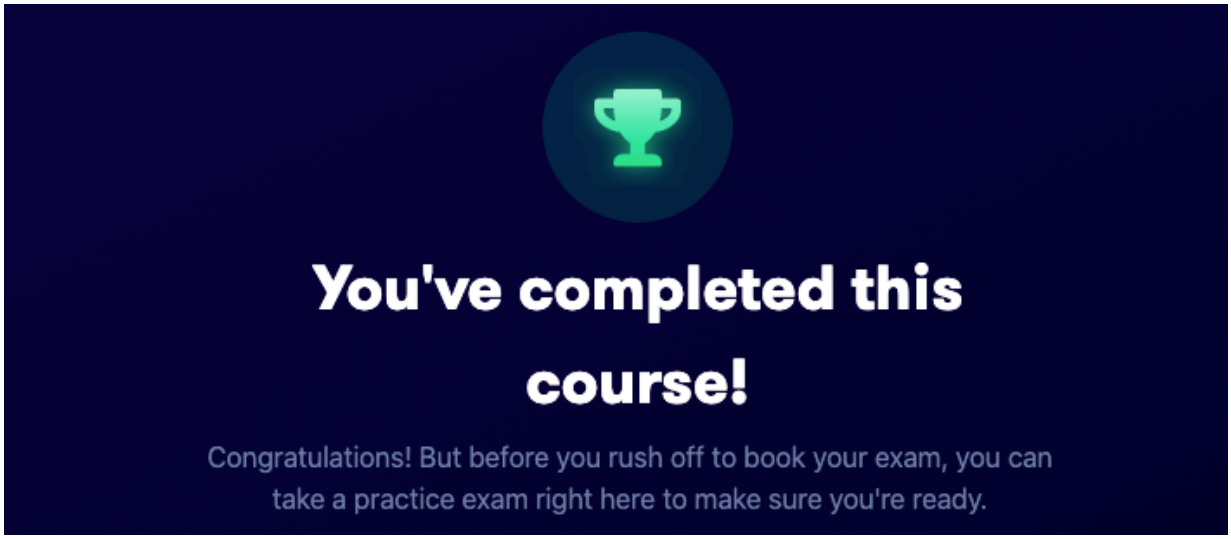
CloudGuru - Practice Exam - part 1

https://acloud.guru/course/aws-certified-machine-learning-specialty/learn/815265f7-45cc-f370-9add-65c035e6df4d/aws-certified-machine-learning-specialty/watch?backUrl=~2Fcourses



A CLOUD GURU

Back to Dashboard →

Practice Exam

AWS Certified Machine Learning - Specialty

3 hours
65 questions

The AWS Certified Machine Learning - Specialty (MLS-C01) is intended for individuals who perform a Development or Data Science role. It validates a candidate's ability to design, implement, deploy, and maintain machine learning solutions for given business problems.

DOMAIN BREAKDOWN

Domain	Percentage
Data Engineering	20%
Exploratory Data Analysis	24%
Modeling	36%
Machine Learning Implementation and Operations	20%

ATTEMPT HISTORY

You haven't attempted this exam yet.

Start Practice Exam

QUESTION 1

You are a machine learning specialist finding ways to detect anomalous data points within a given labeled data set. You've been tasked with creating a model to achieve this and also determine how accurate the model is along with other metrics like precision, recall, and F1-score metrics on the labeled data. How can this easily be achieved?

- (●) Create a model using the Random Cut Forest (RCF) algorithm with both a train and the optional test data channels. Use text/csv for training and validation data. Train the model on an ml.m4 or ml.c4 instance type.
- ( ) Create a model using the XGBoost algorithm with both a train and optional validation channels. Use application/x-recordio-protobuf for training and validation data. Train the model on an ml.c4 or ml.g4 instance type.
- ( ) Create a model using the XGBoost algorithm with both a train and the optional test data channels. Use application/x-recordio-protobuf for training and validation data. Train the model on an ml.m4 or ml.g4 instance type.
- ( ) Create a model using the Random Cut Forest (RCF) algorithm with a single train channel. Use application/x-recordio-protobuf for training and validation data. Train the model on an ml.m4 or ml.c4 instance type.
- ( ) Create a model using the Random Cut Forest (RCF) algorithm with both a train and the optional test data channels. Use application/json for training and validation data. Train the model on an ml.m4 or ml.c4 instance type.

Answer | Flag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 0 of 65 questions.

← Back Next →

QUESTION 2

You are consulting for a logistics company who wants to implement a very specific algorithm for warehouse storage optimization. The algorithm is not part of the currently available SageMaker built-in algorithms. What are your options?

(Choose 2)

- [ ] Post an incendiary message to Twitter hoping to shame AWS into adopting the specialized algorithm.
- [ ] Use a series of existing algorithms to simulate the actions of the unavailable algorithm.
- [ ] Wait until the algorithm is available in SageMaker before further work.
- [x] Search the AWS Marketplace for the algorithm. If it exists, deploy it using SageMaker for inferences.
- [x] Build the algorithm in a docker container and use that custom algorithm for training and inference in SageMaker.

Answer | Flag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 1 of 65 questions.

← Back Next →

QUESTION 3

You are a machine learning engineer working for mortgage lender who loans out money to people who are buying a new home. You need to stream line to application process using historical data from past applications. The historic data consists of attributes like applicants name, job status, marital status