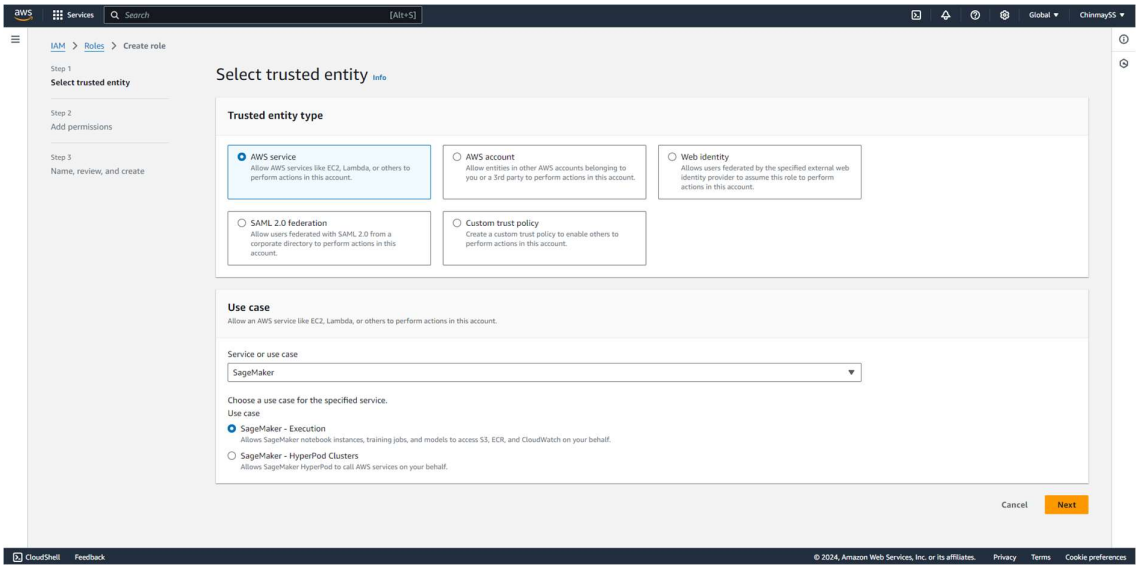
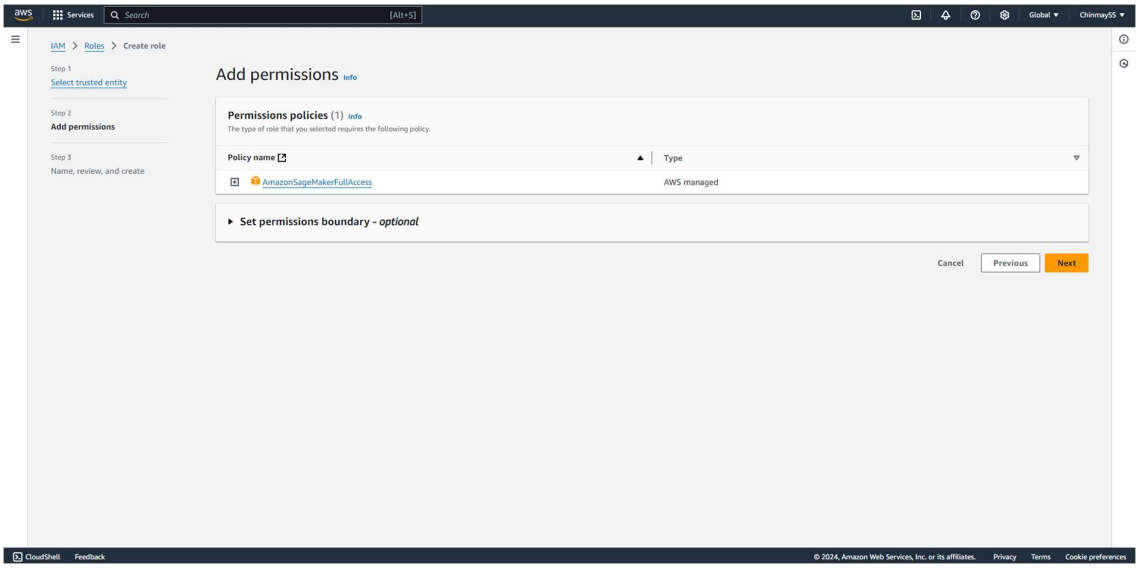
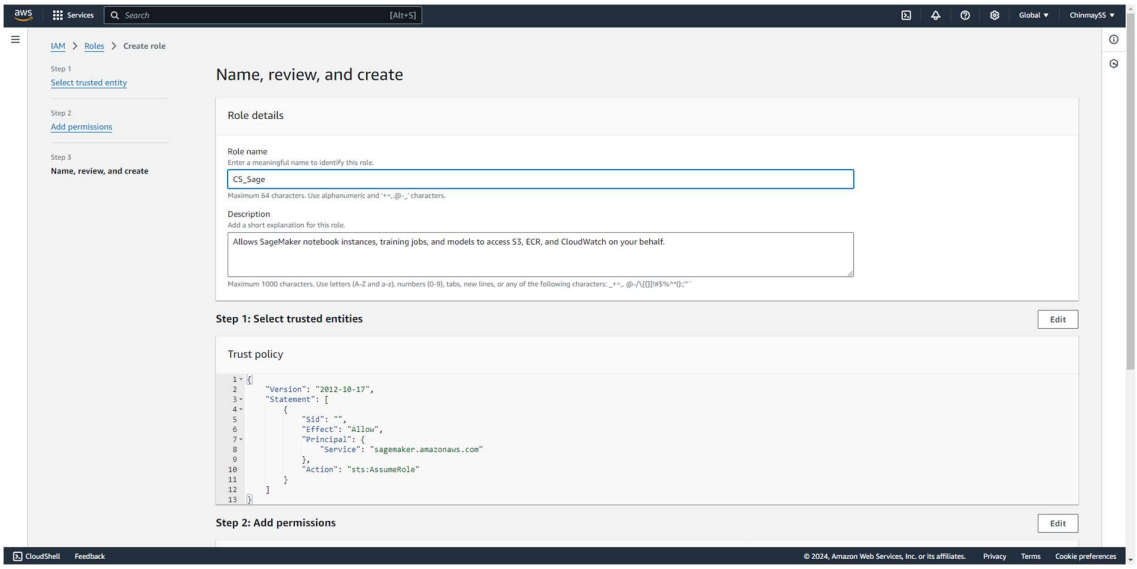
SVKM'S NMIMS Nilkamal School of Mathematics, Applied Statistics & Analytics
Master of Science (Statistics & Data Science)

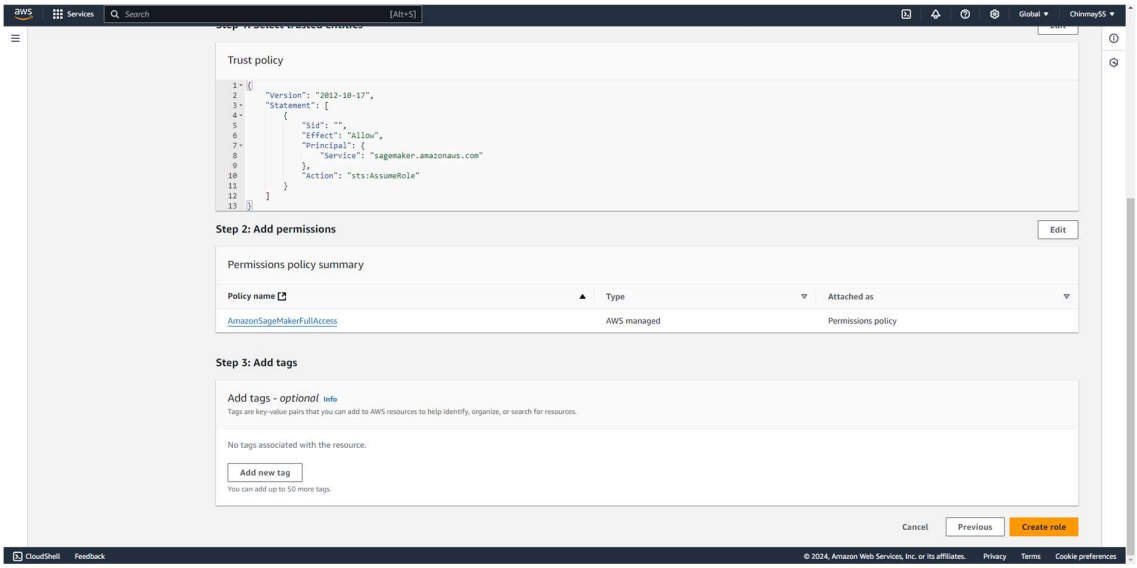
Practical-8: Amazon Sagemaker

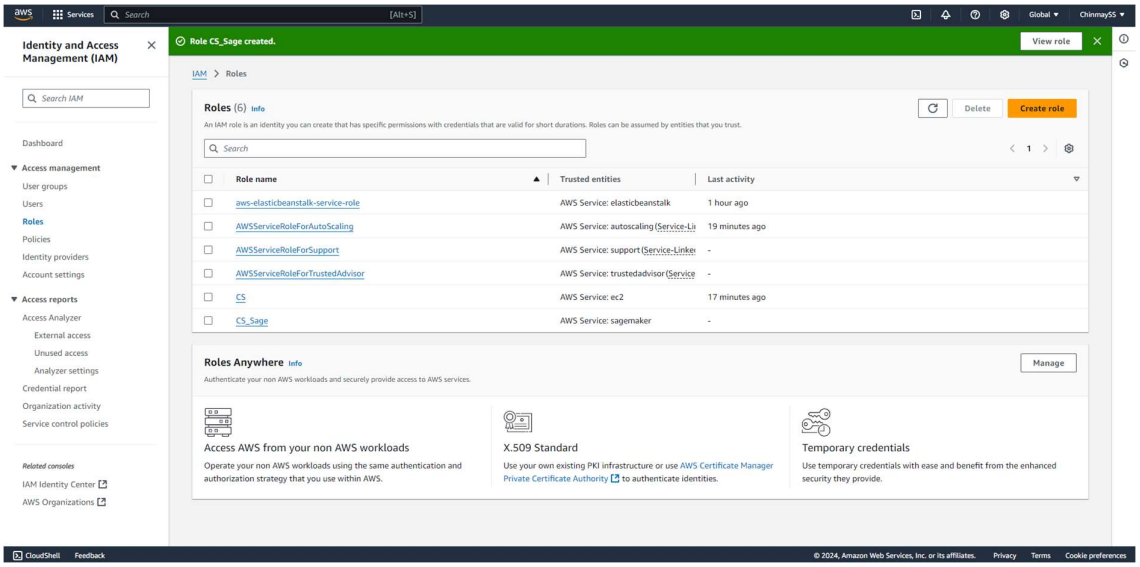
Name: Chinmay Shinde Roll No.: A066 SAP ID: 86062300031

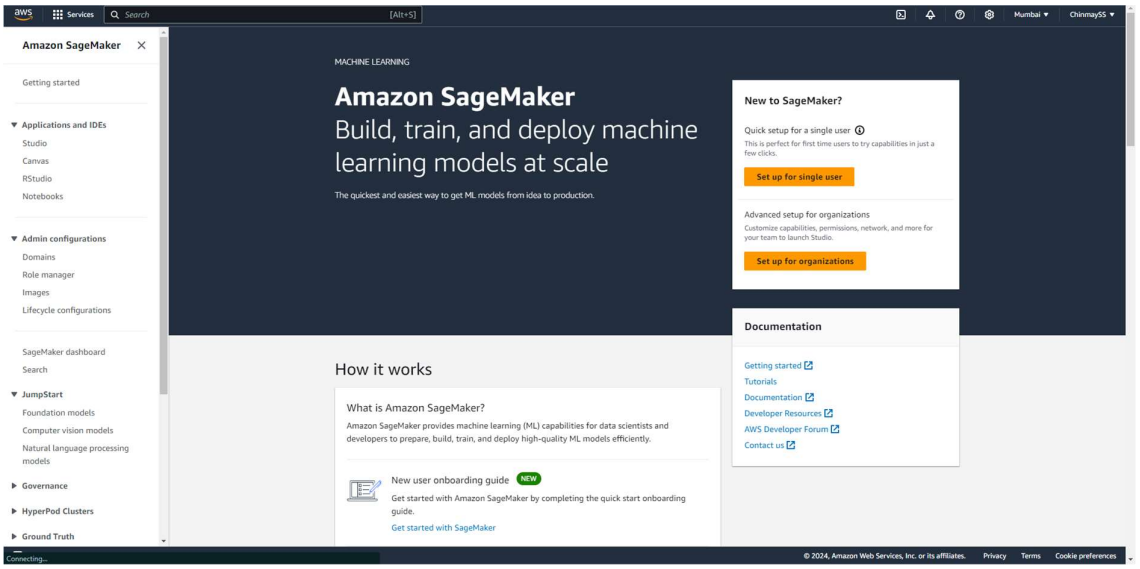


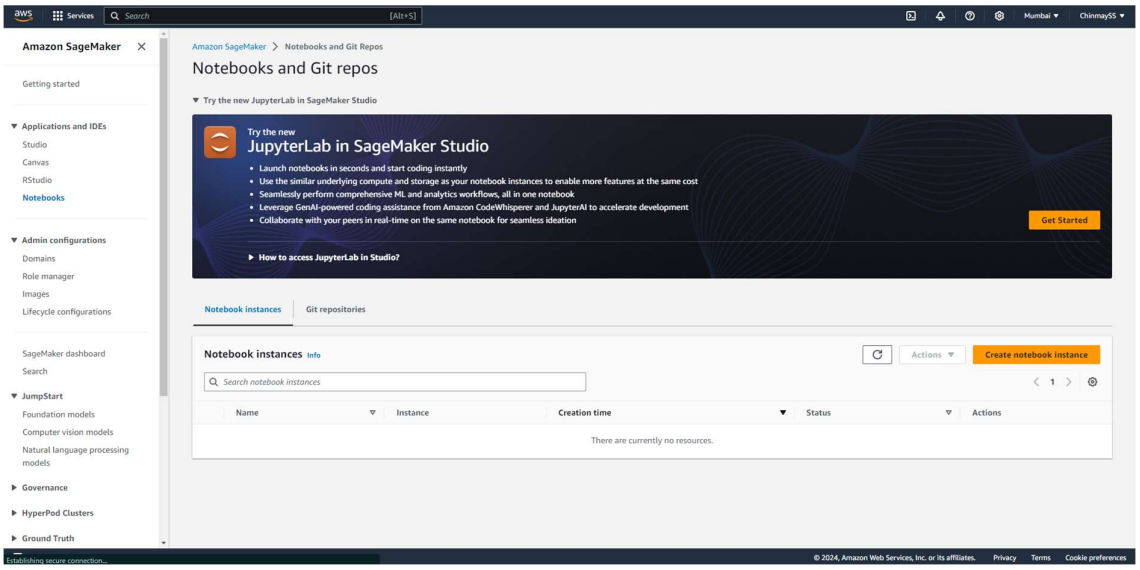


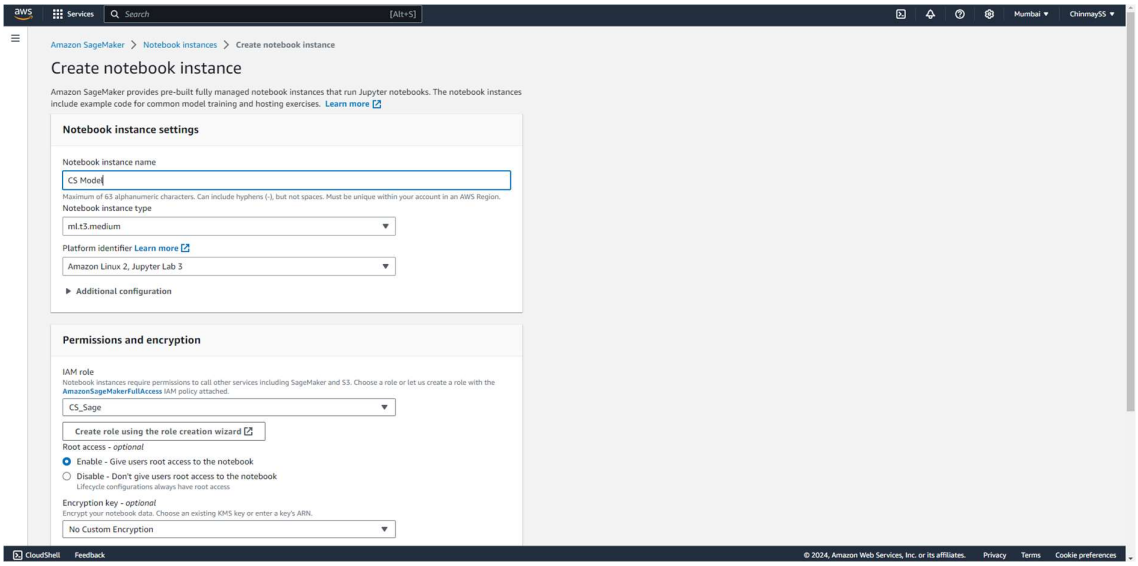

Services
Search
[Alt+S]
Global
ChinmaySS
IAM > Roles > Create role
Step 1
Select trusted entity
Step 2
Add permissions
Step 3
Name, review, and create
Name, review, and create
Role details
Role name
Enter a meaningful name to identify this role.
CS_Sage
Maximum 64 characters. Use alphanumeric and '+=,.@-_' characters.
Description
Add a short explanation for this role.
Allows SageMaker notebook instances, training jobs, and models to access S3, ECR, and CloudWatch on your behalf.
Maximum 1000 characters. Use letters (A-Z and a-z), numbers (0-9), tabs, new lines, or any of the following characters: _+=, .@-/\[{}]!#$%^*();:"'
Step 1: Select trusted entities
Edit
Trust policy
{
"Version": "2012-10-17",
"Statement": [
{
"Sid": "",
"Effect": "Allow",
"Principal": {
"Service": "sagemaker.amazonaws.com"
},
"Action": "sts:AssumeRole"
}
]
}
Step 2: Add permissions
Edit
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

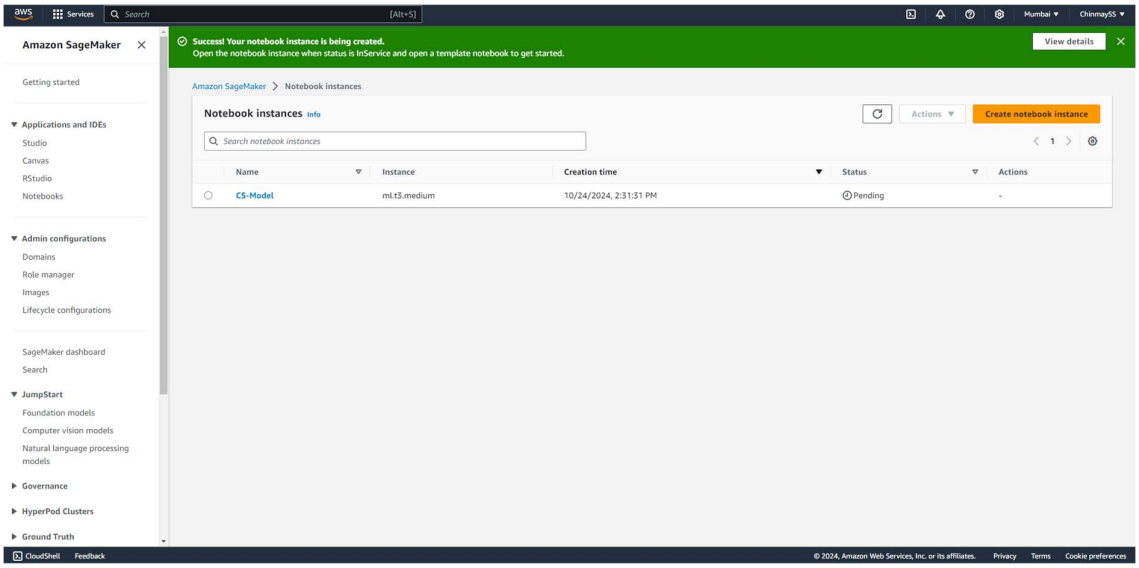

Services
Search
[Alt+S]
Global
ChinmaySS
Trust policy
{
"Version": "2012-10-17",
"Statement": [
{
"Sid": "",
"Effect": "Allow",
"Principal": {
"Service": "sagemaker.amazonaws.com"
},
"Action": "sts:AssumeRole"
}
]
}
Step 2: Add permissions
Edit
Permissions policy summary
Policy name
Type
Attached as
AmazonSageMakerFullAccess
AWS managed
Permissions policy
Step 3: Add tags
Add tags - optional Info
Tags are key-value pairs that you can add to AWS resources to help identify, organize, or search for resources.
No tags associated with the resource.
Add new tag
You can add up to 50 more tags.
Cancel
Previous
Create role
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

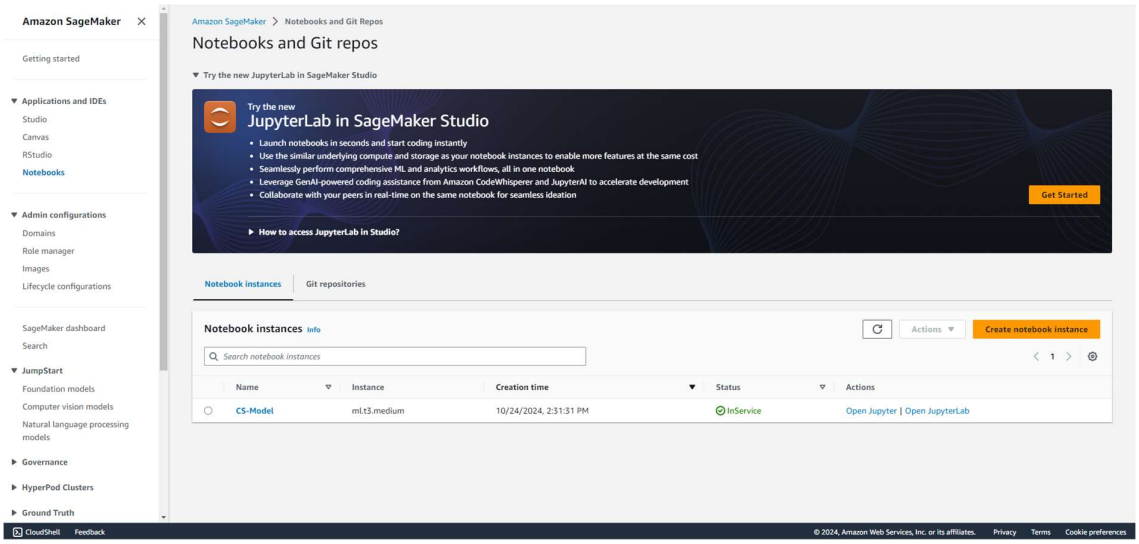

Role CS_Sage created.
View role
Identity and Access Management (IAM)
Search IAM
Dashboard
Access management
User groups
Users
Roles
Policies
Identity providers
Account settings
Access reports
Access Analyzer
External access
Unused access
Analyzer settings
Credential report
Organization activity
Service control policies
Related consoles
IAM Identity Center
AWS Organizations
IAM > Roles
Roles (6) Info
An IAM role is an identity you can create that has specific permissions with credentials that are valid for short durations. Roles can be assumed by entities that you trust.
Delete
Create role
Search
Role name
Trusted entities
Last activity
aws-elasticbeanstalk-service-role
AWS Service: elasticbeanstalk
1 hour ago
AWSServiceRoleForAutoScaling
AWS Service: autoscaling (Service-Li
19 minutes ago
AWSServiceRoleForSupport
AWS Service: support (Service-Linke
-
AWSServiceRoleForTrustedAdvisor
AWS Service: trustedadvisor (Service
-
CS
AWS Service: ec2
17 minutes ago
CS_Sage
AWS Service: sagemaker
-
Roles Anywhere Info
Authenticate your non AWS workloads and securely provide access to AWS services.
Manage
Access AWS from your non AWS workloads
Operate your non AWS workloads using the same authentication and authorization strategy that you use within AWS.
X.509 Standard
Use your own existing PKI infrastructure or use AWS Certificate Manager Private Certificate Authority to authenticate identities.
Temporary credentials
Use temporary credentials with ease and benefit from the enhanced security they provide.
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

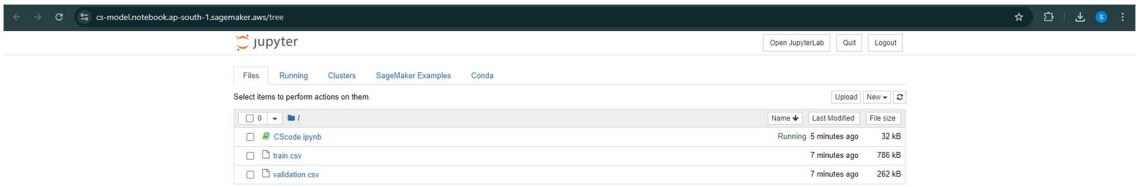

Amazon SageMaker
Getting started
Applications and IDEs
Studio
Canvas
RStudio
Notebooks
Admin configurations
Domains
Role manager
Images
Lifecycle configurations
SageMaker dashboard
Search
JumpStart
Foundation models
Computer vision models
Natural language processing models
Governance
HyperPod Clusters
Ground Truth
MACHINE LEARNING
Amazon SageMaker
Build, train, and deploy machine learning models at scale
The quickest and easiest way to get ML models from idea to production.
New to SageMaker?
Quick setup for a single user
This is perfect for first time users to try capabilities in just a few clicks.
Set up for single user
Advanced setup for organizations
Customize capabilities, permissions, network, and more for your team to launch Studio.
Set up for organizations
Documentation
Getting started
Tutorials
Documentation
Developer Resources
AWS Developer Forum
Contact us
How it works
What is Amazon SageMaker?
Amazon SageMaker provides machine learning (ML) capabilities for data scientists and developers to prepare, build, train, and deploy high-quality ML models efficiently.
New user onboarding guide
NEW
Get started with Amazon SageMaker by completing the quick start onboarding guide.
Get started with SageMaker
Connecting...
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

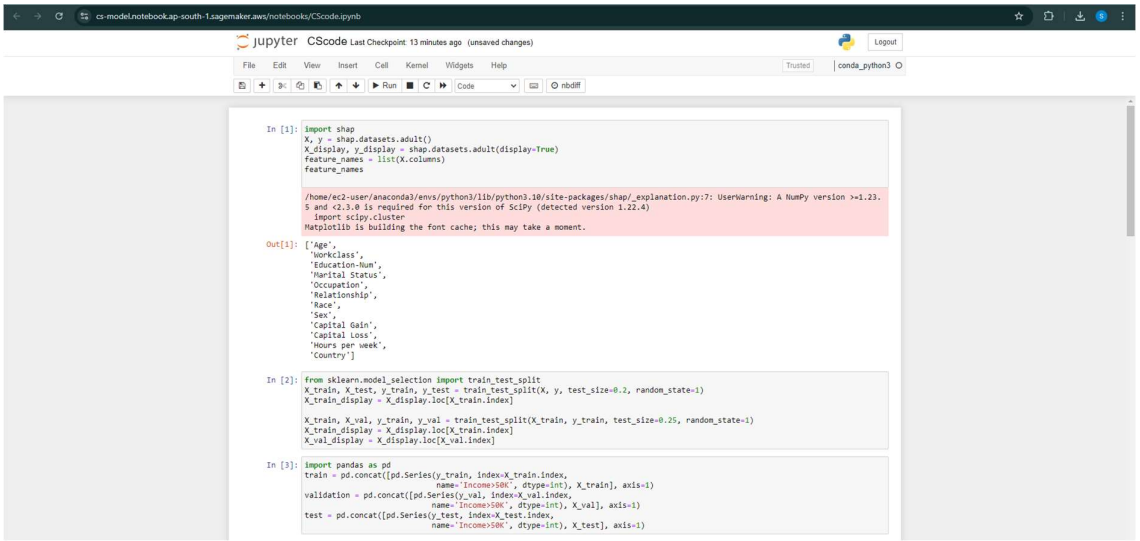

Amazon SageMaker
Getting started
Applications and IDEs
Studio
Canvas
RStudio
Notebooks
Admin configurations
Domains
Role manager
Images
Lifecycle configurations
SageMaker dashboard
Search
JumpStart
Foundation models
Computer vision models
Natural language processing models
Governance
HyperPod Clusters
Ground Truth
Amazon SageMaker > Notebooks and Git Repos
Notebooks and Git repos
Try the new JupyterLab in SageMaker Studio
Try the new
JupyterLab in SageMaker Studio
Launch notebooks in seconds and start coding instantly
Use the similar underlying compute and storage as your notebook instances to enable more features at the same cost
Seamlessly perform comprehensive ML and analytics workflows, all in one notebook
Leverage GenAI-powered coding assistance from Amazon CodeWhisperer and JupyterAI to accelerate development
Collaborate with your peers in real-time on the same notebook for seamless ideation
Get Started
How to access JupyterLab in Studio?
Notebook instances
Git repositories
Notebook instances Info
Actions
Create notebook instance
Search notebook instances
Name
Instance
Creation time
Status
Actions
There are currently no resources.
Establishing secure connection...
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

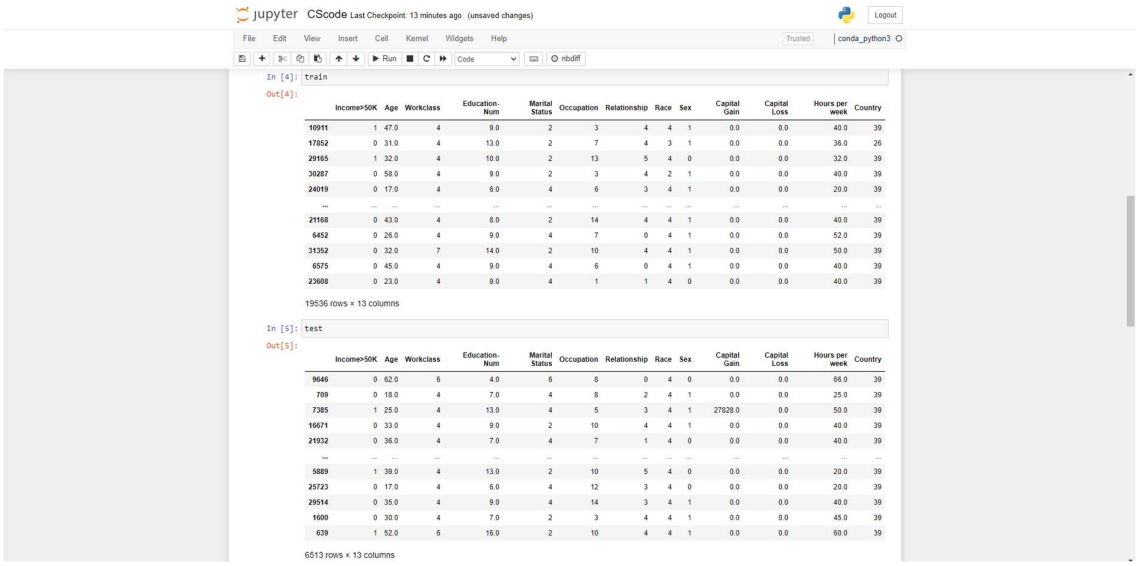

Services
Search
[Alt+S]
Mumbai
ChinmaySS
Amazon SageMaker > Notebook instances > Create notebook instance
Create notebook instance
Amazon SageMaker provides pre-built fully managed notebook instances that run Jupyter notebooks. The notebook instances include example code for common model training and hosting exercises. Learn more
Notebook instance settings
Notebook instance name
CS Model
Maximum of 63 alphanumeric characters. Can include hyphens (-), but not spaces. Must be unique within your account in an AWS Region.
Notebook instance type
ml.t3.medium
Platform identifier Learn more
Amazon Linux 2, Jupyter Lab 3
Additional configuration
Permissions and encryption
IAM role
Notebook instances require permissions to call other services including SageMaker and S3. Choose a role or let us create a role with the AmazonSageMakerFullAccess IAM policy attached.
CS_Sage
Create role using the role creation wizard
Root access - optional
Enable - Give users root access to the notebook
Disable - Don't give users root access to the notebook
Lifecycle configurations always have root access
Encryption key - optional
Encrypt your notebook data. Choose an existing KMS key or enter a key's ARN.
No Custom Encryption
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

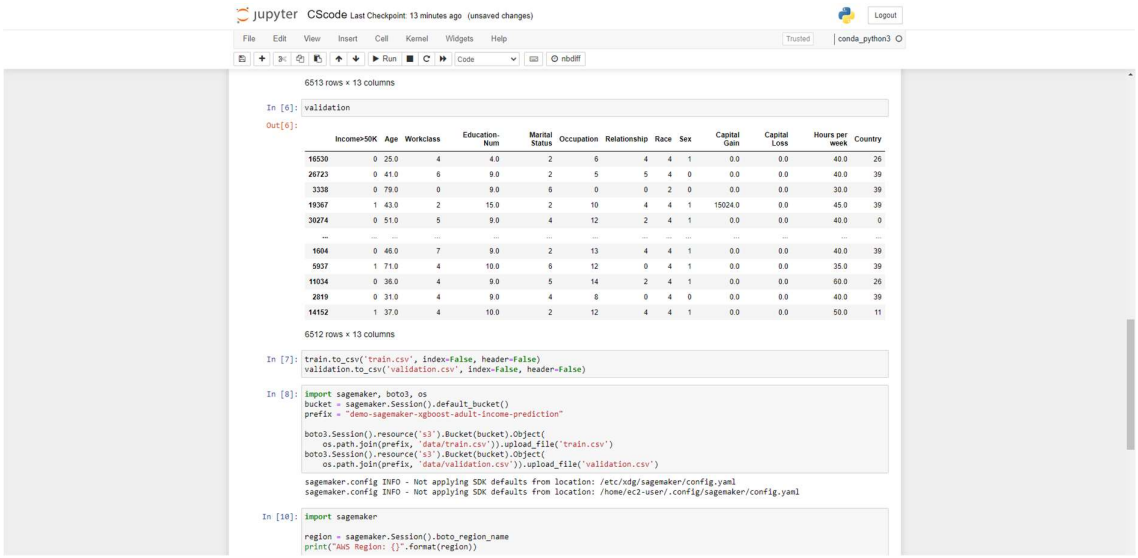

Services
Search
[Alt+S]
Mumbai
ChinmaySS
Amazon SageMaker
Success! Your notebook instance is being created.
Open the notebook instance when status is InService and open a template notebook to get started.
View details
Getting started
Applications and IDEs
Studio
Canvas
RStudio
Notebooks
Admin configurations
Domains
Role manager
Images
Lifecycle configurations
SageMaker dashboard
Search
JumpStart
Foundation models
Computer vision models
Natural language processing models
Governance
HyperPod Clusters
Ground Truth
Amazon SageMaker > Notebook instances
Notebook instances Info
Actions
Create notebook instance
Search notebook instances
Name
Instance
Creation time
Status
Actions
CS-Model
ml.t3.medium
10/24/2024, 2:31:31 PM
Pending
-
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

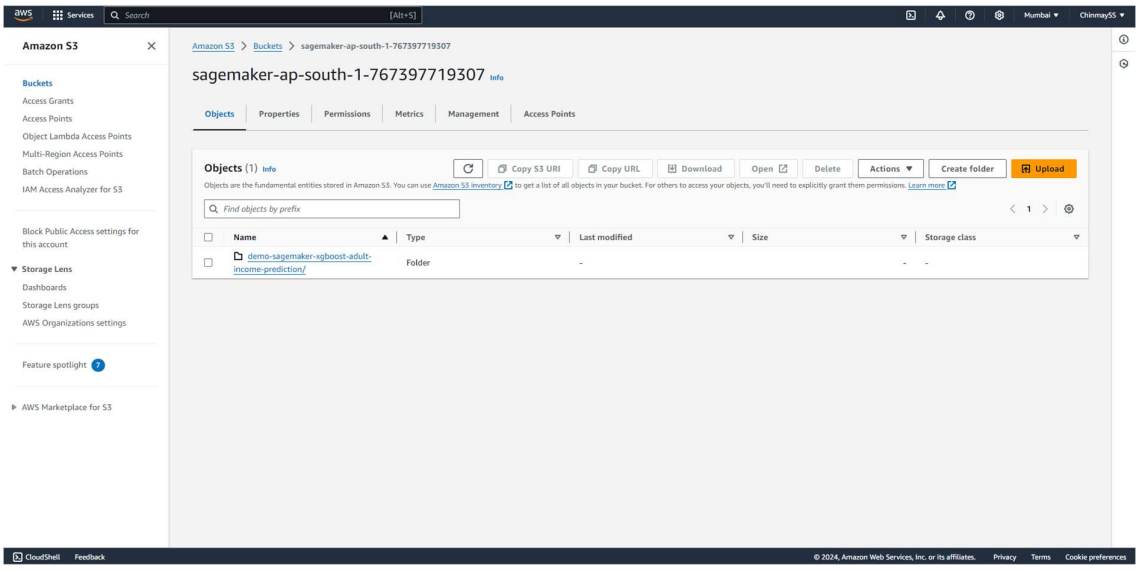

Amazon SageMaker
Getting started
Applications and IDEs
Studio
Canvas
RStudio
Notebooks
Admin configurations
Domains
Role manager
Images
Lifecycle configurations
SageMaker dashboard
Search
JumpStart
Foundation models
Computer vision models
Natural language processing models
Governance
HyperPod Clusters
Ground Truth
Amazon SageMaker > Notebooks and Git Repos
Notebooks and Git repos
Try the new JupyterLab in SageMaker Studio
Try the new
JupyterLab in SageMaker Studio
Launch notebooks in seconds and start coding instantly
Use the similar underlying compute and storage as your notebook instances to enable more features at the same cost
Seamlessly perform comprehensive ML and analytics workflows, all in one notebook
Leverage GenAI-powered coding assistance from Amazon CodeWhisperer and JupyterAI to accelerate development
Collaborate with your peers in real-time on the same notebook for seamless ideation
Get Started
How to access JupyterLab in Studio?
Notebook instances
Git repositories
Notebook instances Info
Actions
Create notebook instance
Search notebook instances
Name
Instance
Creation time
Status
Actions
CS-Model
ml.t3.medium
10/24/2024, 2:31:31 PM
InService
Open Jupyter | Open JupyterLab
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

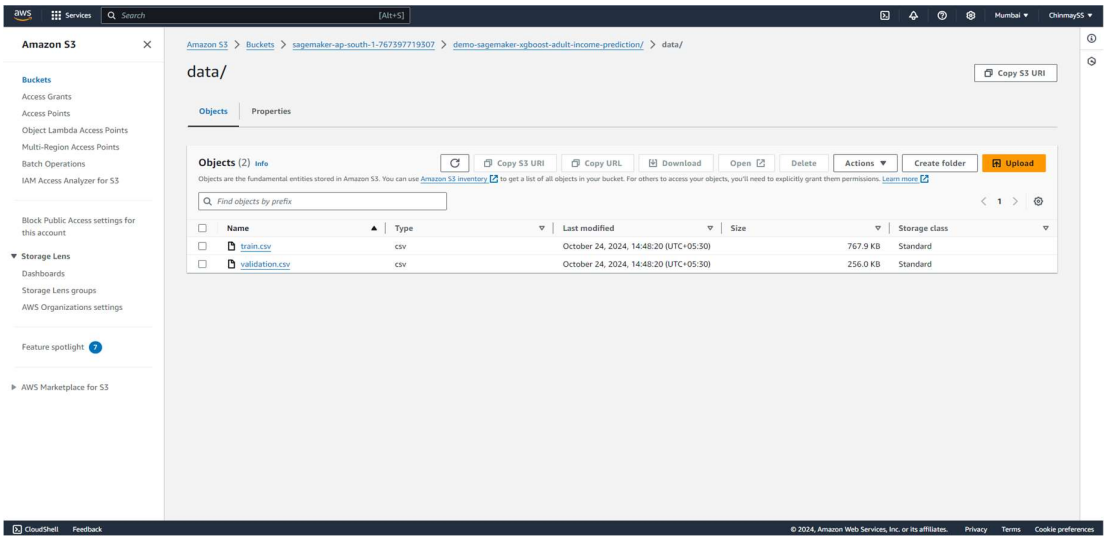

cs-model.notebook.ap-south-1.sagemaker.aws/tree
jupyter
Open JupyterLab
Quit
Logout
Files
Running
Clusters
SageMaker Examples
Conda
Select items to perform actions on them.
Upload
New
Name
Last Modified
File size
CScode.ipynb
Running 5 minutes ago
32 kB
train.csv
7 minutes ago
786 kB
validation.csv
7 minutes ago
262 kB

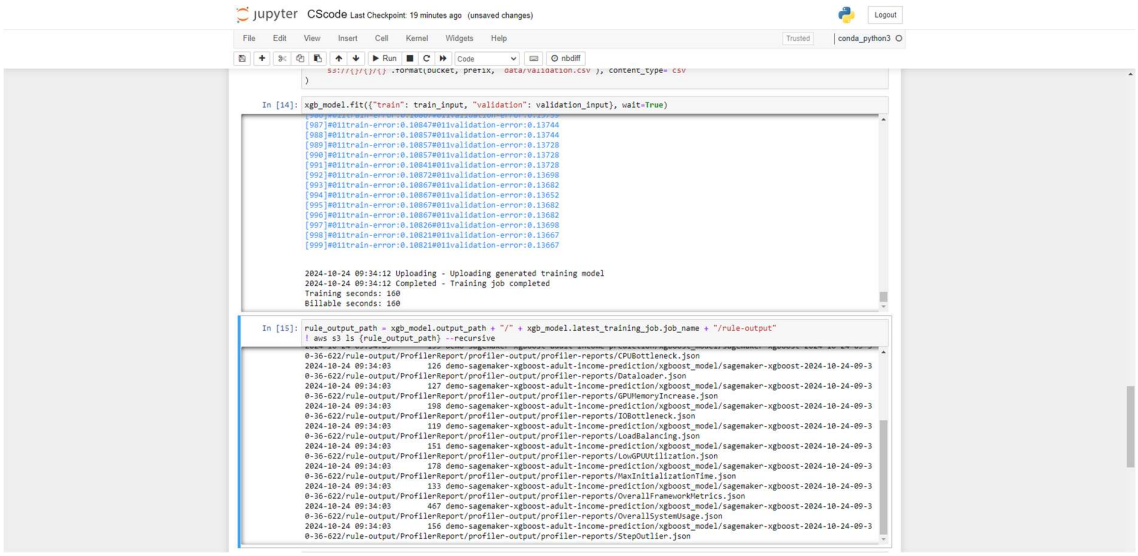

cs-model.notebook.ap-south-1.sagemaker.aws/notebooks/CScode.ipynb
jupyter CScode Last Checkpoint: 13 minutes ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
conda_python3
Run
Code
nbdiff
In [1]:
import shap
X, y = shap.datasets.adult()
X_display, y_display = shap.datasets.adult(display=True)
feature_names = list(X.columns)
feature_names
/home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages/shap/_explanation.py:7: UserWarning: A NumPy version >=1.23.5 and <2.3.0 is required for this version of SciPy (detected version 1.22.4)
import scipy.cluster
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 [2]:
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1)
X_train_display = X_display.loc[X_train.index]
X_train, X_val, y_train, y_val = train_test_split(X_train, y_train, test_size=0.25, random_state=1)
X_train_display = X_display.loc[X_train.index]
X_val_display = X_display.loc[X_val.index]
In [3]:
import pandas as pd
train = pd.concat([pd.Series(y_train, index=X_train.index,
name='Income>50K', dtype=int), X_train], axis=1)
validation = pd.concat([pd.Series(y_val, index=X_val.index,
name='Income>50K', dtype=int), X_val], axis=1)
test = pd.concat([pd.Series(y_test, index=X_test.index,
name='Income>50K', dtype=int), X_test], axis=1)

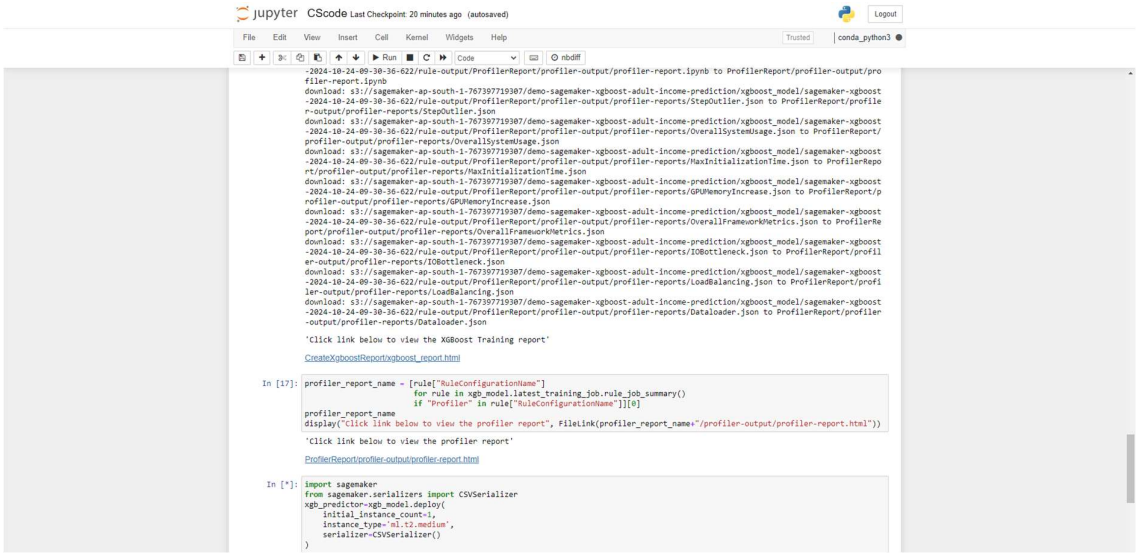

jupyter CScode Last Checkpoint: 13 minutes ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
conda_python3
Run
Code
nbdiff
In [4]:
train
Out[4]:
Income>50K Age Workclass Education-Num Marital Status Occupation Relationship Race Sex Capital Gain Capital Loss Hours per week Country
10911 1 47.0 4 9.0 2 3 4 4 1 0.0 0.0 40.0 39
17852 0 31.0 4 13.0 2 7 4 3 1 0.0 0.0 36.0 26
29165 1 32.0 4 10.0 2 13 5 4 0 0.0 0.0 32.0 39
30287 0 58.0 4 9.0 2 3 4 2 1 0.0 0.0 40.0 39
24019 0 17.0 4 6.0 4 6 3 4 1 0.0 0.0 20.0 39
... ...
21168 0 43.0 4 8.0 2 14 4 4 1 0.0 0.0 40.0 39
6452 0 26.0 4 9.0 4 7 0 4 1 0.0 0.0 52.0 39
31352 0 32.0 7 14.0 2 10 4 4 1 0.0 0.0 50.0 39
6575 0 45.0 4 9.0 4 6 0 4 1 0.0 0.0 40.0 39
23608 0 23.0 4 9.0 4 1 1 4 0 0.0 0.0 40.0 39
19536 rows × 13 columns
In [5]:
test
Out[5]:
Income>50K Age Workclass Education-Num Marital Status Occupation Relationship Race Sex Capital Gain Capital Loss Hours per week Country
9646 0 62.0 6 4.0 6 8 0 4 0 0.0 0.0 66.0 39
709 0 18.0 4 7.0 4 8 2 4 1 0.0 0.0 25.0 39
7385 1 25.0 4 13.0 4 5 3 4 1 27828.0 0.0 50.0 39
16671 0 33.0 4 9.0 2 10 4 4 1 0.0 0.0 40.0 39
21932 0 36.0 4 7.0 4 7 1 4 0 0.0 0.0 40.0 39
... ...
5889 1 39.0 4 13.0 2 10 5 4 0 0.0 0.0 20.0 39
25723 0 17.0 4 6.0 4 12 3 4 0 0.0 0.0 20.0 39
29514 0 35.0 4 9.0 4 14 3 4 1 0.0 0.0 40.0 39
1600 0 30.0 4 7.0 2 3 4 4 1 0.0 0.0 45.0 39
639 1 52.0 6 16.0 2 10 4 4 1 0.0 0.0 60.0 39
6513 rows × 13 columns

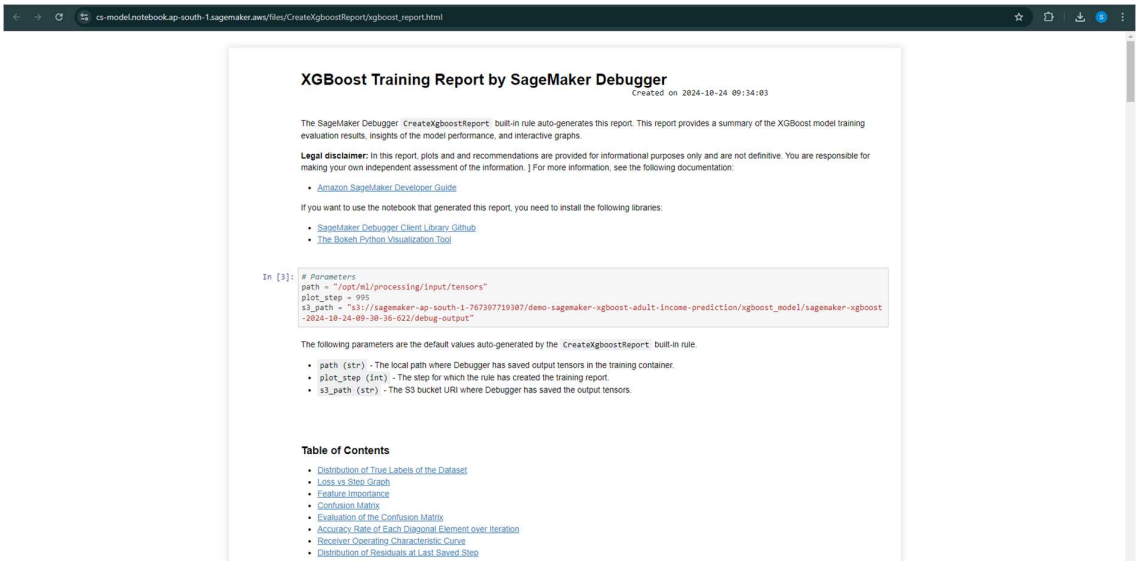

jupyter CScode Last Checkpoint: 13 minutes ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
conda_python3
Run
Code
nbdiff
6513 rows × 13 columns
In [6]: validation
Out[6]:
Income>50K Age Workclass Education-Num Marital Status Occupation Relationship Race Sex Capital Gain Capital Loss Hours per week Country
16530 0 25.0 4 4.0 2 6 4 4 1 0.0 0.0 40.0 26
26723 0 41.0 6 9.0 2 5 5 4 0 0.0 0.0 40.0 39
3338 0 79.0 0 9.0 6 0 0 2 0 0.0 0.0 30.0 39
19367 1 43.0 2 15.0 2 10 4 4 1 15024.0 0.0 45.0 39
30274 0 51.0 5 9.0 4 12 2 4 1 0.0 0.0 40.0 0
... ...
1604 0 46.0 7 9.0 2 13 4 4 1 0.0 0.0 40.0 39
5937 1 71.0 4 10.0 6 12 0 4 1 0.0 0.0 35.0 39
11034 0 36.0 4 9.0 5 14 2 4 1 0.0 0.0 60.0 26
2819 0 31.0 4 9.0 4 8 0 4 0 0.0 0.0 40.0 39
14152 1 37.0 4 10.0 2 12 4 4 1 0.0 0.0 50.0 11
6512 rows × 13 columns
In [7]:
train.to_csv('train.csv', index=False, header=False)
validation.to_csv('validation.csv', index=False, header=False)
In [8]:
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 [10]:
import sagemaker
region = sagemaker.Session().boto_region_name
print("AWS Region: {}".format(region))

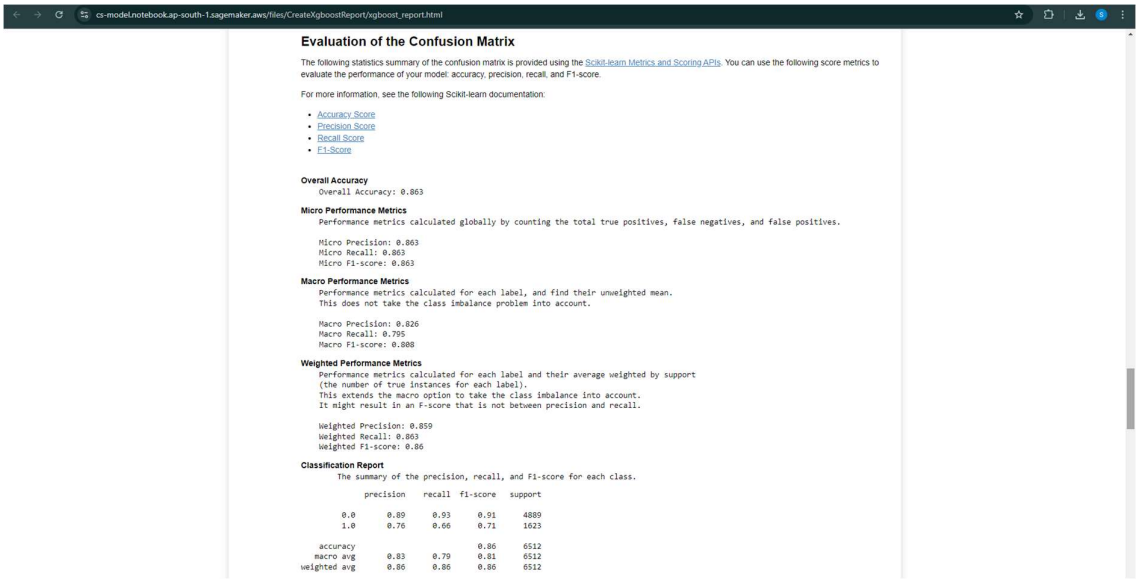

aws Services Search [Alt+S] Mumbai ChinmaySS
Amazon S3
Buckets
Access Grants
Access Points
Object Lambda Access Points
Multi-Region Access Points
Batch Operations
IAM Access Analyzer for S3
Block Public Access settings for this account
Storage Lens
Dashboards
Storage Lens groups
AWS Organizations settings
Feature spotlight 7
AWS Marketplace for S3
Amazon S3 > Buckets > sagemaker-ap-south-1-767397719307
sagemaker-ap-south-1-767397719307 Info
Objects Properties Permissions Metrics Management Access Points
Objects (1) Info
Copy S3 URI Copy URL Download Open Delete Actions Create folder Upload
Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more
Find objects by prefix
Name Type Last modified Size Storage class
demo-sagemaker-xgboost-adult-income-prediction/ Folder - - -
CloudShell Feedback
© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

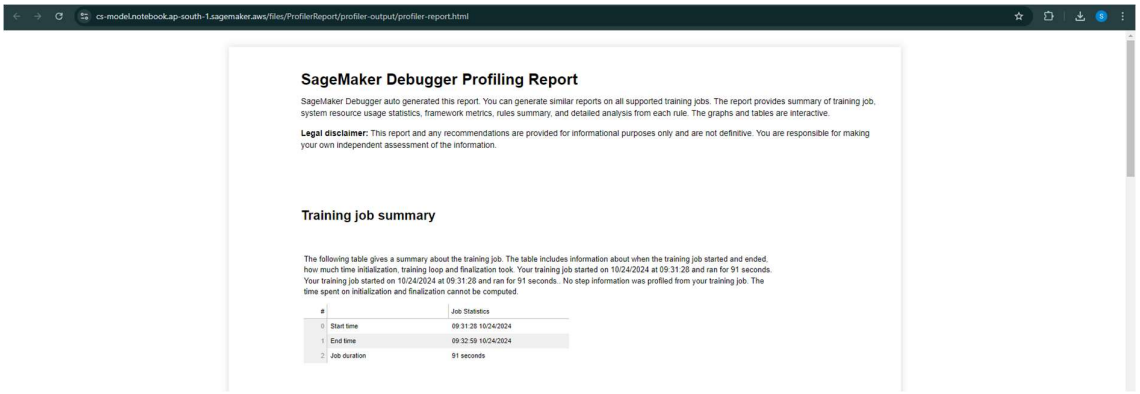

aws Services Search [Alt+S] Mumbai ChinmaySS
Amazon S3
Buckets
Access Grants
Access Points
Object Lambda Access Points
Multi-Region Access Points
Batch Operations
IAM Access Analyzer for S3
Block Public Access settings for this account
Storage Lens
Dashboards
Storage Lens groups
AWS Organizations settings
Feature spotlight 7
AWS Marketplace for S3
Amazon S3 > Buckets > sagemaker-ap-south-1-767397719307 > demo-sagemaker-xgboost-adult-income-prediction/ > data/
data/
Copy S3 URI
Objects Properties
Objects (2) Info
Copy S3 URI Copy URL Download Open Delete Actions Create folder Upload
Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more
Find objects by prefix
Name Type Last modified Size Storage class
train.csv csv October 24, 2024, 14:48:20 (UTC+05:30) 767.9 KB Standard
validation.csv csv October 24, 2024, 14:48:20 (UTC+05:30) 256.0 KB Standard
CloudShell Feedback
© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

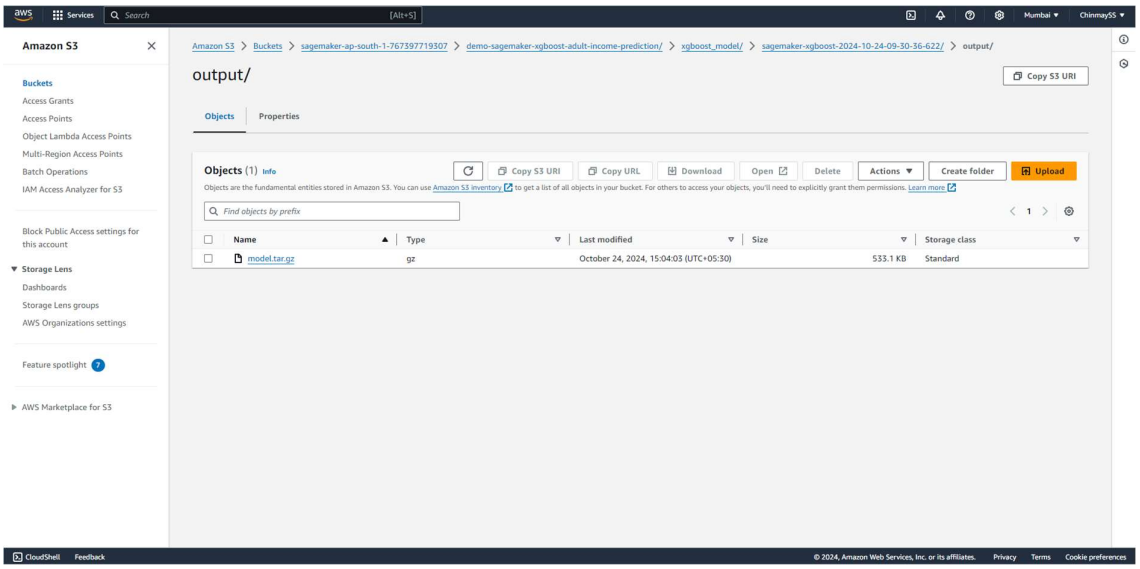

jupyter CScode Last Checkpoint: 19 minutes ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
conda_python3
Run
Code
nbdiff
)
In [14]:
xgb_model.fit({"train": train_input, "validation": validation_input}, wait=True)
[987]#011train-error:0.10847#011validation-error:0.13744
[988]#011train-error:0.10857#011validation-error:0.13744
[989]#011train-error:0.10857#011validation-error:0.13728
[990]#011train-error:0.10857#011validation-error:0.13728
[991]#011train-error:0.10841#011validation-error:0.13728
[992]#011train-error:0.10872#011validation-error:0.13698
[993]#011train-error:0.10867#011validation-error:0.13682
[994]#011train-error:0.10867#011validation-error:0.13652
[995]#011train-error:0.10867#011validation-error:0.13682
[996]#011train-error:0.10867#011validation-error:0.13682
[997]#011train-error:0.10826#011validation-error:0.13698
[998]#011train-error:0.10821#011validation-error:0.13667
[999]#011train-error:0.10821#011validation-error:0.13667
2024-10-24 09:34:12 Uploading - Uploading generated training model
2024-10-24 09:34:12 Completed - Training job completed
Training seconds: 160
Billable seconds: 160
In [15]:
rule_output_path = xgb_model.output_path + "/" + xgb_model.latest_training_job.job_name + "/rule-output"
! aws s3 ls {rule_output_path} --recursive
0-36-622/rule-output/ProfilerReport/profiler-output/profiler-reports/CPUBottleneck.json
2024-10-24 09:34:03 126 demo-sagemaker-xgboost-adult-income-prediction/xgboost_model/sagemaker-xgboost-2024-10-24-09-3
0-36-622/rule-output/ProfilerReport/profiler-output/profiler-reports/Dataloader.json
2024-10-24 09:34:03 127 demo-sagemaker-xgboost-adult-income-prediction/xgboost_model/sagemaker-xgboost-2024-10-24-09-3
0-36-622/rule-output/ProfilerReport/profiler-output/profiler-reports/GPUMemoryIncrease.json
2024-10-24 09:34:03 198 demo-sagemaker-xgboost-adult-income-prediction/xgboost_model/sagemaker-xgboost-2024-10-24-09-3
0-36-622/rule-output/ProfilerReport/profiler-output/profiler-reports/IOBottleneck.json
2024-10-24 09:34:03 119 demo-sagemaker-xgboost-adult-income-prediction/xgboost_model/sagemaker-xgboost-2024-10-24-09-3
0-36-622/rule-output/ProfilerReport/profiler-output/profiler-reports/LoadBalancing.json
2024-10-24 09:34:03 151 demo-sagemaker-xgboost-adult-income-prediction/xgboost_model/sagemaker-xgboost-2024-10-24-09-3
0-36-622/rule-output/ProfilerReport/profiler-output/profiler-reports/LowGPUUtilization.json
2024-10-24 09:34:03 178 demo-sagemaker-xgboost-adult-income-prediction/xgboost_model/sagemaker-xgboost-2024-10-24-09-3
0-36-622/rule-output/ProfilerReport/profiler-output/profiler-reports/MaxInitializationTime.json
2024-10-24 09:34:03 133 demo-sagemaker-xgboost-adult-income-prediction/xgboost_model/sagemaker-xgboost-2024-10-24-09-3
0-36-622/rule-output/ProfilerReport/profiler-output/profiler-reports/OverallFrameworkMetrics.json
2024-10-24 09:34:03 467 demo-sagemaker-xgboost-adult-income-prediction/xgboost_model/sagemaker-xgboost-2024-10-24-09-3
0-36-622/rule-output/ProfilerReport/profiler-output/profiler-reports/OverallSystemUsage.json
2024-10-24 09:34:03 156 demo-sagemaker-xgboost-adult-income-prediction/xgboost_model/sagemaker-xgboost-2024-10-24-09-3
0-36-622/rule-output/ProfilerReport/profiler-output/profiler-reports/StepOutlier.json


jupyter CScode Last Checkpoint: 20 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
conda_python3
Run
Code
nbdiff
-2024-10-24-09-30-36-622/rule-output/ProfilerReport/profiler-output/profiler-report.ipynb to ProfilerReport/profiler-output/profiler-report.ipynb
download: s3://sagemaker-ap-south-1-767397719307/demo-sagemaker-xgboost-adult-income-prediction/xgboost_model/sagemaker-xgboost-2024-10-24-09-30-36-622/rule-output/ProfilerReport/profiler-output/profiler-reports/StepOutlier.json to ProfilerReport/profiler-output/profiler-reports/StepOutlier.json
download: s3://sagemaker-ap-south-1-767397719307/demo-sagemaker-xgboost-adult-income-prediction/xgboost_model/sagemaker-xgboost-2024-10-24-09-30-36-622/rule-output/ProfilerReport/profiler-output/profiler-reports/OverallSystemUsage.json to ProfilerReport/profiler-output/profiler-reports/OverallSystemUsage.json
download: s3://sagemaker-ap-south-1-767397719307/demo-sagemaker-xgboost-adult-income-prediction/xgboost_model/sagemaker-xgboost-2024-10-24-09-30-36-622/rule-output/ProfilerReport/profiler-output/profiler-reports/MaxInitializationTime.json to ProfilerReport/profiler-output/profiler-reports/MaxInitializationTime.json
download: s3://sagemaker-ap-south-1-767397719307/demo-sagemaker-xgboost-adult-income-prediction/xgboost_model/sagemaker-xgboost-2024-10-24-09-30-36-622/rule-output/ProfilerReport/profiler-output/profiler-reports/GPUMemoryIncrease.json to ProfilerReport/profiler-output/profiler-reports/GPUMemoryIncrease.json
download: s3://sagemaker-ap-south-1-767397719307/demo-sagemaker-xgboost-adult-income-prediction/xgboost_model/sagemaker-xgboost-2024-10-24-09-30-36-622/rule-output/ProfilerReport/profiler-output/profiler-reports/OverallFrameworkMetrics.json to ProfilerReport/profiler-output/profiler-reports/OverallFrameworkMetrics.json
download: s3://sagemaker-ap-south-1-767397719307/demo-sagemaker-xgboost-adult-income-prediction/xgboost_model/sagemaker-xgboost-2024-10-24-09-30-36-622/rule-output/ProfilerReport/profiler-output/profiler-reports/IOBottleneck.json to ProfilerReport/profiler-output/profiler-reports/IOBottleneck.json
download: s3://sagemaker-ap-south-1-767397719307/demo-sagemaker-xgboost-adult-income-prediction/xgboost_model/sagemaker-xgboost-2024-10-24-09-30-36-622/rule-output/ProfilerReport/profiler-output/profiler-reports/LoadBalancing.json to ProfilerReport/profiler-output/profiler-reports/LoadBalancing.json
download: s3://sagemaker-ap-south-1-767397719307/demo-sagemaker-xgboost-adult-income-prediction/xgboost_model/sagemaker-xgboost-2024-10-24-09-30-36-622/rule-output/ProfilerReport/profiler-output/profiler-reports/Dataloader.json to ProfilerReport/profiler-output/profiler-reports/Dataloader.json
'Click link below to view the XGBoost Training report'
CreateXgboostReport/xgboost_report.html
In [17]:
profiler_report_name = [rule["RuleConfigurationName"]
for rule in xgb_model.latest_training_job.rule_job_summary()
if "Profiler" in rule["RuleConfigurationName"]][0]
profiler_report_name
display("Click link below to view the profiler report", FileLink(profiler_report_name+"/profiler-output/profiler-report.html"))
'Click link below to view the profiler report'
ProfilerReport/profiler-output/profiler-report.html
In [*]:
import sagemaker
from sagemaker.serializers import CSVSerializer
xgb_predictor=xgb_model.deploy(
initial_instance_count=1,
instance_type='ml.t2.medium',
serializer=CSVSerializer()
)


cs-model.notebook.ap-south-1.sagemaker.aws/files/CreateXgboostReport/xgboost_report.html
XGBoost Training Report by SageMaker Debugger
Created on 2024-10-24 09:34:03
The SageMaker Debugger CreateXgboostReport built-in rule auto-generates this report. This report provides a summary of the XGBoost model training evaluation results, insights of the model performance, and interactive graphs.
Legal disclaimer: In this report, plots and and recommendations are provided for informational purposes only and are not definitive. You are responsible for making your own independent assessment of the information.] For more information, see the following documentation:
Amazon SageMaker Developer Guide
If you want to use the notebook that generated this report, you need to install the following libraries:
SageMaker Debugger Client Library Github
The Bokeh Python Visualization Tool
In [3]:
# Parameters
path = "/opt/ml/processing/input/tensors"
plot_step = 995
s3_path = "s3://sagemaker-ap-south-1-767397719307/demo-sagemaker-xgboost-adult-income-prediction/xgboost_model/sagemaker-xgboost-2024-10-24-09-30-36-622/debug-output"
The following parameters are the default values auto-generated by the CreateXgboostReport built-in rule.
path (str) - The local path where Debugger has saved output tensors in the training container.
plot_step (int) - The step for which the rule has created the training report.
s3_path (str) - The S3 bucket URI where Debugger has saved the output tensors.
Table of Contents
Distribution of True Labels of the Dataset
Loss vs Step Graph
Feature Importance
Confusion Matrix
Evaluation of the Confusion Matrix
Accuracy Rate of Each Diagonal Element over Iteration
Receiver Operating Characteristic Curve
Distribution of Residuals at Last Saved Step


cs-model.notebook.ap-south-1.sagemaker.aws/files/CreateXgboostReport/xgboost_report.html
Evaluation of the Confusion Matrix
The following statistics summary of the confusion matrix is provided using the Scikit-learn Metrics and Scoring APIs. You can use the following score metrics to evaluate the performance of your model: accuracy, precision, recall, and F1-score.
For more information, see the following Scikit-learn documentation:
Accuracy Score
Precision Score
Recall Score
F1-Score
Overall Accuracy
Overall Accuracy: 0.863
Micro Performance Metrics
Performance metrics calculated globally by counting the total true positives, false negatives, and false positives.
Micro Precision: 0.863
Micro Recall: 0.863
Micro F1-score: 0.863
Macro Performance Metrics
Performance metrics calculated for each label, and find their unweighted mean.
This does not take the class imbalance problem into account.
Macro Precision: 0.826
Macro Recall: 0.795
Macro F1-score: 0.808
Weighted Performance Metrics
Performance metrics calculated for each label and their average weighted by support
(the number of true instances for each label).
This extends the macro option to take the class imbalance into account.
It might result in an F-score that is not between precision and recall.
Weighted Precision: 0.859
Weighted Recall: 0.863
Weighted F1-score: 0.86
Classification Report
The summary of the precision, recall, and F1-score for each class.
precision recall f1-score support
0.0 0.89 0.93 0.91 4889
1.0 0.76 0.66 0.71 1623
accuracy 0.86 6512
macro avg 0.83 0.79 0.81 6512
weighted avg 0.86 0.86 0.86 6512


cs-model.notebook.ap-south-1.sagemaker.aws/files/ProfilerReport/profiler-output/profiler-report.html
SageMaker Debugger Profiling Report
SageMaker Debugger auto generated this report. You can generate similar reports on all supported training jobs. The report provides summary of training job, system resource usage statistics, framework metrics, rules summary, and detailed analysis from each rule. The graphs and tables are interactive.
Legal disclaimer: This report and any recommendations are provided for informational purposes only and are not definitive. You are responsible for making your own independent assessment of the information.
Training job summary
The following table gives a summary about the training job. The table includes information about when the training job started and ended, how much time initialization, training loop and finalization took. Your training job started on 10/24/2024 at 09:31:28 and ran for 91 seconds. Your training job started on 10/24/2024 at 09:31:28 and ran for 91 seconds.. No step information was profiled from your training job. The time spent on initialization and finalization cannot be computed.
Job Statistics
0 Start time 09:31:28 10/24/2024
1 End time 09:32:59 10/24/2024
2 Job duration 91 seconds


aws
Services
Search
[Alt+S]
Mumbai
ChinmaySS
Amazon S3
Buckets
Access Grants
Access Points
Object Lambda Access Points
Multi-Region Access Points
Batch Operations
IAM Access Analyzer for S3
Block Public Access settings for this account
Storage Lens
Dashboards
Storage Lens groups
AWS Organizations settings
Feature spotlight
AWS Marketplace for S3
Amazon S3 > Buckets > sagemaker-ap-south-1-767397719307 > demo-sagemaker-xgboost-adult-income-prediction/ > xgboost_model/ > sagemaker-xgboost-2024-10-24-09-30-36-622/ > output/
output/
Copy S3 URI
Objects
Properties
Objects (1) Info
Copy S3 URI
Copy URL
Download
Open
Delete
Actions
Create folder
Upload
Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more
Find objects by prefix
Name
Type
Last modified
Size
Storage class
model.tar.gz
gz
October 24, 2024, 15:04:03 (UTC+05:30)
533.1 KB
Standard
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

jupyter CScode Last Checkpoint: 26 minutes ago (unsaved changes)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | conda_python3

Run Code nbdiff

```
----------!
```

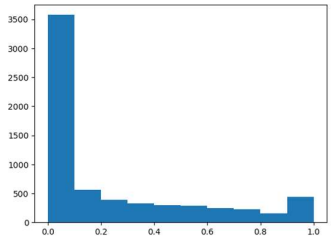
In [19]:
```
xgb_predictor.endpoint_name
```

Out[19]:
```
'sagemaker-xgboost-2024-10-24-09-35-33-384'
```

In [20]:
```
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 [21]:
```
import matplotlib.pyplot as plt

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



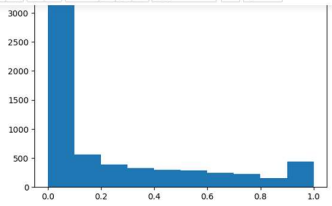
jupyter CScode Last Checkpoint: 26 minutes ago (unsaved changes)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | conda_python3

Run Code nbdiff



In [22]:
```
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
```

In [ ]: