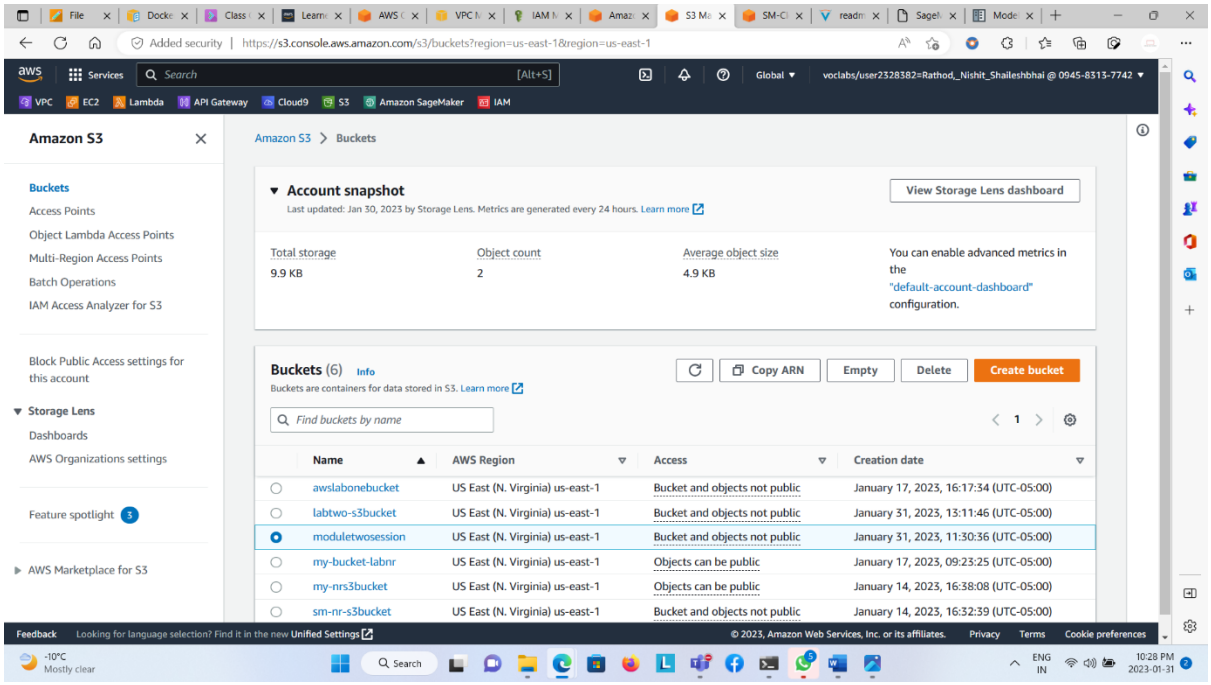
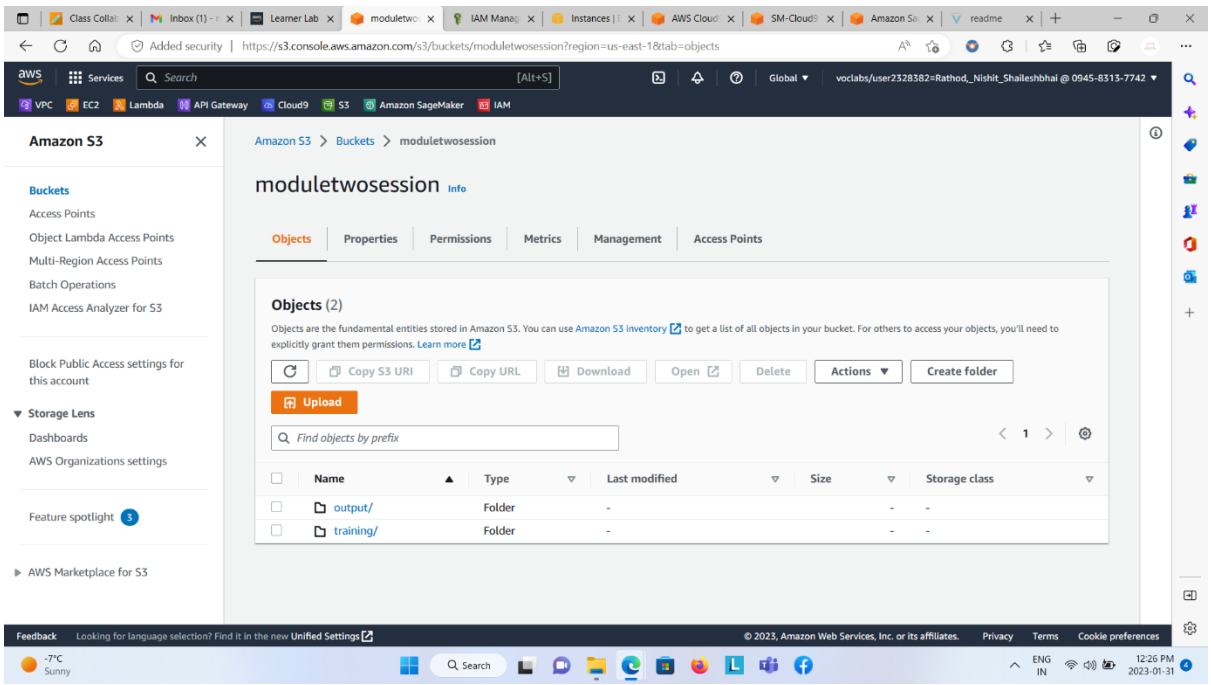
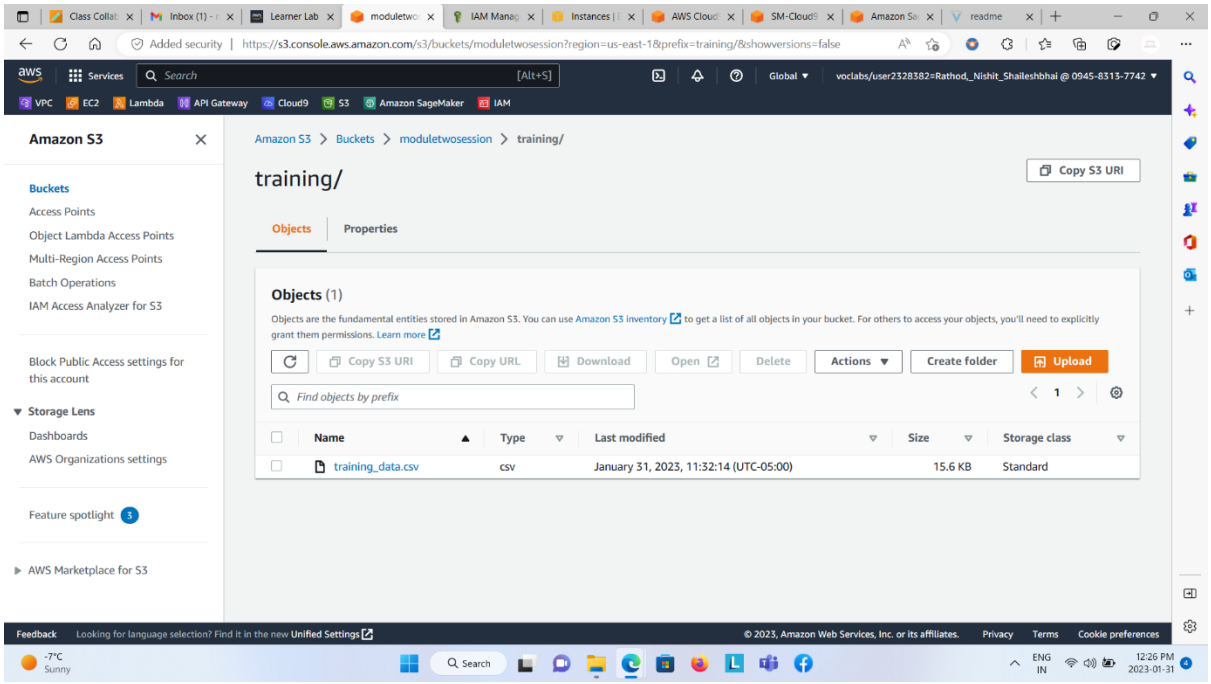
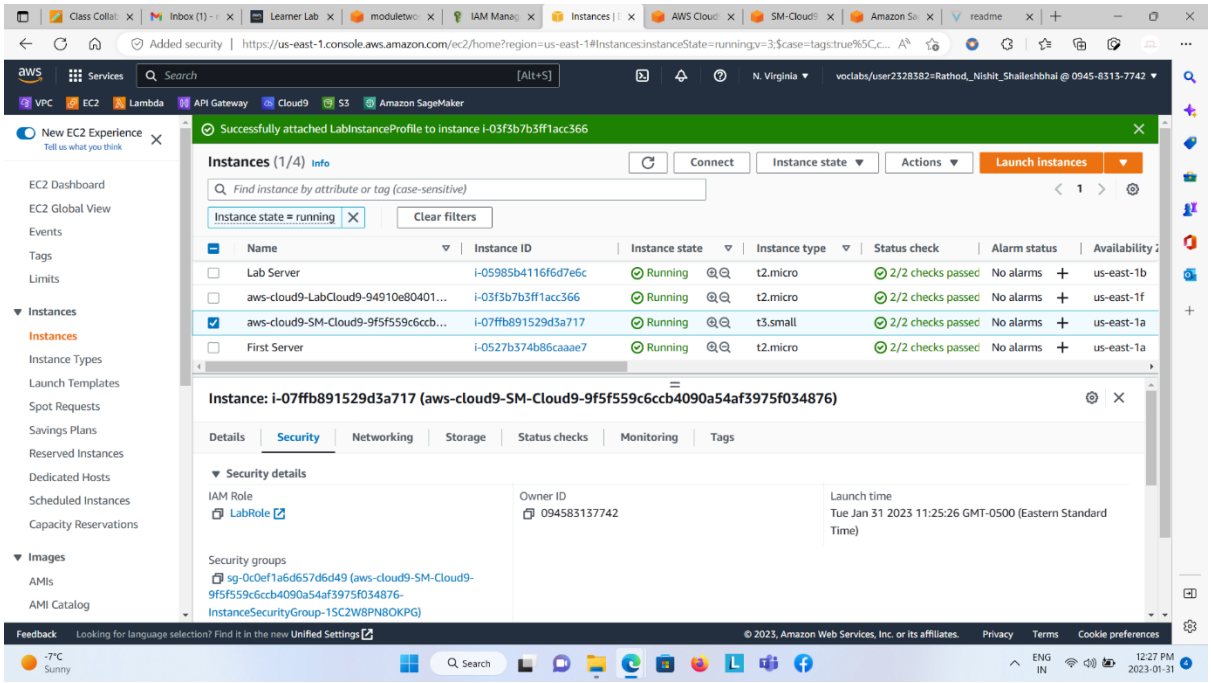
Assignment option (1): create a model

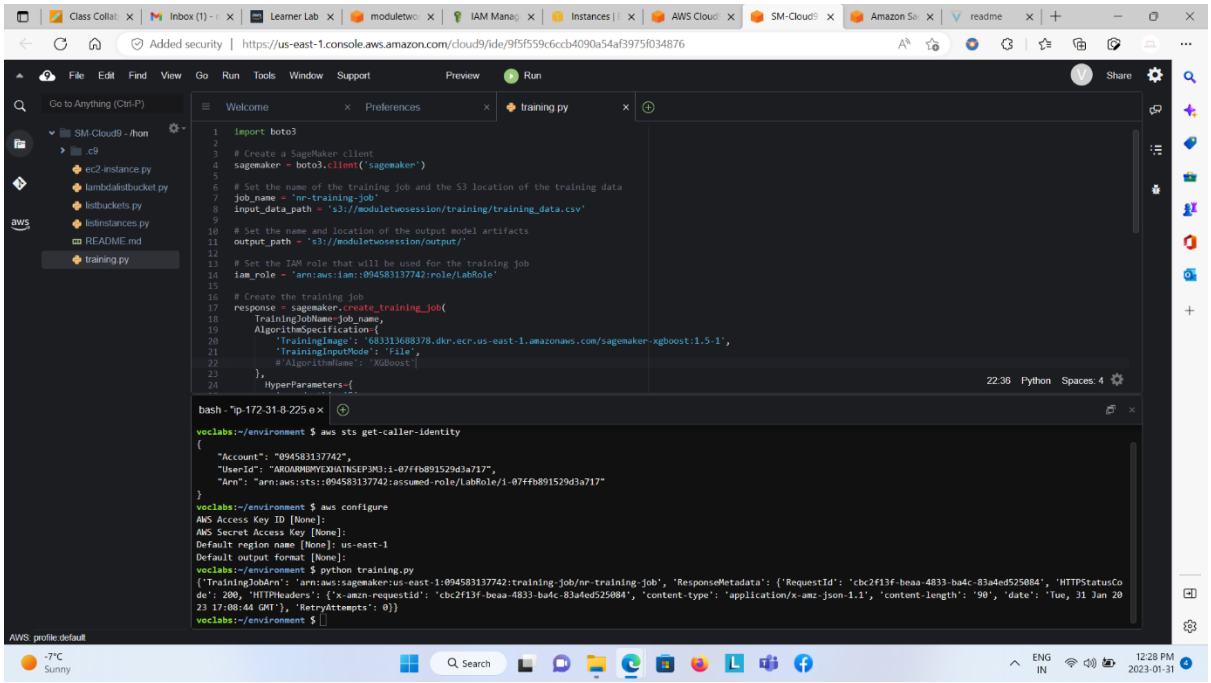
- Use the model.tar.gz model artifact file that has been created and create a SageMaker model.
- The skeleton code is provided to use as a start point but that code will not run as-is, you need to modify it.
- The file name is model.py
- (Optional) I have also provided another skeleton to deploy the code by create_endpoint.py file.

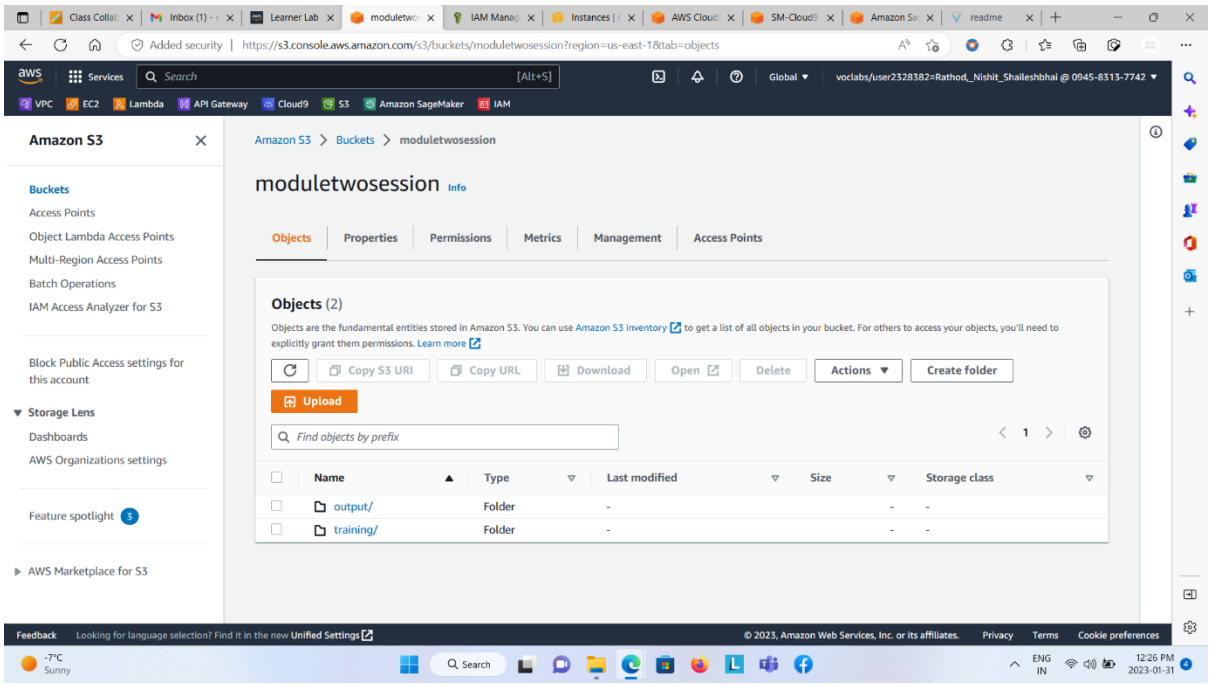


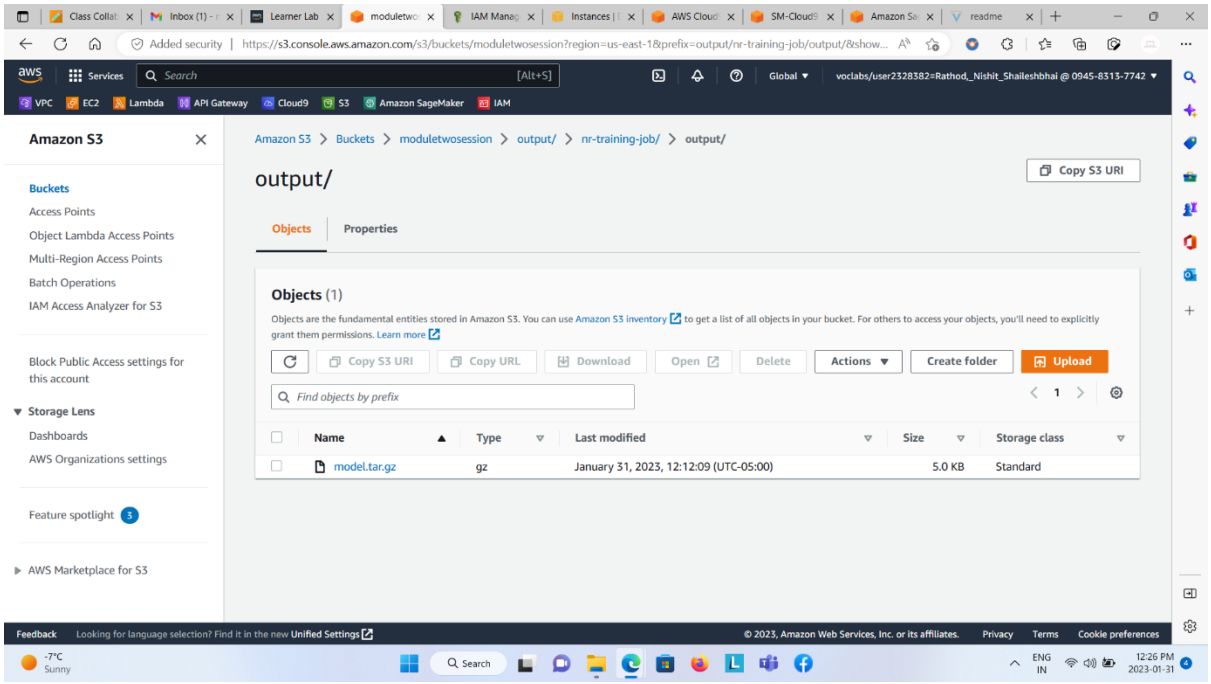


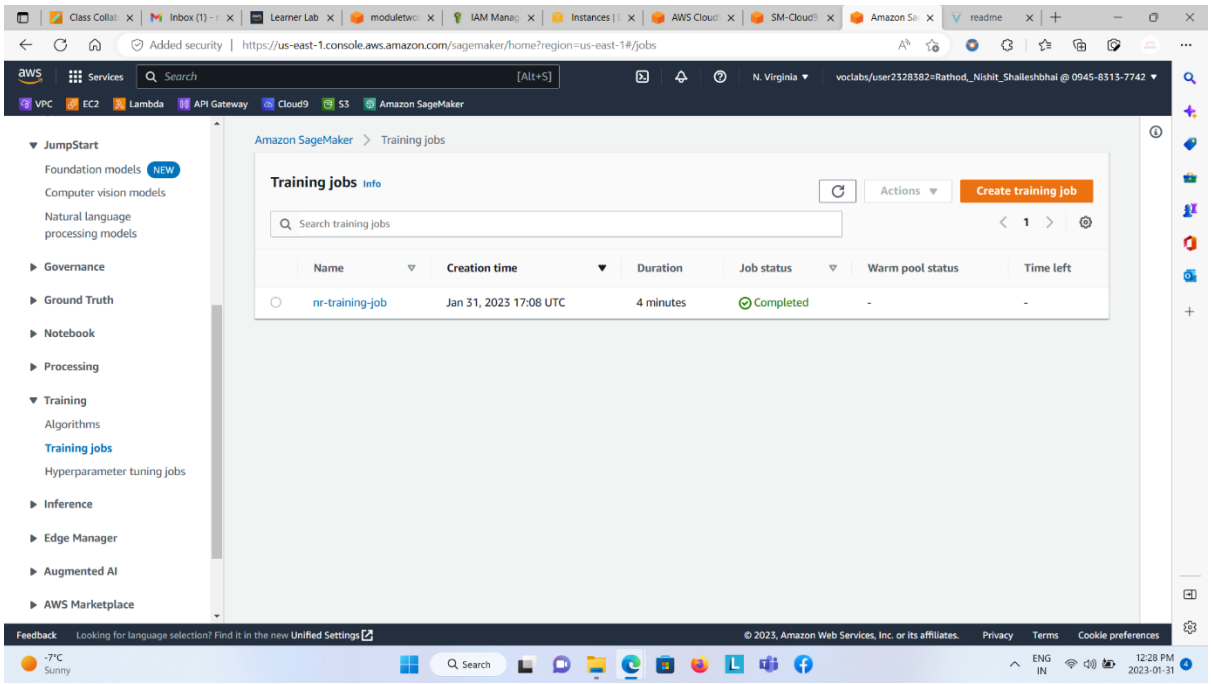

Amazon S3
Buckets
Access Points
Object Lambda Access Points
Multi-Region Access Points
Batch Operations
IAM Access Analyzer for S3
Block Public Access settings for this account
Storage Lens
Dashboards
AWS Organizations settings
Feature spotlight
AWS Marketplace for S3
Amazon S3 > Buckets > moduletwosession > training/
training/
Copy S3 URI
Objects
Properties
Objects (1)
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
Copy S3 URI
Copy URL
Download
Open
Delete
Actions
Create folder
Upload
Find objects by prefix
Name
Type
Last modified
Size
Storage class
training_data.csv
csv
January 31, 2023, 11:32:14 (UTC-05:00)
15.6 KB
Standard

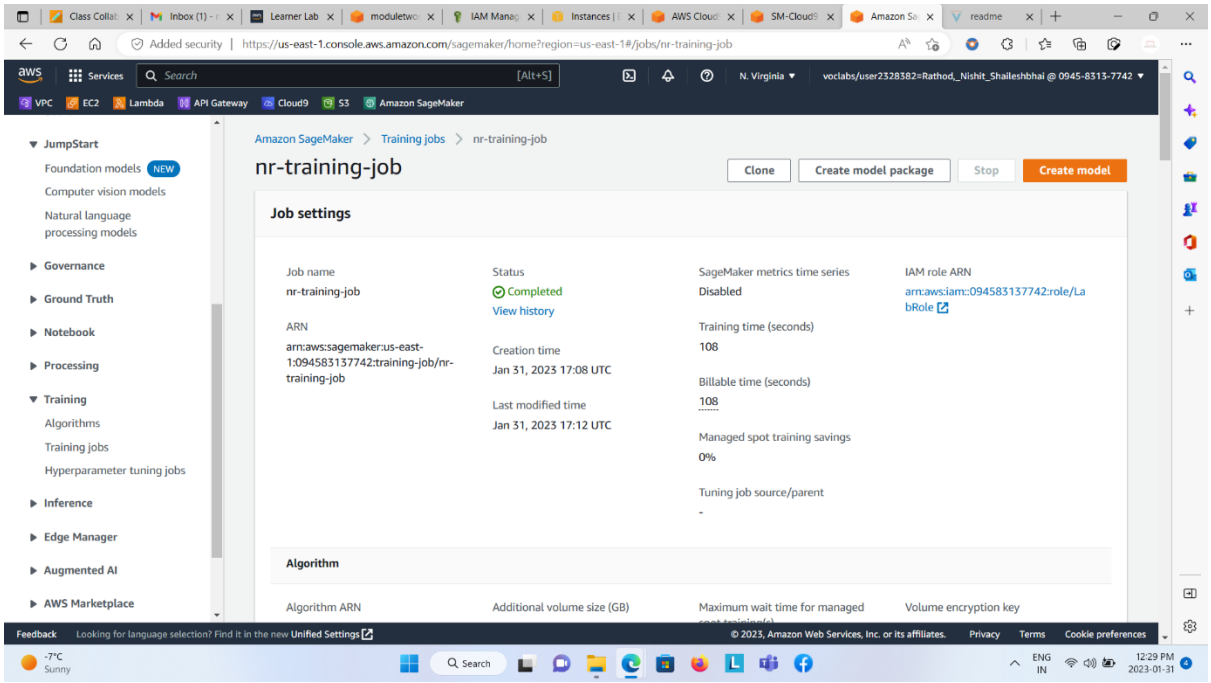

Successfully attached LabInstanceProfile to instance i-03f3b7b3ff1acc366
Instances (1/4) Info
Connect
Instance state
Actions
Launch instances
Find instance by attribute or tag (case-sensitive)
Instance state = running
Clear filters
Name
Instance ID
Instance state
Instance type
Status check
Alarm status
Availability
Lab Server
i-05985b4116f6d7e6c
Running
t2.micro
2/2 checks passed
No alarms
us-east-1b
aws-cloud9-LabCloud9-94910e80401...
i-03f3b7b3ff1acc366
Running
t2.micro
2/2 checks passed
No alarms
us-east-1f
aws-cloud9-SM-Cloud9-9f5f559c6ccb...
i-07ffb891529d3a717
Running
t3.small
2/2 checks passed
No alarms
us-east-1a
First Server
i-0527b374b86caaae7
Running
t2.micro
2/2 checks passed
No alarms
us-east-1a
Instance: i-07ffb891529d3a717 (aws-cloud9-SM-Cloud9-9f5f559c6ccb4090a54af3975f034876)
Details
Security
Networking
Storage
Status checks
Monitoring
Tags
Security details
IAM Role
LabRole
Owner ID
094583137742
Launch time
Tue Jan 31 2023 11:25:26 GMT-0500 (Eastern Standard Time)
Security groups
sg-0c0ef1a6d657d6d49 (aws-cloud9-SM-Cloud9-9f5f559c6ccb4090a54af3975f034876-InstanceSecurityGroup-1SC2W8PN8OKPG)

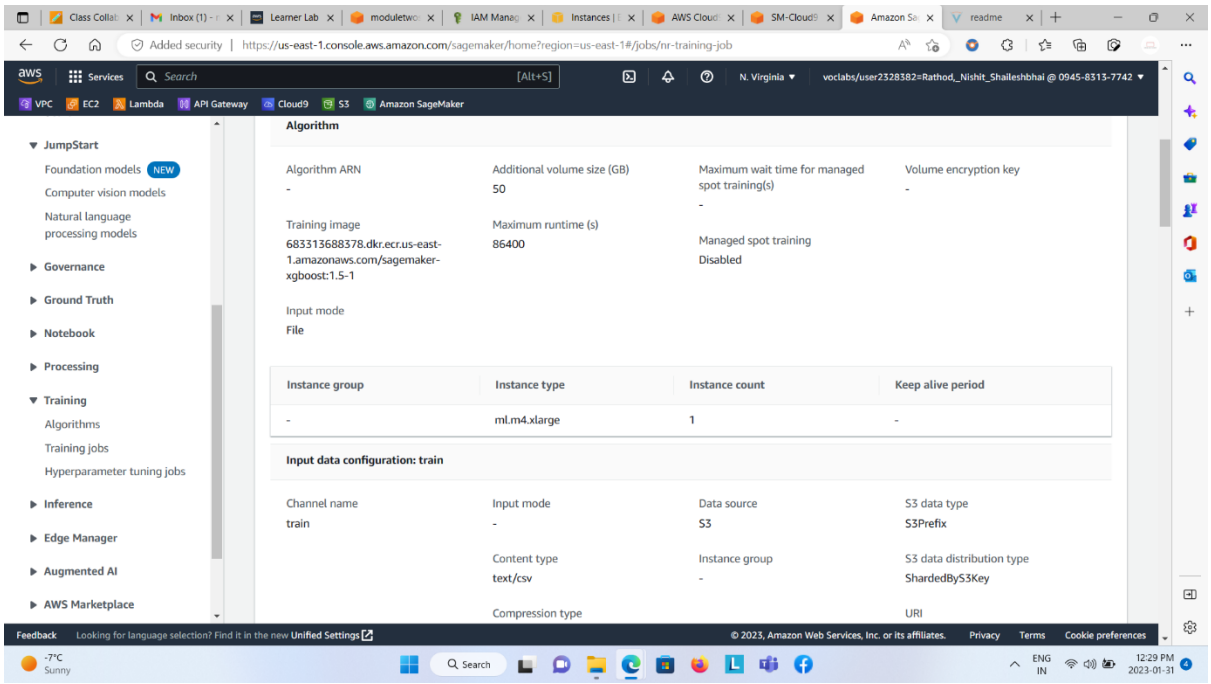

https://us-east-1.console.aws.amazon.com/cloud9/ide/9f5f559c6ccb4090a54af3975f034876
File Edit Find View Go Run Tools Window Support Preview Run Share
Welcome Preferences training.py
import boto3
# Create a SageMaker client
sagemaker = boto3.client('sagemaker')
# Set the name of the training job and the S3 location of the training data
job_name = 'nr-training-job'
input_data_path = 's3://moduletwosession/training/training_data.csv'
# Set the name and location of the output model artifacts
output_path = 's3://moduletwosession/output/'
# Set the IAM role that will be used for the training job
iam_role = 'arn:aws:iam::094583137742:role/LabRole'
# Create the training job
response = sagemaker.create_training_job(
TrainingJobName=job_name,
AlgorithmSpecification={
'TrainingImage': '683313688378.dkr.ecr.us-east-1.amazonaws.com/sagemaker-xgboost:1.5-1',
'TrainingInputMode': 'File',
'AlgorithmName': 'XGBoost'
},
HyperParameters={
22:36 Python Spaces: 4
bash - "ip-172-31-8-225.e
voclabs:~/environment $ aws sts get-caller-identity
{
"Account": "094583137742",
"UserId": "AROAAMBMYEJHATNSEP3MJ:i-07ffb891529d3a717",
"Arn": "arn:aws:sts::094583137742:assumed-role/LabRole/i-07ffb891529d3a717"
}
voclabs:~/environment $ aws configure
AWS Access Key ID [None]:
AWS Secret Access Key [None]:
Default region name [None]: us-east-1
Default output format [None]:
voclabs:~/environment $ python training.py
{'TrainingJobArn': 'arn:aws:sagemaker:us-east-1:094583137742:training-job/nr-training-job', 'ResponseMetadata': {'RequestId': 'cbc2f13f-beaa-4833-ba4c-83a4ed525084', 'HTTPStatusCode': 200, 'HTTPHeaders': {'x-amzn-requestid': 'cbc2f13f-beaa-4833-ba4c-83a4ed525084', 'content-type': 'application/x-amz-json-1.1', 'content-length': '90', 'date': 'Tue, 31 Jan 2023 17:08:44 GMT'}, 'RetryAttempts': 0}}
voclabs:~/environment $
AWS: profile:default
-7°C Sunny
Search
ENG IN
12:28 PM 2023-01-31

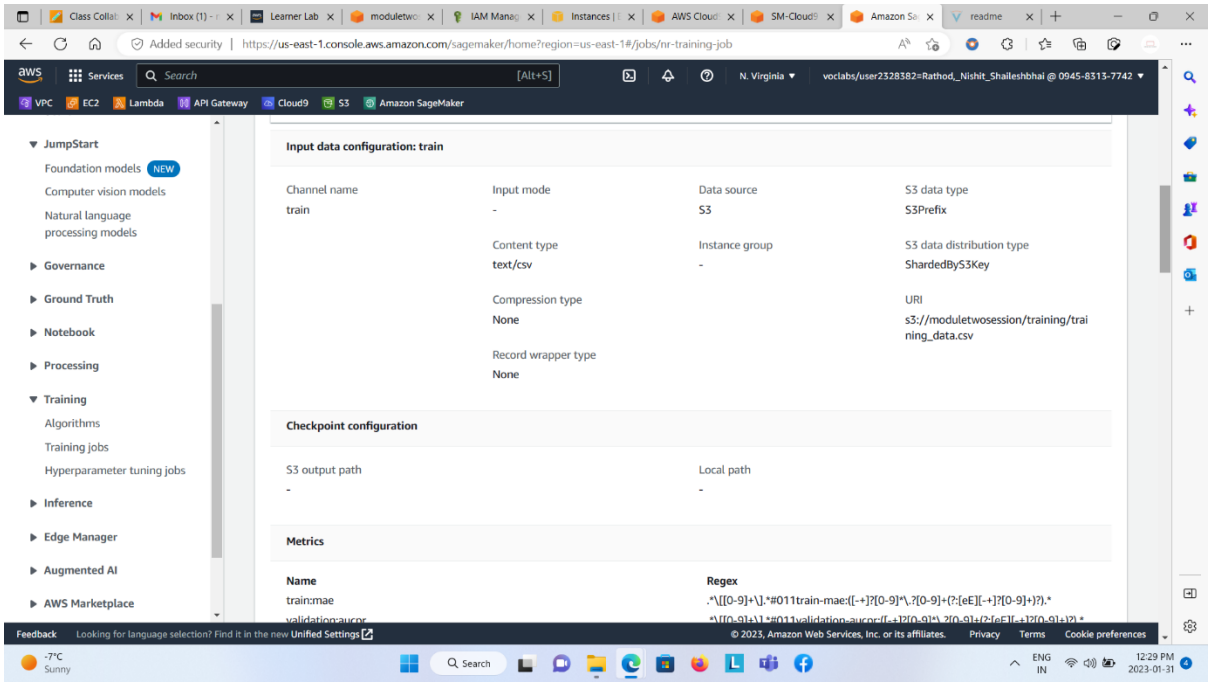

https://s3.console.aws.amazon.com/s3/buckets/moduletwosession?region=us-east-1&tab=objects
Services Search [Alt+S] Global voclabs/user2328382=Rathod,_Nishit_Shaileshbhai @ 0945-8313-7742
VPC EC2 Lambda API Gateway Cloud9 S3 Amazon SageMaker IAM
Amazon S3
Buckets
Access Points
Object Lambda Access Points
Multi-Region Access Points
Batch Operations
IAM Access Analyzer for S3
Block Public Access settings for this account
Storage Lens
Dashboards
AWS Organizations settings
Feature spotlight
AWS Marketplace for S3
Amazon S3 > Buckets > moduletwosession
moduletwosession Info
Objects Properties Permissions Metrics Management Access Points
Objects (2)
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
Copy S3 URI Copy URL Download Open Delete Actions Create folder
Upload
Find objects by prefix
Name Type Last modified Size Storage class
output/ Folder - - -
training/ Folder - - -
Feedback Looking for language selection? Find it in the new Unified Settings
© 2023, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences
-7°C Sunny
Search
ENG IN
12:26 PM 2023-01-31

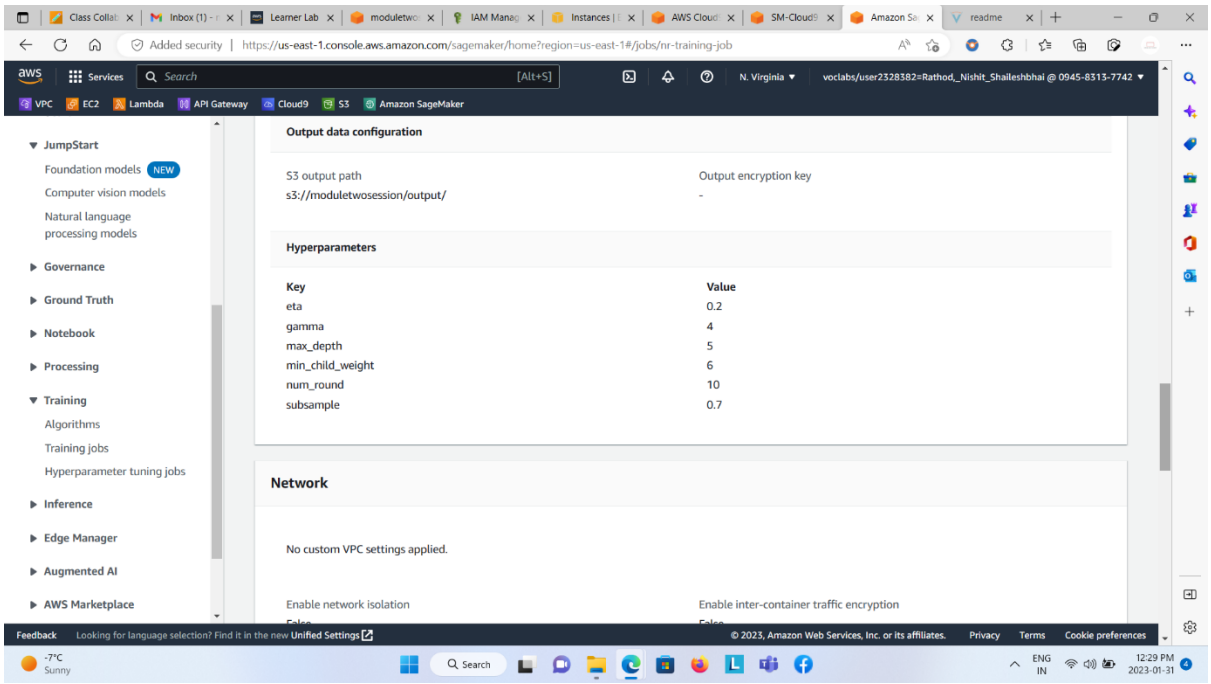

Amazon S3
Buckets
Access Points
Object Lambda Access Points
Multi-Region Access Points
Batch Operations
IAM Access Analyzer for S3
Block Public Access settings for this account
Storage Lens
Dashboards
AWS Organizations settings
Feature spotlight
AWS Marketplace for S3
Amazon S3 > Buckets > moduletwosession > output/ > nr-training-job/ > output/
output/
Copy S3 URI
Objects
Properties
Objects (1)
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
Copy S3 URI
Copy URL
Download
Open
Delete
Actions
Create folder
Upload
Find objects by prefix
Name
Type
Last modified
Size
Storage class
model.tar.gz
gz
January 31, 2023, 12:12:09 (UTC-05:00)
5.0 KB
Standard
Feedback
Looking for language selection? Find it in the new Unified Settings
© 2023, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

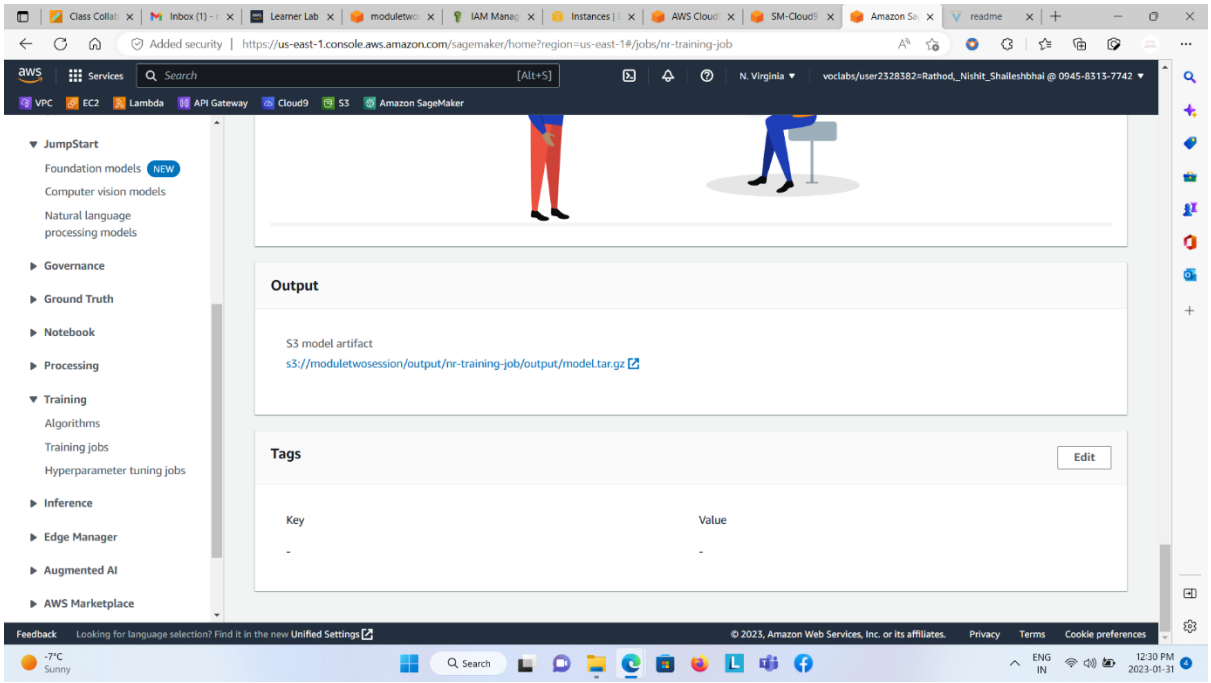

JumpStart
Foundation models
NEW
Computer vision models
Natural language processing models
Governance
Ground Truth
Notebook
Processing
Training
Algorithms
Training jobs
Hyperparameter tuning jobs
Inference
Edge Manager
Augmented AI
AWS Marketplace
Amazon SageMaker > Training jobs
Training jobs Info
Actions
Create training job
Search training jobs
Name
Creation time
Duration
Job status
Warm pool status
Time left
nr-training-job
Jan 31, 2023 17:08 UTC
4 minutes
Completed
Feedback
Looking for language selection? Find it in the new Unified Settings
© 2023, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

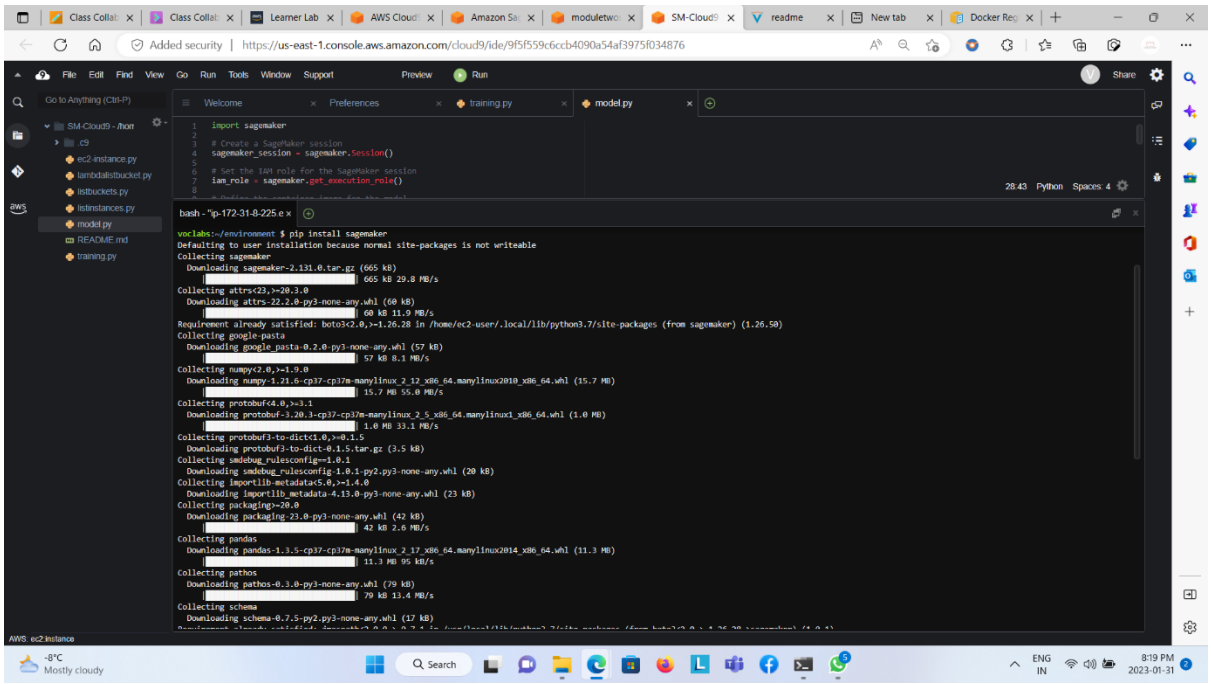

Amazon SageMaker > Training jobs > nr-training-job
nr-training-job
Clone
Create model package
Stop
Create model
Job settings
Job name
nr-training-job
Status
Completed
View history
SageMaker metrics time series
Disabled
IAM role ARN
arn:aws:iam::094583137742:role/LabRole
ARN
arn:aws:sagemaker:us-east-1:094583137742:training-job/nr-training-job
Creation time
Jan 31, 2023 17:08 UTC
Training time (seconds)
108
Last modified time
Jan 31, 2023 17:12 UTC
Billable time (seconds)
108
Managed spot training savings
0%
Tuning job source/parent
-
Algorithm
Algorithm ARN
Additional volume size (GB)
Maximum wait time for managed
Volume encryption key

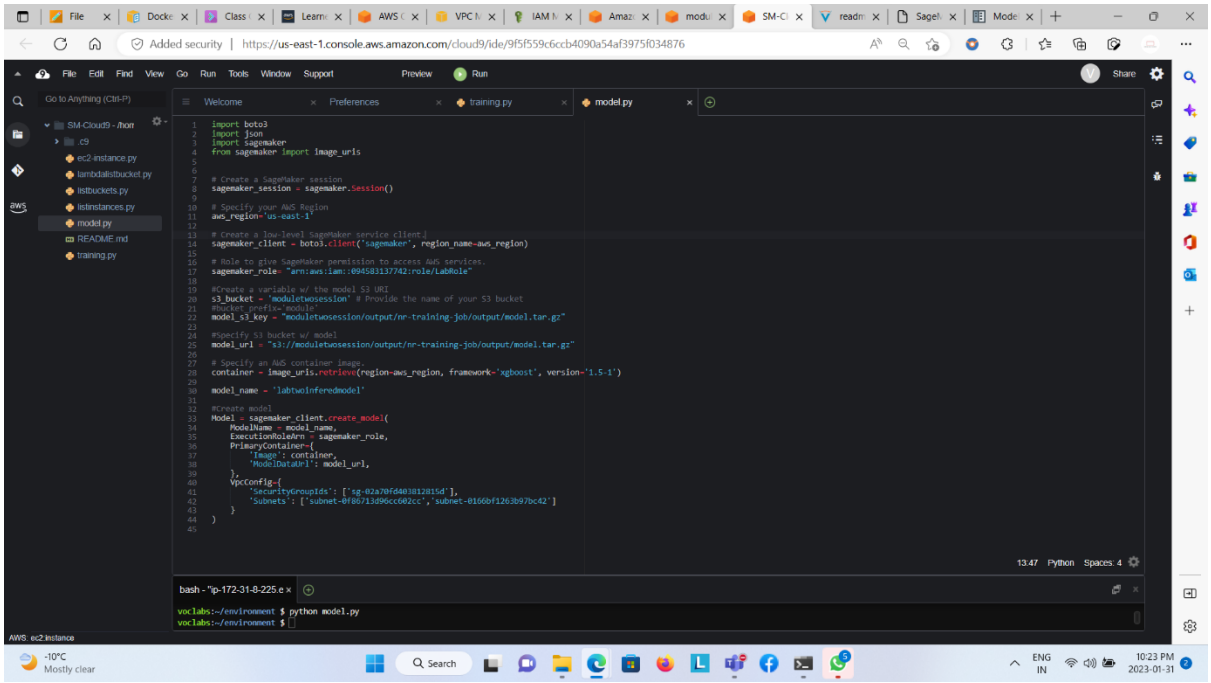

Algorithm
Algorithm ARN
-
Additional volume size (GB)
50
Maximum wait time for managed spot training(s)
-
Volume encryption key
-
Training image
683313688378.dkr.ecr.us-east-1.amazonaws.com/sagemaker-xgboost:1.5-1
Maximum runtime (s)
86400
Managed spot training
Disabled
Input mode
File
Instance group
Instance type
Instance count
Keep alive period
-
ml.m4.xlarge
1
-
Input data configuration: train
Channel name
train
Input mode
-
Data source
S3
S3 data type
S3Prefix
Content type
text/csv
Instance group
-
S3 data distribution type
ShardedByS3Key
Compression type
URI

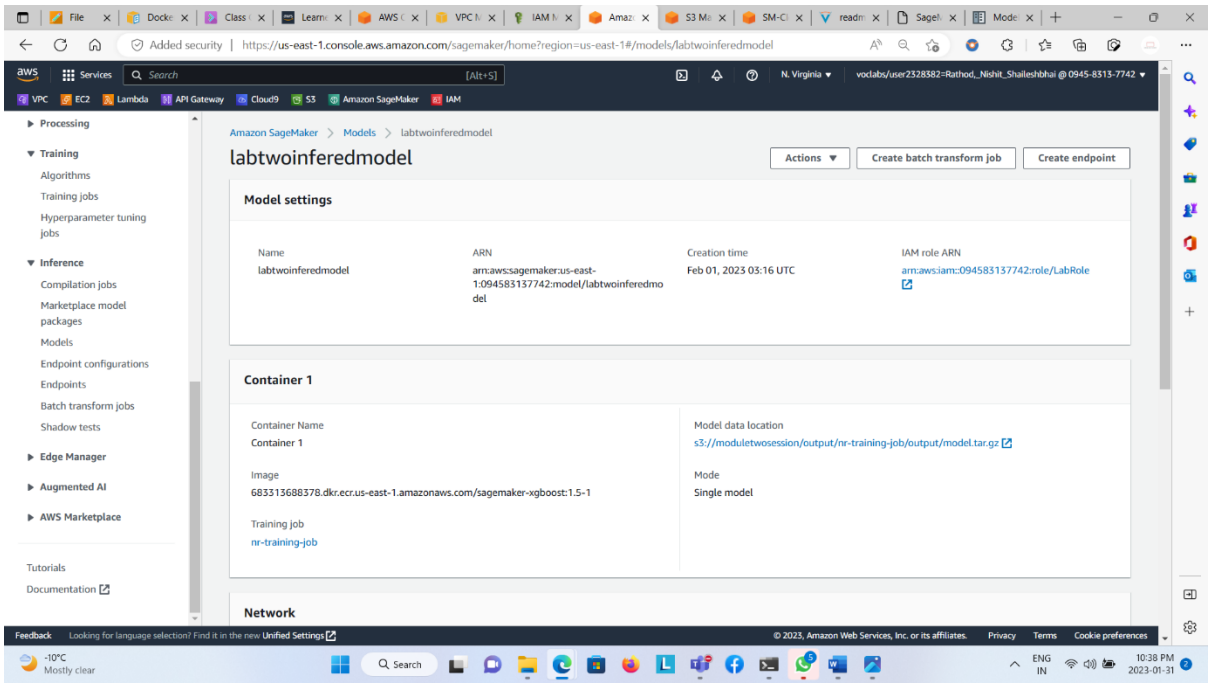

Input data configuration: train
Channel name
train
Input mode
-
Data source
S3
S3 data type
S3Prefix
Content type
text/csv
Instance group
-
S3 data distribution type
ShardedByS3Key
Compression type
None
URI
s3://moduletwosession/training/training_data.csv
Record wrapper type
None
Checkpoint configuration
S3 output path
-
Local path
-
Metrics
Name
train:mae
Regex
.*\[[0-9]+\].*#011train-mae:([-+]?[0-9]*\.?[0-9]+(?:[eE][-+]?[0-9]+)?).*

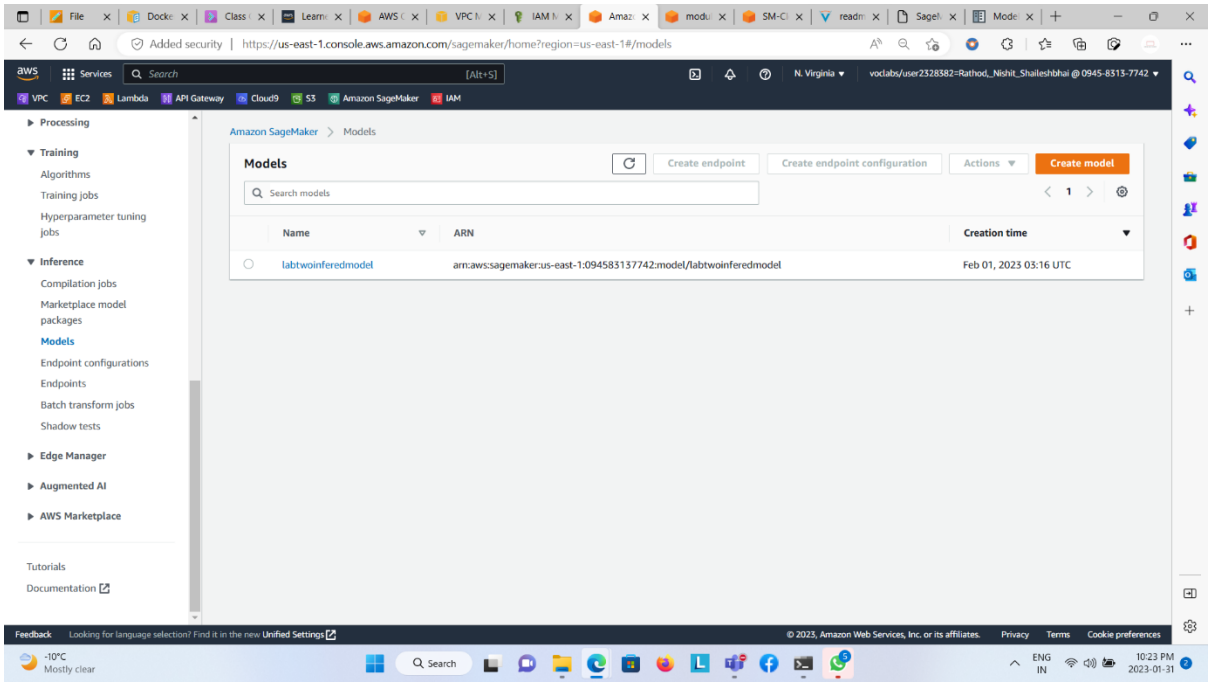

Output data configuration
S3 output path
s3://moduletwosession/output/
Output encryption key
-
Hyperparameters
Key
Value
eta
0.2
gamma
4
max_depth
5
min_child_weight
6
num_round
10
subsample
0.7
Network
No custom VPC settings applied.
Enable network isolation
Enable inter-container traffic encryption

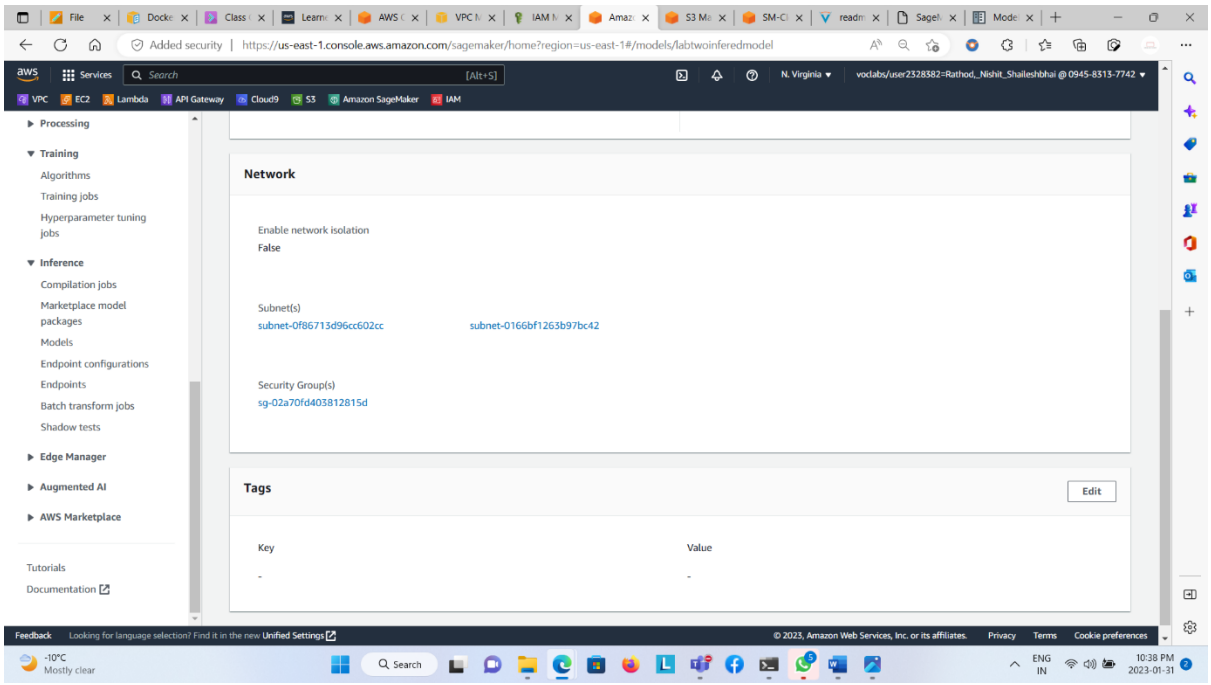

Output
S3 model artifact
s3://moduletwosession/output/nr-training-job/output/model.tar.gz
Tags
Edit
Key
Value
JumpStart
Foundation models
NEW
Computer vision models
Natural language processing models
Governance
Ground Truth
Notebook
Processing
Training
Algorithms
Training jobs
Hyperparameter tuning jobs
Inference
Edge Manager
Augmented AI
AWS Marketplace


import sagemaker
# Create a SageMaker session
sagemaker_session = sagemaker.Session()
# Set the IAM role for the SageMaker session
iam_role = sagemaker.get_execution_role()
voclabs:~/environment $ pip install sagemaker
Defaulting to user installation because normal site-packages is not writeable
Collecting sagemaker
Downloading sagemaker-2.131.0.tar.gz (665 kB)
Collecting attrs<23,>=20.3.0
Downloading attrs-22.2.0-py3-none-any.whl (60 kB)
Requirement already satisfied: boto3<2.0,>=1.26.28 in /home/ec2-user/.local/lib/python3.7/site-packages (from sagemaker) (1.26.50)
Collecting google-pasta
Downloading google_pasta-0.2.0-py3-none-any.whl (57 kB)
Collecting numpy<2.0,>=1.9.0
Downloading numpy-1.21.6-cp37-cp37m-manylinux_2_12_x86_64.manylinux2010_x86_64.whl (15.7 MB)
Collecting protobuf<4.0,>=3.1
Downloading protobuf-3.20.3-cp37-cp37m-manylinux_2_5_x86_64.manylinux1_x86_64.whl (1.0 MB)
Collecting protobuf3-to-dict<1.0,>=0.1.5
Downloading protobuf3-to-dict-0.1.5.tar.gz (3.5 kB)
Collecting smdebug_rulesconfig==1.0.1
Downloading smdebug_rulesconfig-1.0.1-py2.py3-none-any.whl (20 kB)
Collecting importlib-metadata<5.0,>=1.4.0
Downloading importlib_metadata-4.13.0-py3-none-any.whl (23 kB)
Collecting packaging>=20.0
Downloading packaging-23.0-py3-none-any.whl (42 kB)
Collecting pandas
Downloading pandas-1.3.5-cp37-cp37m-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (11.3 MB)
Collecting pathos
Downloading pathos-0.3.0-py3-none-any.whl (79 kB)
Collecting schema
Downloading schema-0.7.5-py2.py3-none-any.whl (17 kB)


File Edit Find View Go Run Tools Window Support Preview Run Share
Welcome Preferences training.py model.py
import boto3
import json
import sagemaker
from sagemaker import image_uris
# Create a SageMaker session
sagemaker_session = sagemaker.Session()
# Specify your AWS Region
aws_region='us-east-1'
# Create a low-level SageMaker service client.
sagemaker_client = boto3.client('sagemaker', region_name=aws_region)
# Role to give SageMaker permission to access AWS services.
sagemaker_role= "arn:aws:iam::094583137742:role/LabRole"
#Create a variable w/ the model S3 URI
s3_bucket = 'moduletwosession' # Provide the name of your S3 bucket
#bucket_prefix='module'
model_s3_key = "moduletwosession/output/nr-training-job/output/model.tar.gz"
#Specify S3 bucket w/ model
model_url = "s3://moduletwosession/output/nr-training-job/output/model.tar.gz"
# Specify an AWS container image.
container = image_uris.retrieve(region=aws_region, framework='xgboost', version='1.5-1')
model_name = 'labtwoinferedmodel'
#Create model
Model = sagemaker_client.create_model(
ModelName = model_name,
ExecutionRoleArn = sagemaker_role,
PrimaryContainer={
'Image': container,
'ModelDataUrl': model_url,
},
VpcConfig={
'SecurityGroupIds': ['sg-02a70fd4038128154'],
'Subnets': ['subnet-0f85713d96cc602cc','subnet-0166bf1263b97bc42']
}
)
bash - "ip-172-31-8-225.e
voclabs:~/environment $ python model.py
voclabs:~/environment $
AWS: ec2.instance
-10°C Mostly clear
Search
ENG IN
10:23 PM 2023-01-31


Added security https://us-east-1.console.aws.amazon.com/sagemaker/home?region=us-east-1#/models/labtwoinferedmodel
Services Search [Alt+S]
N. Virginia
voclabs/user2328382=Rathod,_Nishit_Shaileshbhai @ 0945-8313-7742
VPC EC2 Lambda API Gateway Cloud9 S3 Amazon SageMaker IAM
Processing
Training
Algorithms
Training jobs
Hyperparameter tuning jobs
Inference
Compilation jobs
Marketplace model packages
Models
Endpoint configurations
Endpoints
Batch transform jobs
Shadow tests
Edge Manager
Augmented AI
AWS Marketplace
Tutorials
Documentation
Amazon SageMaker > Models > labtwoinferedmodel
labtwoinferedmodel
Actions
Create batch transform job
Create endpoint
Model settings
Name
labtwoinferedmodel
ARN
arn:aws:sagemaker:us-east-1:094583137742:model/labtwoinferedmodel
Creation time
Feb 01, 2023 03:16 UTC
IAM role ARN
arn:aws:iam::094583137742:role/LabRole
Container 1
Container Name
Container 1
Image
683313688378.dkr.ecr.us-east-1.amazonaws.com/sagemaker-xgboost:1.5-1
Training job
nr-training-job
Model data location
s3://moduletwosession/output/nr-training-job/output/model.tar.gz
Mode
Single model
Network
Feedback
Looking for language selection? Find it in the new Unified Settings
© 2023, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences
-10°C Mostly clear
Search
ENG IN
10:38 PM 2023-01-31


Amazon SageMaker
Models
Create endpoint
Create endpoint configuration
Actions
Create model
Search models
Name
ARN
Creation time
labtwoinferedmodel
arn:aws:sagemaker:us-east-1:094583137742:model/labtwoinferedmodel
Feb 01, 2023 03:16 UTC


Network
Enable network isolation
False
Subnet(s)
subnet-0f86713d96cc602cc
subnet-0166bf1263b97bc42
Security Group(s)
sg-02a70fd403812815d
Tags
Edit
Key
Value