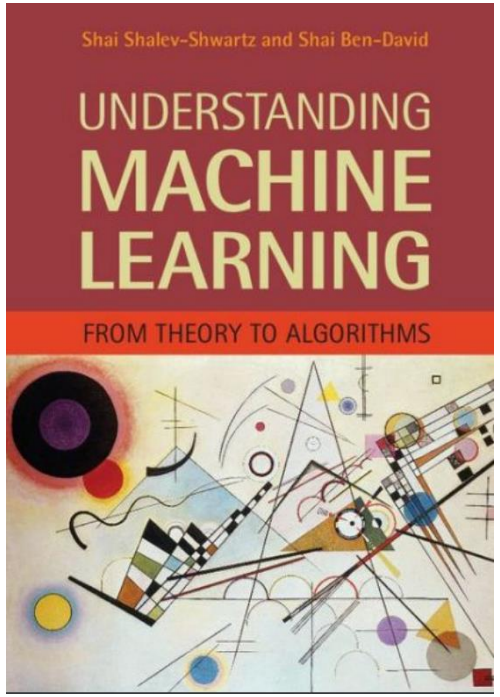
# Cloud Computing Linear Regression-AWS sagemaker

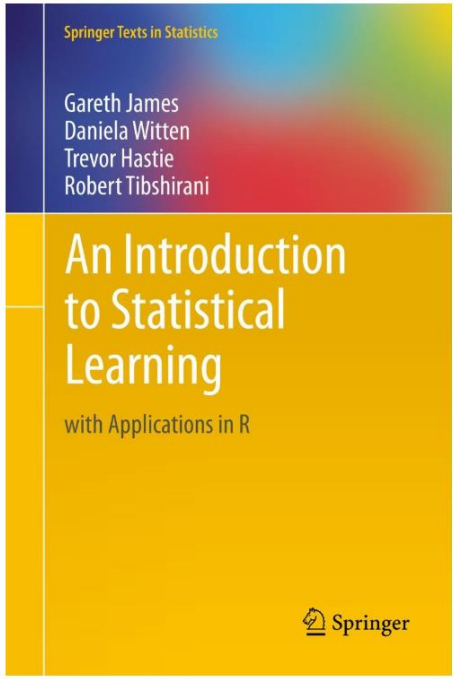
Farid Afzali, Ph.D., P.Eng.

# Exploratory Data Analysis

Additional Resources, Page #123:
http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/understanding-machine-learning-theory-algorithms.pdf

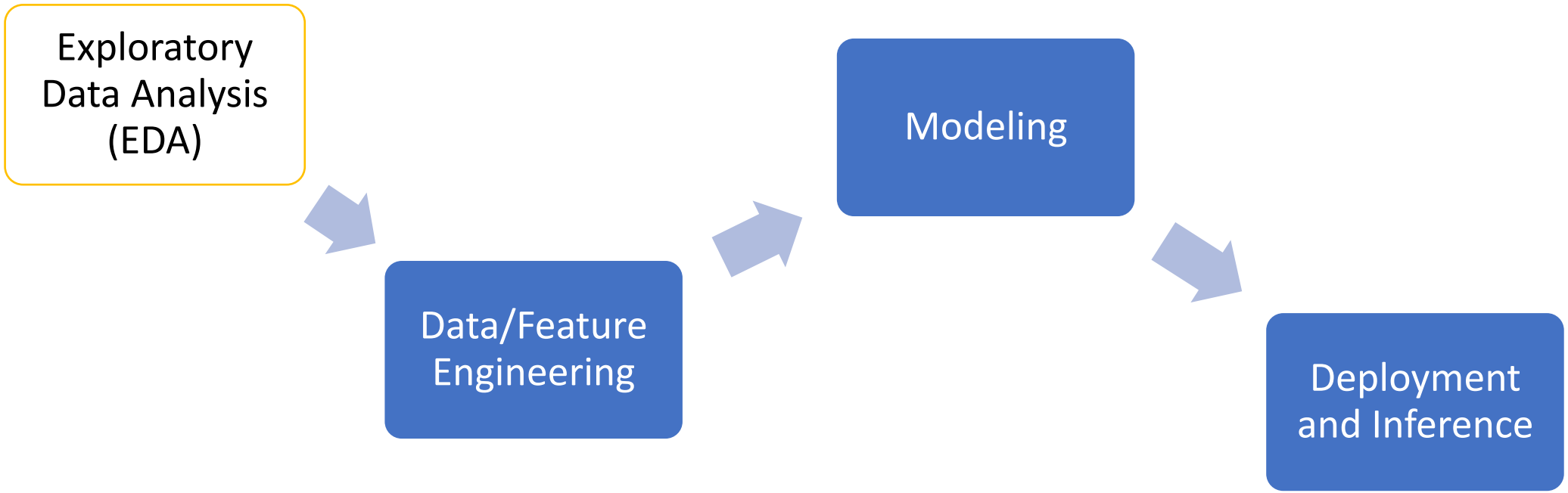



Additional Resources, Page #61:
http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Seventh%20Printing.pdf




Dr. Ryan Ahmed

# EDA


Exploratory Data Analysis (EDA)
Data/Feature Engineering
Modeling
Deployment and Inference

# University admission project

- **Goal**: The project's objective is to build, train, test, and deploy a machine learning model that can predict the likelihood of university admission based on a student's profile.
- **Tool**: AWS SageMaker is mentioned as the tool for launching a training job from the Management Console, which is a part of AWS that allows you to build, train, and deploy machine learning models at scale.
- **Practical Real-World Application**: The machine learning model is intended to be used by university admissions departments to identify the most qualified students.
- **Data**:
  - **Inputs (Features)**: These are the parameters the model will use to make its predictions. They include GRE scores, TOEFL scores, University Rating, Statement of Purpose (SOP) quality, Letter of Recommendation (LOR) strength, Undergraduate GPA, and Research Experience.
  - **Outputs**: The model will output the probability of admission, which will range from 0 to 1 (with 1 likely indicating certain admission and 0 indicating no chance of admission).
- The **Data Source** for the machine learning model is provided as a URL, which is a Kaggle link. Kaggle is a popular platform for data science competitions and datasets.

# University admission project

- **Inputs**: These are the variables or features the model will use to make its predictions. They include:
  - ❑ GRE Score
  - ❑ TOEFL Score
  - ❑ University Rating
  - ❑ Statement of Purpose (SOP)
  - ❑ Letter of Recommendation (LOR)
  - ❑ Cumulative Grade Point Average (CGPA)
  - ❑ Research (likely a binary indicator of research experience)
- **Machine Learning Model**: This is the core computational component that processes the inputs to make a prediction. This component represents the algorithms and statistical methods that learn from the input data to make predictions or decisions without being explicitly programmed to perform the task.
- **Output**: The result of the machine learning model's processing of the inputs, which in this case is the "Chance of Admission." This is likely a probability score between 0 and 1 where a higher score indicates a higher likelihood of admission.

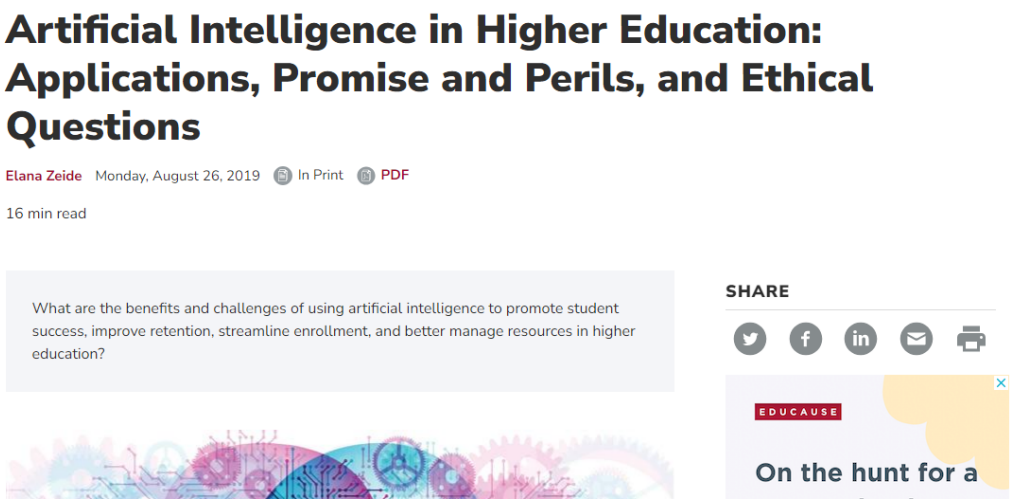
# The impact of Machine Learning (ML) and Artificial Intelligence (AI) on various facets of higher education

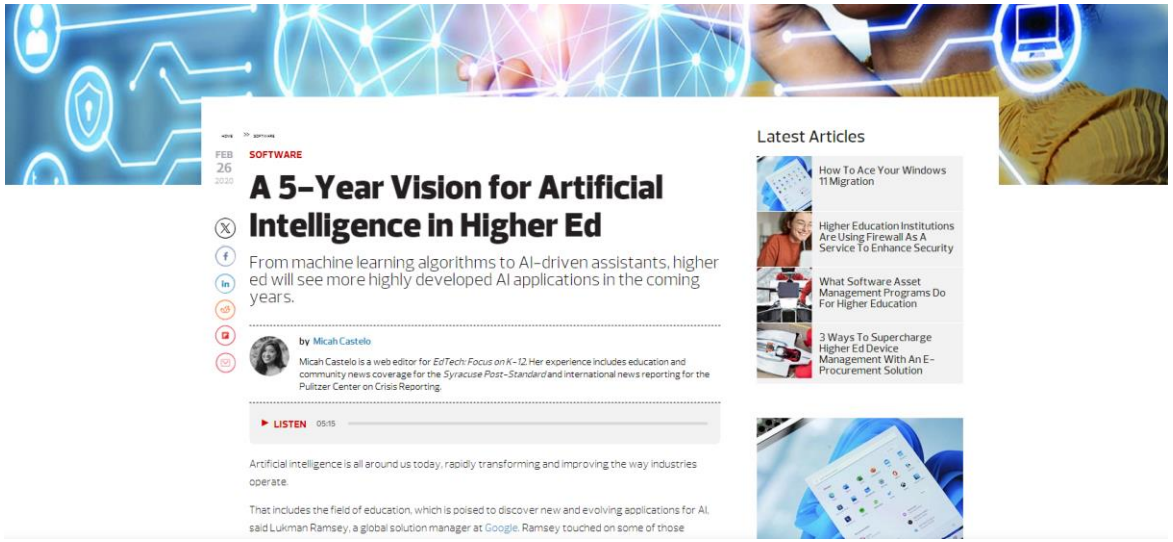
Link

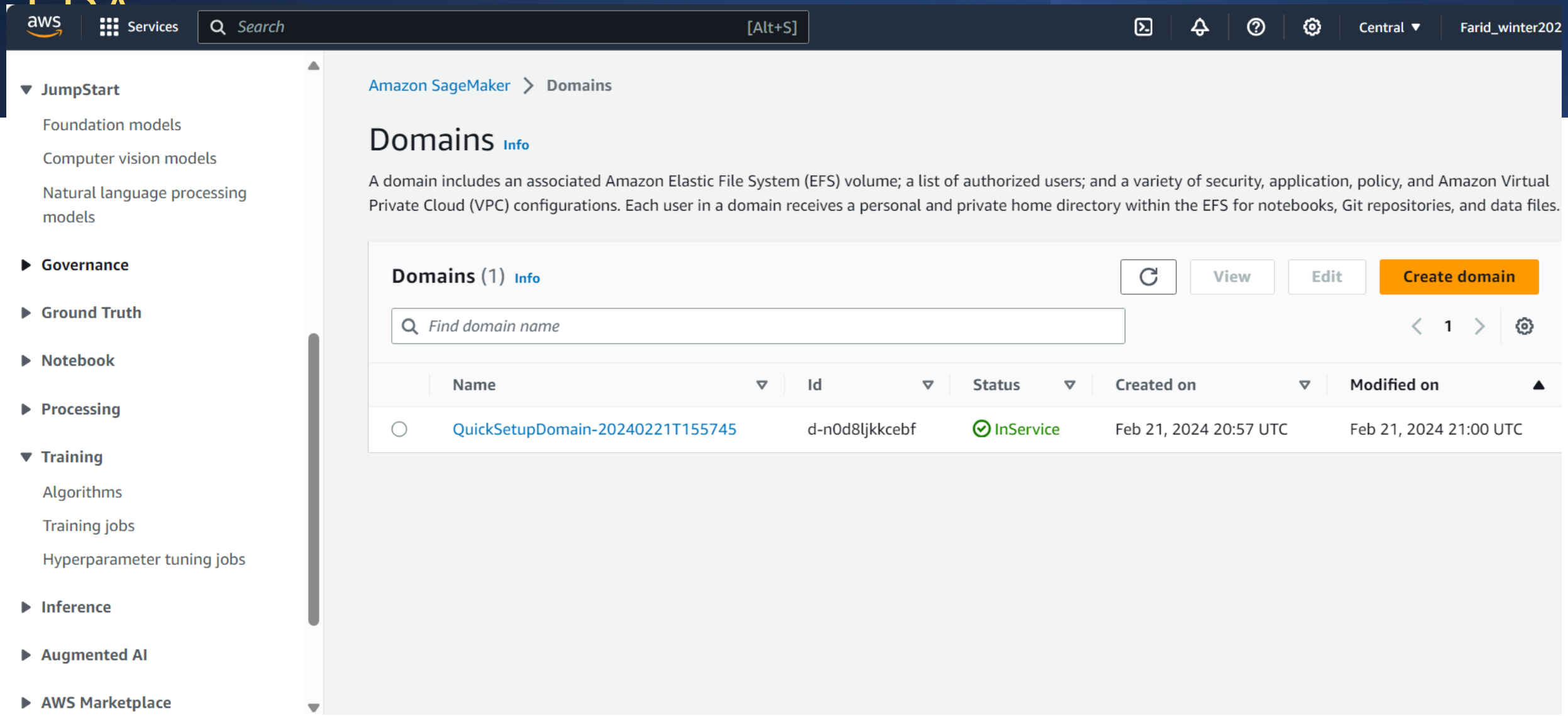
- ML and AI are presented as transformative forces in higher education, impacting several areas.
- The specific areas affected include:
  - **Marketing and Recruiting**: Using ML and AI to better target potential students and understand the effectiveness of marketing strategies.
  - **Students Admission and Enrollment**: Applying ML models, like the previously discussed predictive models, to streamline the admissions process and predict enrollment numbers.
  - **Curriculum and Resources Planning and Forecasting**: Utilizing AI for optimizing curriculum design based on trends and predicting resource requirements.
  - **Pedagogy and Personalized Students Learning Experience**: Implementing AI to customize the learning experience for students, adapting to their individual learning styles and needs.
  - **Students Support (AI-Powered Counseling)**: Providing student support services using AI, such as virtual advisors for academic counseling.

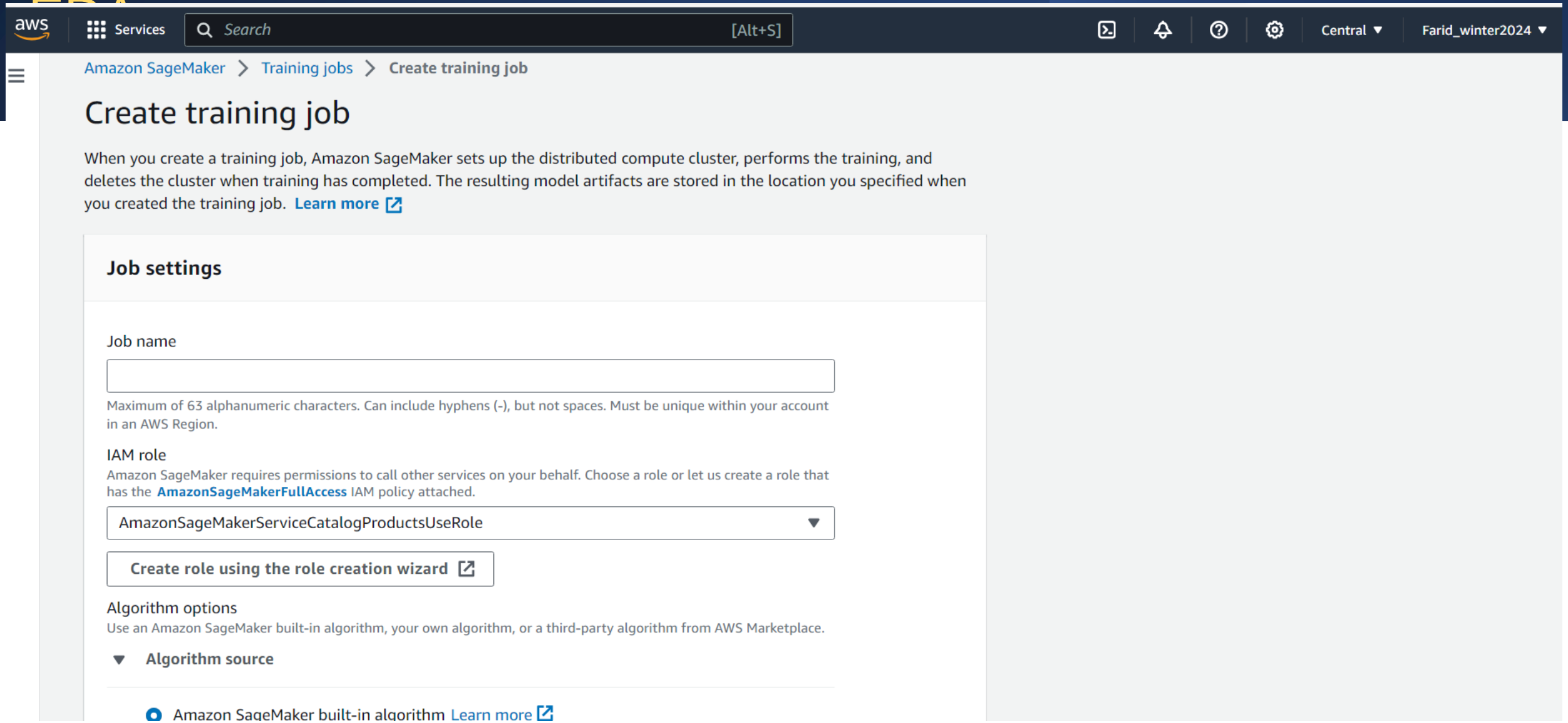
# More Resources

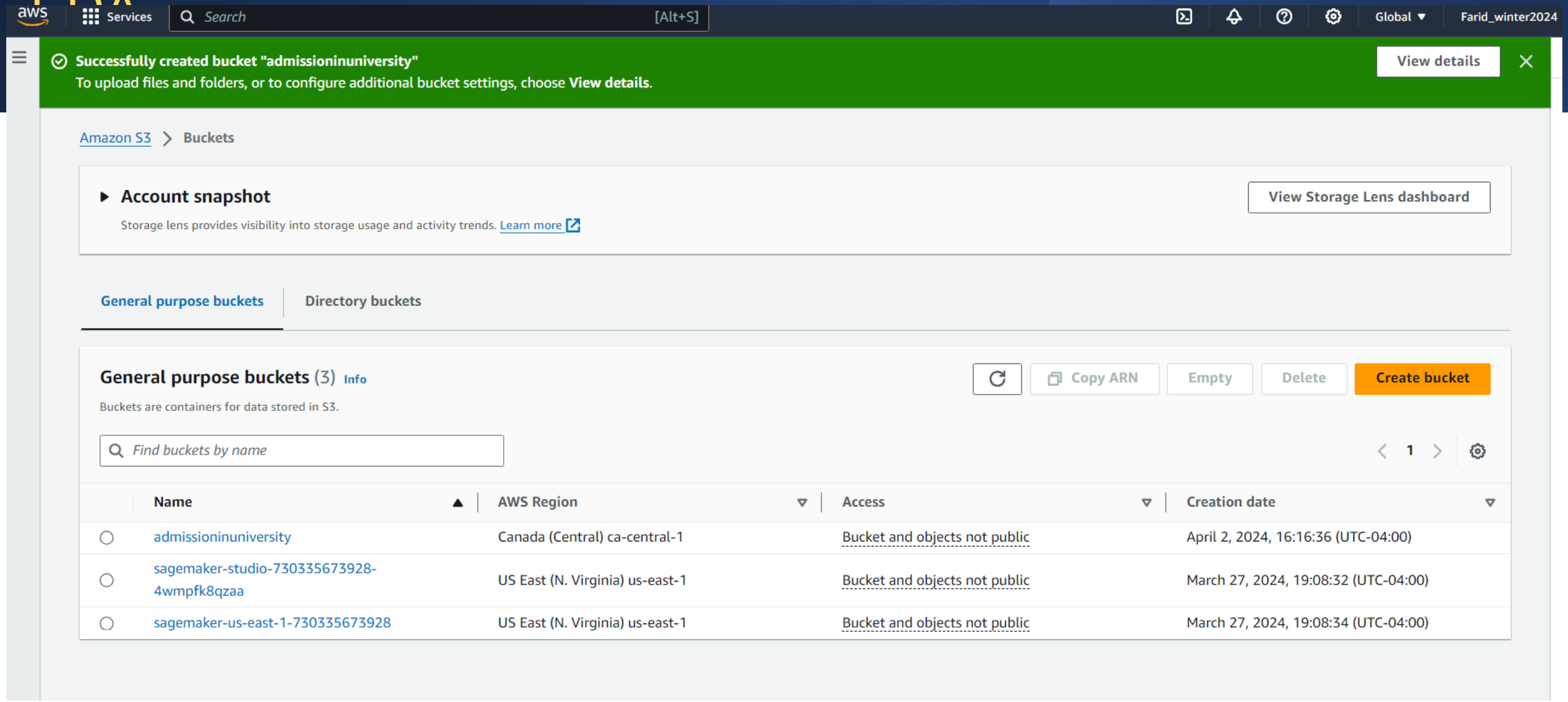
- Please read the 2 articles below and answer the following quiz.
  - https://edtechmagazine.com/higher/article/2020/02/5-year-vision-artificial-intelligence-higher-ed
  - https://er.educause.edu/articles/2019/8/artificial-intelligence-in-higher-education-applications-promise-and-perils-and-ethical-questions



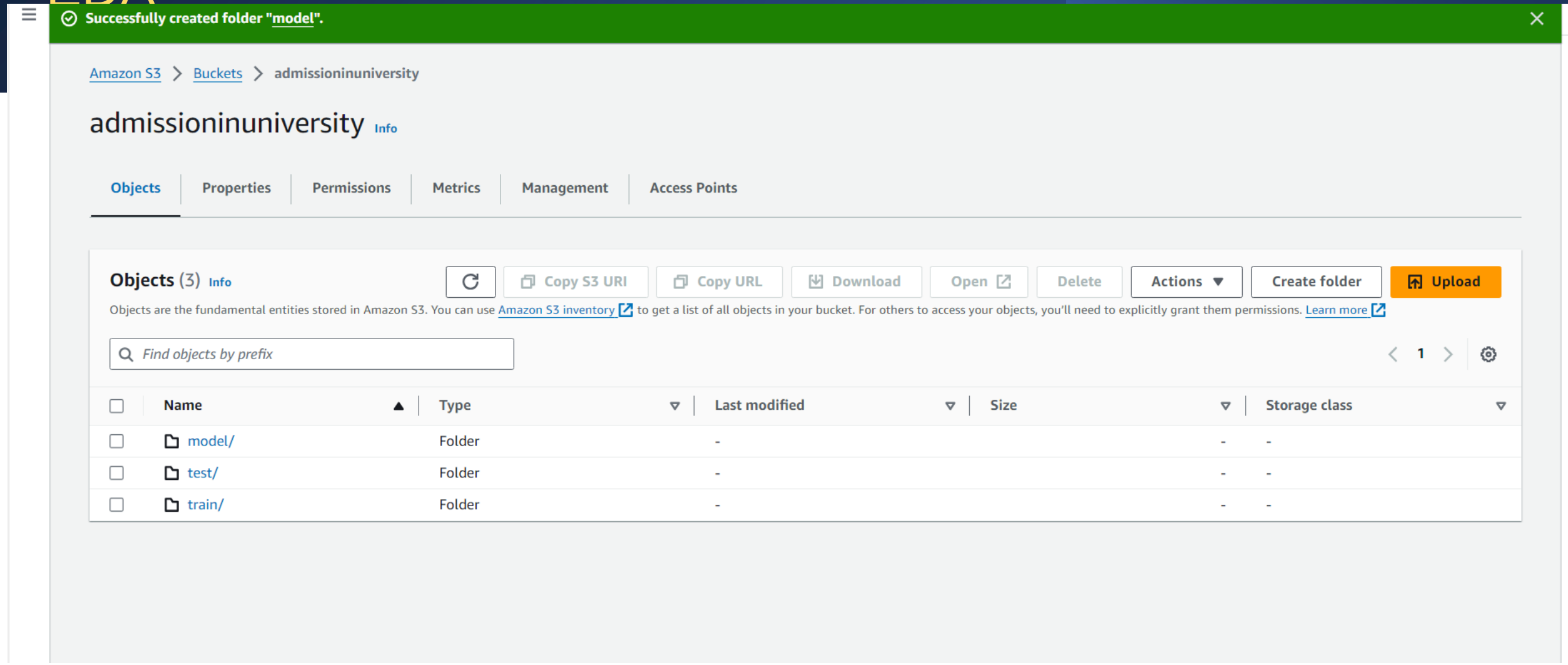




aws
Services
Search
[Alt+S]
Central
Farid_winter202
JumpStart
Foundation models
Computer vision models
Natural language processing models
Governance
Ground Truth
Notebook
Processing
Training
Algorithms
Training jobs
Hyperparameter tuning jobs
Inference
Augmented AI
AWS Marketplace
Amazon SageMaker > Domains
Domains Info
A domain includes an associated Amazon Elastic File System (EFS) volume; a list of authorized users; and a variety of security, application, policy, and Amazon Virtual Private Cloud (VPC) configurations. Each user in a domain receives a personal and private home directory within the EFS for notebooks, Git repositories, and data files.
Domains (1) Info
View
Edit
Create domain
Find domain name
1
Name
Id
Status
Created on
Modified on
QuickSetupDomain-20240221T155745
d-n0d8ljkkcebf
InService
Feb 21, 2024 20:57 UTC
Feb 21, 2024 21:00 UTC

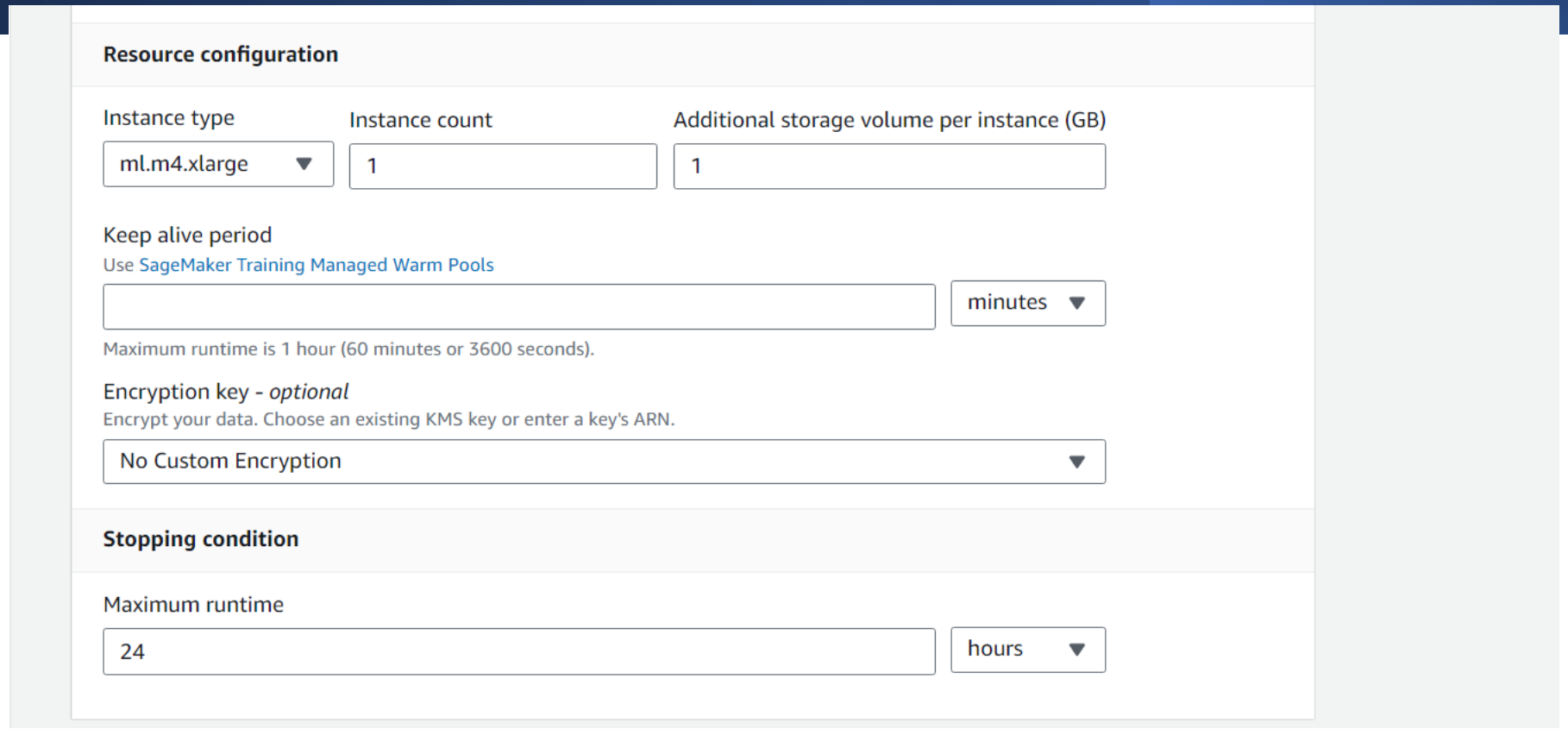

aws
Services
Search
[Alt+S]
Central
Farid_winter2024
Amazon SageMaker
Training jobs
Create training job
Create training job
When you create a training job, Amazon SageMaker sets up the distributed compute cluster, performs the training, and deletes the cluster when training has completed. The resulting model artifacts are stored in the location you specified when you created the training job. Learn more
Job settings
Job name
Maximum of 63 alphanumeric characters. Can include hyphens (-), but not spaces. Must be unique within your account in an AWS Region.
IAM role
Amazon SageMaker requires permissions to call other services on your behalf. Choose a role or let us create a role that has the AmazonSageMakerFullAccess IAM policy attached.
AmazonSageMakerServiceCatalogProductsUseRole
Create role using the role creation wizard
Algorithm options
Use an Amazon SageMaker built-in algorithm, your own algorithm, or a third-party algorithm from AWS Marketplace.
Algorithm source
Amazon SageMaker built-in algorithm Learn more


aws
Services
Search
[Alt+S]
Global
Farid_winter2024
Successfully created bucket "admissioninuniversity"
To upload files and folders, or to configure additional bucket settings, choose View details.
View details
Amazon S3
Buckets
Account snapshot
Storage lens provides visibility into storage usage and activity trends. Learn more
View Storage Lens dashboard
General purpose buckets
Directory buckets
General purpose buckets (3) Info
Buckets are containers for data stored in S3.
Copy ARN
Empty
Delete
Create bucket
Find buckets by name
1
Name
AWS Region
Access
Creation date
admissioninuniversity
Canada (Central) ca-central-1
Bucket and objects not public
April 2, 2024, 16:16:36 (UTC-04:00)
sagemaker-studio-730335673928-4wmpfk8qzaa
US East (N. Virginia) us-east-1
Bucket and objects not public
March 27, 2024, 19:08:32 (UTC-04:00)
sagemaker-us-east-1-730335673928
US East (N. Virginia) us-east-1
Bucket and objects not public
March 27, 2024, 19:08:34 (UTC-04:00)

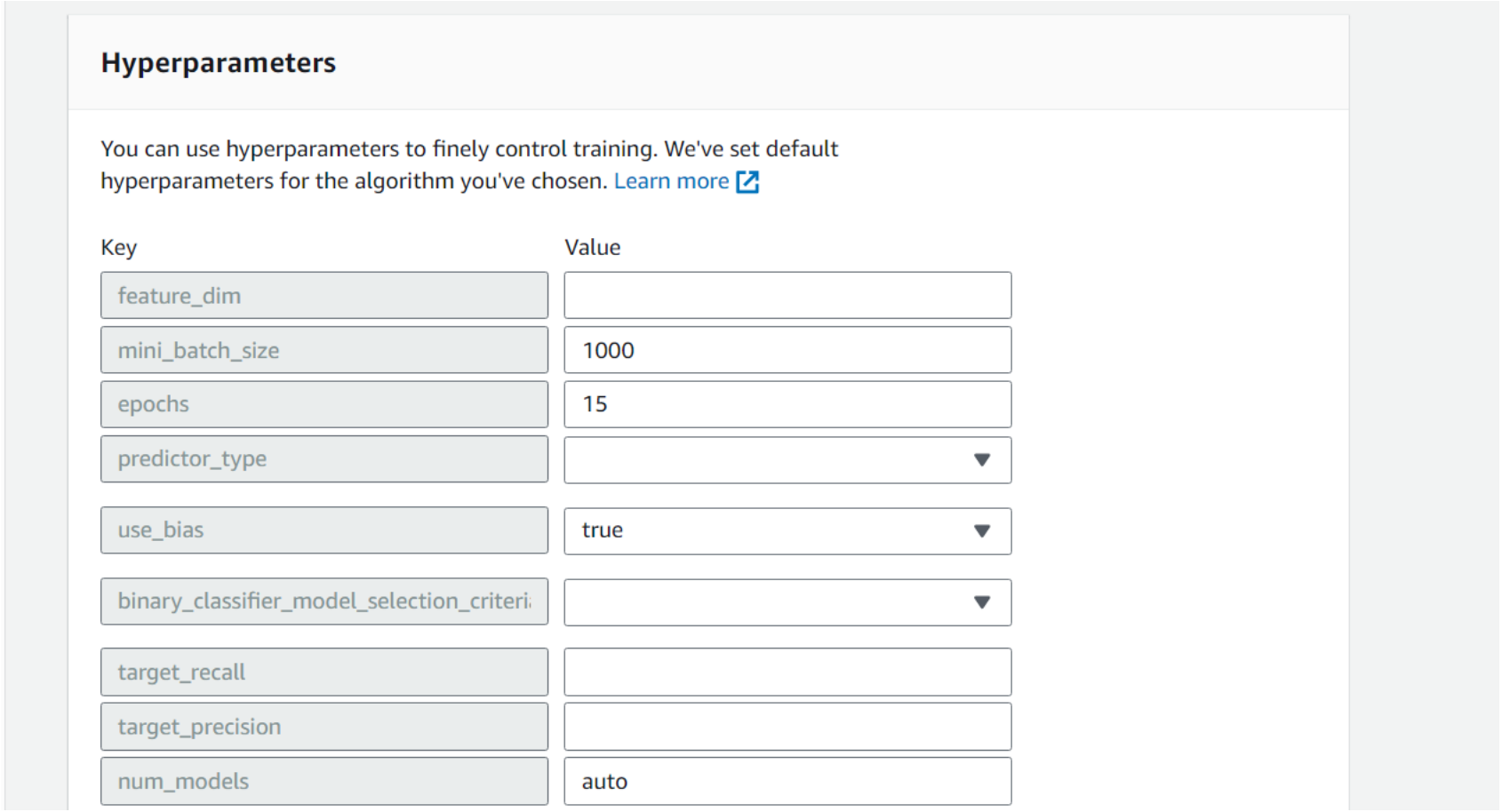
# EDA


Successfully created folder "model".
Amazon S3 > Buckets > admissioninuniversity
admissioninuniversity Info
Objects
Properties
Permissions
Metrics
Management
Access Points
Objects (3) 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
1
Name
Type
Last modified
Size
Storage class
model/ Folder - - -
test/ Folder - - -
train/ Folder - - -

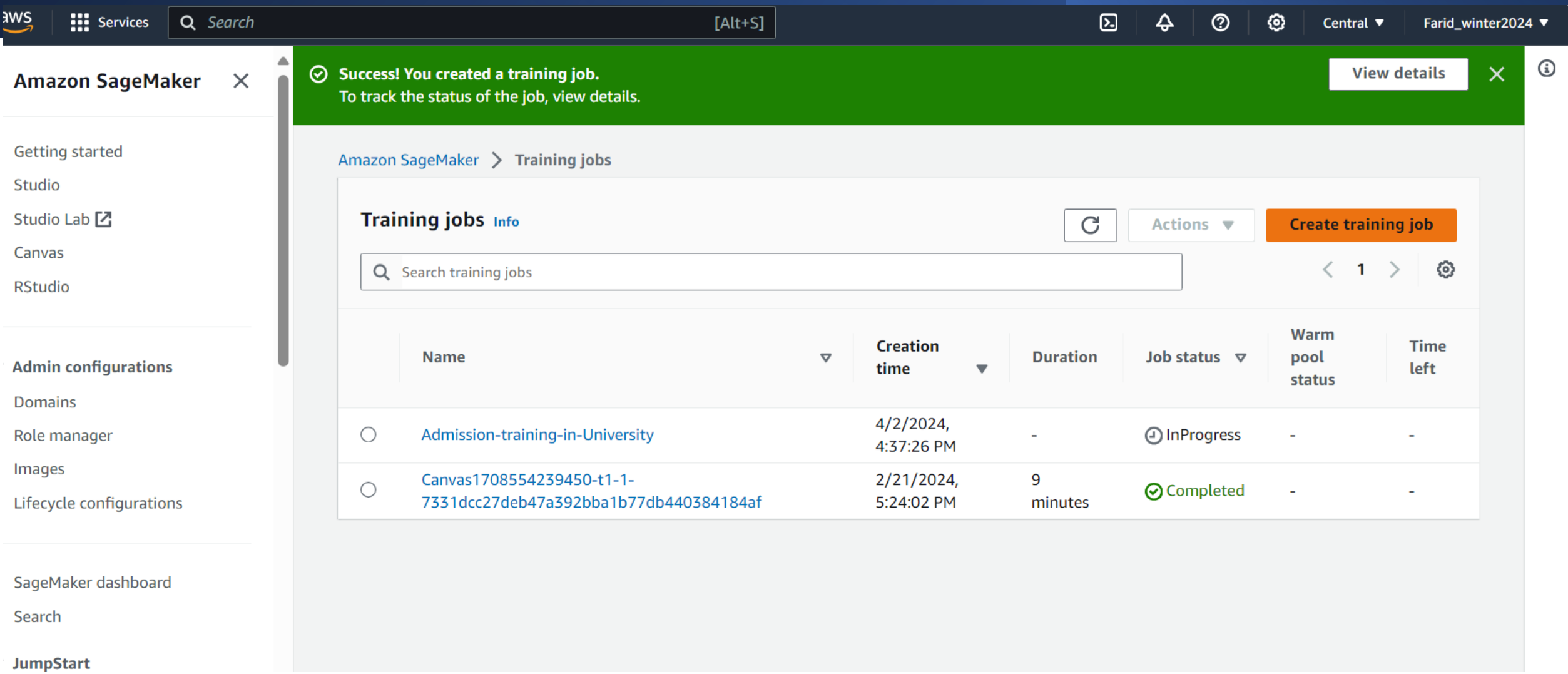
# EDA


Resource configuration
Instance type
ml.m4.xlarge
Instance count
1
Additional storage volume per instance (GB)
1
Keep alive period
Use SageMaker Training Managed Warm Pools
minutes
Maximum runtime is 1 hour (60 minutes or 3600 seconds).
Encryption key - optional
Encrypt your data. Choose an existing KMS key or enter a key's ARN.
No Custom Encryption
Stopping condition
Maximum runtime
24
hours

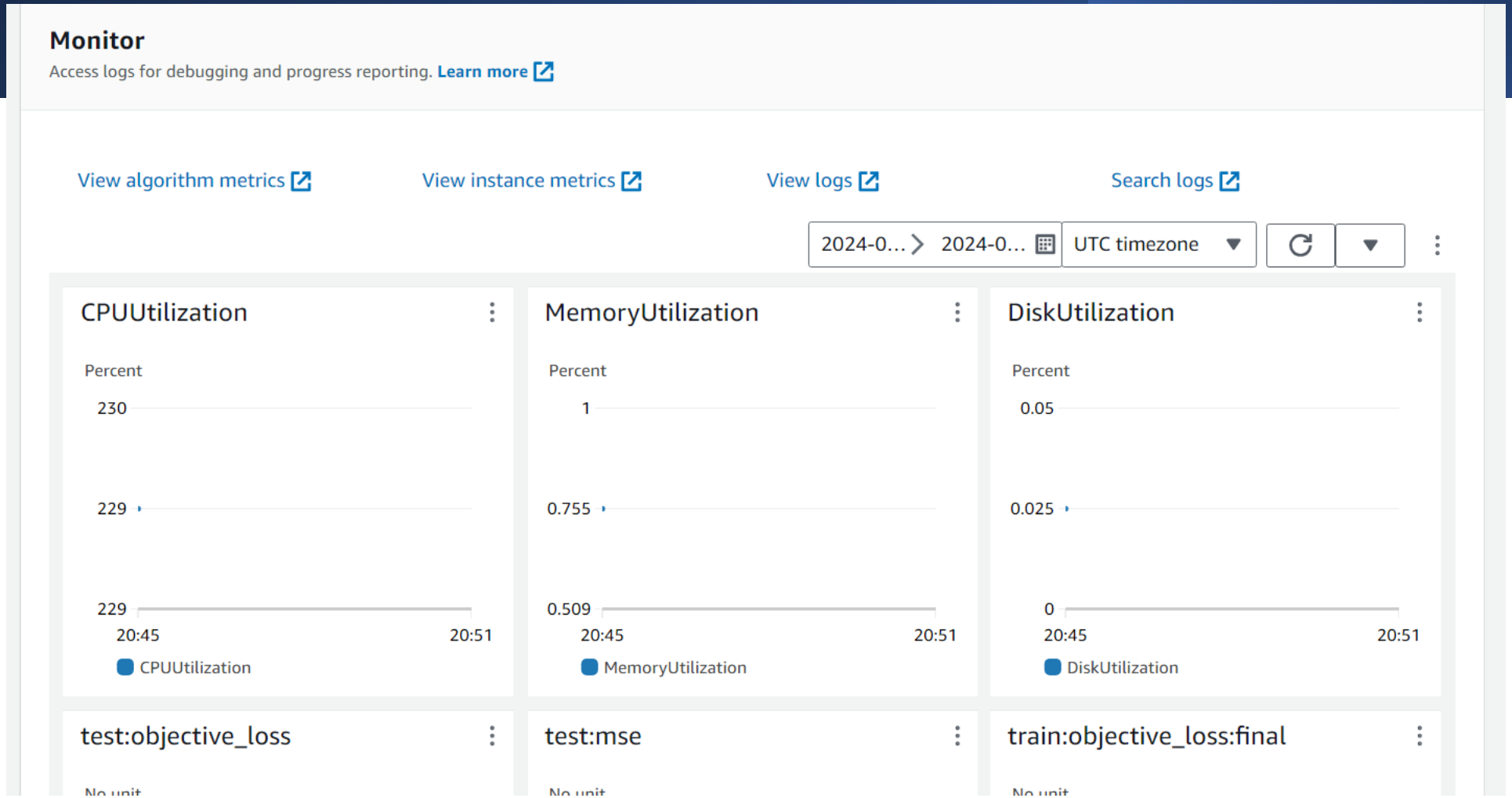
# EDA


Hyperparameters
You can use hyperparameters to finely control training. We've set default hyperparameters for the algorithm you've chosen. Learn more
Key
Value
feature_dim
mini_batch_size
1000
epochs
15
predictor_type
use_bias
true
binary_classifier_model_selection_criteri
target_recall
target_precision
num_models
auto

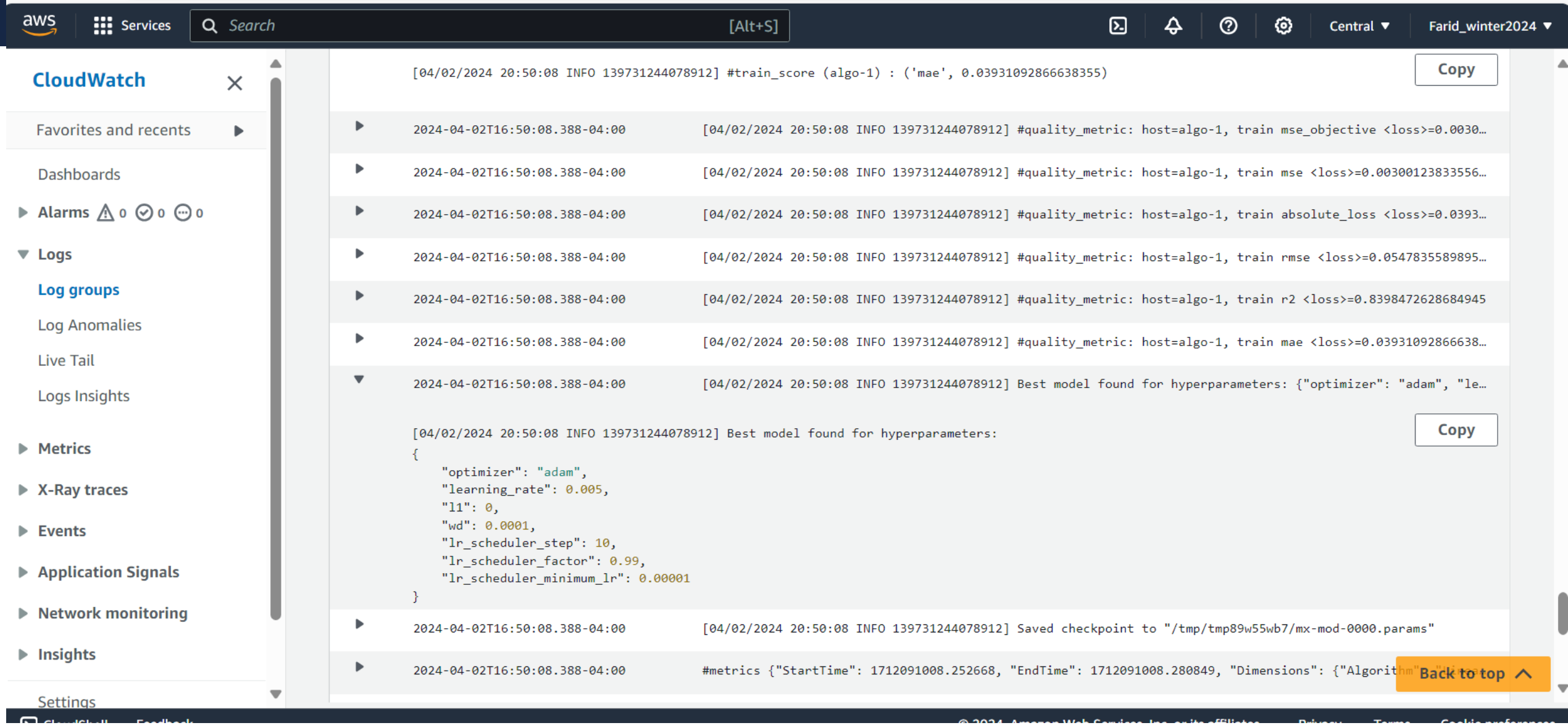
# EDA

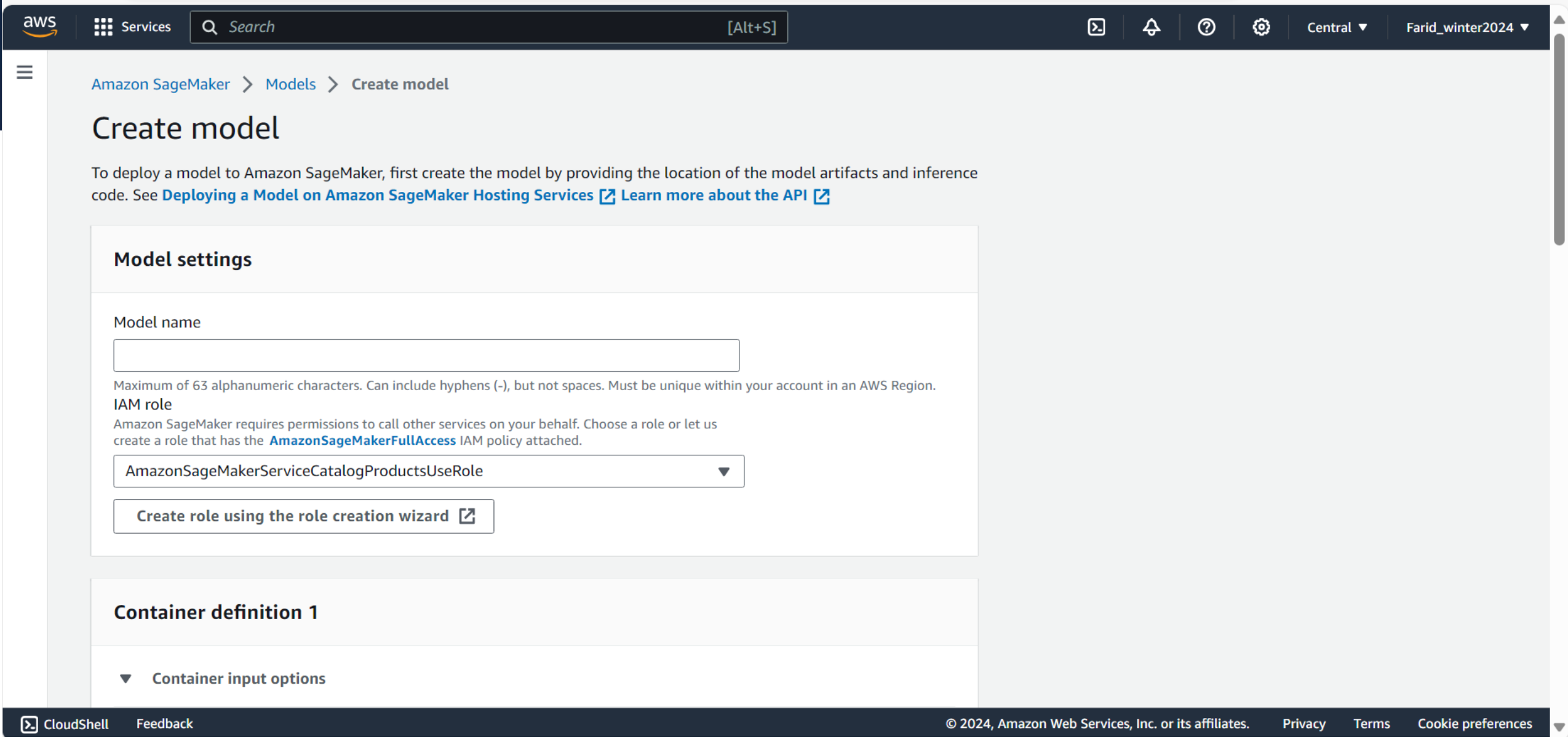

Services
Search
[Alt+S]
Central
Farid_winter2024
Amazon SageMaker
Success! You created a training job.
To track the status of the job, view details.
View details
Getting started
Studio
Studio Lab
Canvas
RStudio
Admin configurations
Domains
Role manager
Images
Lifecycle configurations
SageMaker dashboard
Search
JumpStart
Amazon SageMaker > Training jobs
Training jobs Info
Actions
Create training job
Search training jobs
1
Name
Creation time
Duration
Job status
Warm pool status
Time left
Admission-training-in-University
4/2/2024, 4:37:26 PM
-
InProgress
-
-
Canvas1708554239450-t1-1-7331dcc27deb47a392bba1b77db440384184af
2/21/2024, 5:24:02 PM
9 minutes
Completed
-
-

# EDA

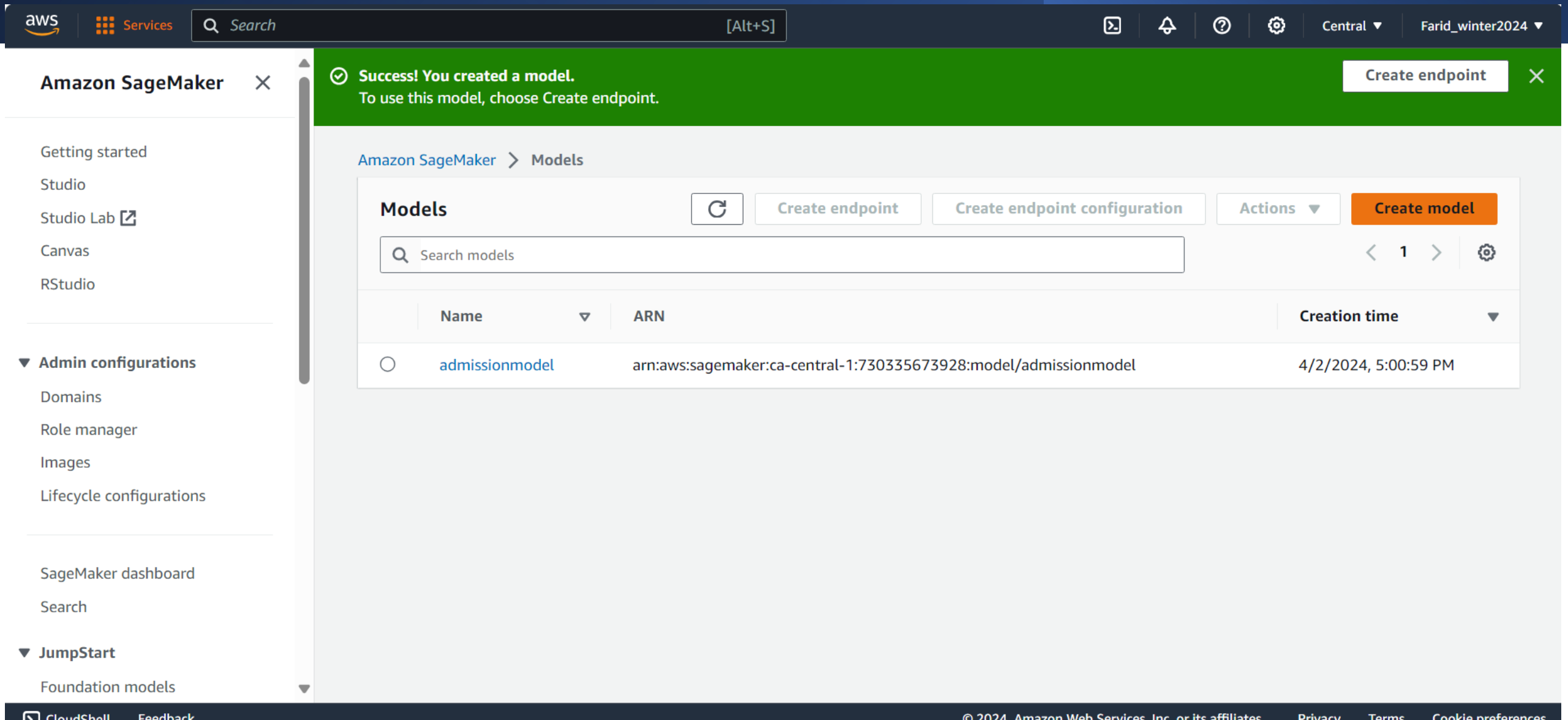

Monitor
Access logs for debugging and progress reporting. Learn more
View algorithm metrics
View instance metrics
View logs
Search logs
2024-0…
2024-0…
UTC timezone
CPUUtilization
Percent
230
229
229
20:45
20:51
CPUUtilization
MemoryUtilization
Percent
1
0.755
0.509
20:45
20:51
MemoryUtilization
DiskUtilization
Percent
0.05
0.025
0
20:45
20:51
DiskUtilization
test:objective_loss
No unit
test:mse
No unit
train:objective_loss:final
No unit

# EDA

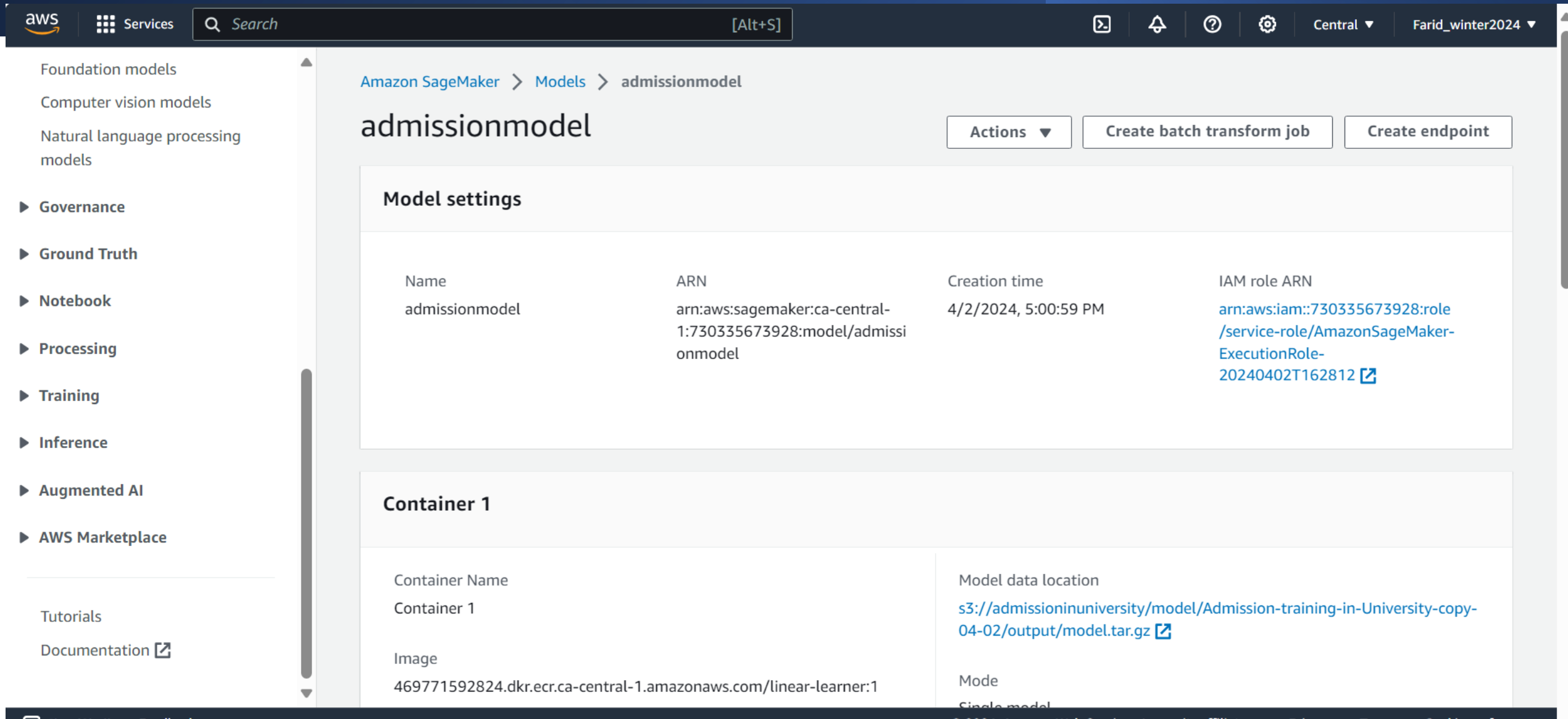

aws
Services
Search
[Alt+S]
Central
Farid_winter2024
CloudWatch
Favorites and recents
Dashboards
Alarms 0 0 0
Logs
Log groups
Log Anomalies
Live Tail
Logs Insights
Metrics
X-Ray traces
Events
Application Signals
Network monitoring
Insights
Settings
[04/02/2024 20:50:08 INFO 139731244078912] #train_score (algo-1) : ('mae', 0.03931092866638355)
Copy
2024-04-02T16:50:08.388-04:00 [04/02/2024 20:50:08 INFO 139731244078912] #quality_metric: host=algo-1, train mse_objective <loss>=0.0030…
2024-04-02T16:50:08.388-04:00 [04/02/2024 20:50:08 INFO 139731244078912] #quality_metric: host=algo-1, train mse <loss>=0.00300123833556…
2024-04-02T16:50:08.388-04:00 [04/02/2024 20:50:08 INFO 139731244078912] #quality_metric: host=algo-1, train absolute_loss <loss>=0.0393…
2024-04-02T16:50:08.388-04:00 [04/02/2024 20:50:08 INFO 139731244078912] #quality_metric: host=algo-1, train rmse <loss>=0.0547835589895…
2024-04-02T16:50:08.388-04:00 [04/02/2024 20:50:08 INFO 139731244078912] #quality_metric: host=algo-1, train r2 <loss>=0.8398472628684945
2024-04-02T16:50:08.388-04:00 [04/02/2024 20:50:08 INFO 139731244078912] #quality_metric: host=algo-1, train mae <loss>=0.03931092866638…
2024-04-02T16:50:08.388-04:00 [04/02/2024 20:50:08 INFO 139731244078912] Best model found for hyperparameters: {"optimizer": "adam", "le…
[04/02/2024 20:50:08 INFO 139731244078912] Best model found for hyperparameters:
{
"optimizer": "adam",
"learning_rate": 0.005,
"l1": 0,
"wd": 0.0001,
"lr_scheduler_step": 10,
"lr_scheduler_factor": 0.99,
"lr_scheduler_minimum_lr": 0.00001
}
Copy
2024-04-02T16:50:08.388-04:00 [04/02/2024 20:50:08 INFO 139731244078912] Saved checkpoint to "/tmp/tmp89w55wb7/mx-mod-0000.params"
2024-04-02T16:50:08.388-04:00 #metrics {"StartTime": 1712091008.252668, "EndTime": 1712091008.280849, "Dimensions": {"Algorith
Back to top


aws
Services
Search
[Alt+S]
Central
Farid_winter2024
Amazon SageMaker
Models
Create model
Create model
To deploy a model to Amazon SageMaker, first create the model by providing the location of the model artifacts and inference code. See Deploying a Model on Amazon SageMaker Hosting Services Learn more about the API
Model settings
Model name
Maximum of 63 alphanumeric characters. Can include hyphens (-), but not spaces. Must be unique within your account in an AWS Region.
IAM role
Amazon SageMaker requires permissions to call other services on your behalf. Choose a role or let us create a role that has the AmazonSageMakerFullAccess IAM policy attached.
AmazonSageMakerServiceCatalogProductsUseRole
Create role using the role creation wizard
Container definition 1
Container input options
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

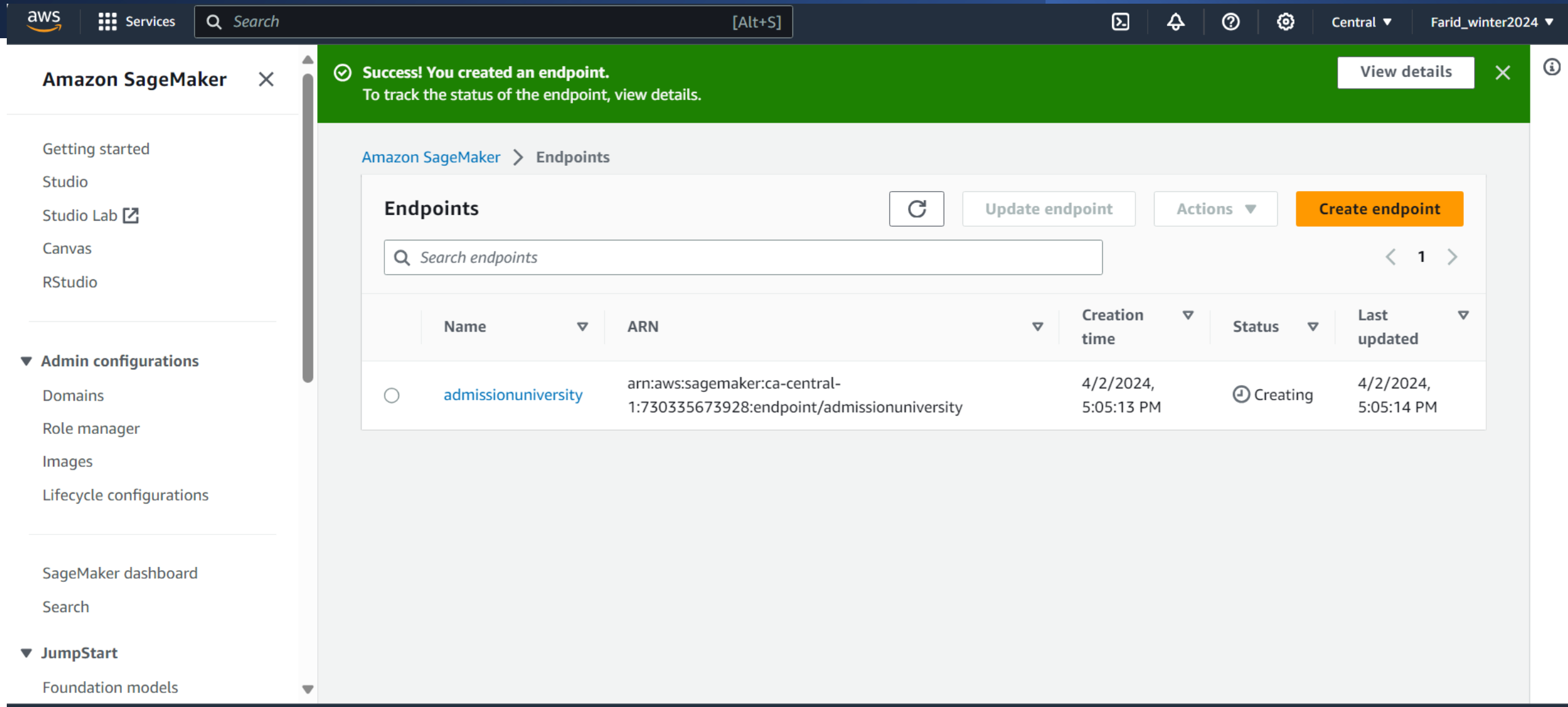
# EDA


aws
Services
Search
[Alt+S]
Central
Farid_winter2024
Amazon SageMaker
Getting started
Studio
Studio Lab
Canvas
RStudio
Admin configurations
Domains
Role manager
Images
Lifecycle configurations
SageMaker dashboard
Search
JumpStart
Foundation models
Success! You created a model.
To use this model, choose Create endpoint.
Create endpoint
Amazon SageMaker > Models
Models
Create endpoint
Create endpoint configuration
Actions
Create model
Search models
1
Name
ARN
Creation time
admissionmodel
arn:aws:sagemaker:ca-central-1:730335673928:model/admissionmodel
4/2/2024, 5:00:59 PM
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

# EDA


aws
Services
Search
[Alt+S]
Central
Farid_winter2024
Foundation models
Computer vision models
Natural language processing models
Governance
Ground Truth
Notebook
Processing
Training
Inference
Augmented AI
AWS Marketplace
Tutorials
Documentation
Amazon SageMaker > Models > admissionmodel
admissionmodel
Actions
Create batch transform job
Create endpoint
Model settings
Name
admissionmodel
ARN
arn:aws:sagemaker:ca-central-1:730335673928:model/admissionmodel
Creation time
4/2/2024, 5:00:59 PM
IAM role ARN
arn:aws:iam::730335673928:role/service-role/AmazonSageMaker-ExecutionRole-20240402T162812
Container 1
Container Name
Container 1
Image
469771592824.dkr.ecr.ca-central-1.amazonaws.com/linear-learner:1
Model data location
s3://admissioninuniversity/model/Admission-training-in-University-copy-04-02/output/model.tar.gz
Mode

# EDA


aws
Services
Search
[Alt+S]
Central
Farid_winter2024
Amazon SageMaker
Getting started
Studio
Studio Lab
Canvas
RStudio
Admin configurations
Domains
Role manager
Images
Lifecycle configurations
SageMaker dashboard
Search
JumpStart
Foundation models
Success! You created an endpoint.
To track the status of the endpoint, view details.
View details
Amazon SageMaker > Endpoints
Endpoints
Update endpoint
Actions
Create endpoint
Search endpoints
1
Name
ARN
Creation time
Status
Last updated
admissionuniversity
arn:aws:sagemaker:ca-central-1:730335673928:endpoint/admissionuniversity
4/2/2024, 5:05:13 PM
Creating
4/2/2024, 5:05:14 PM
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences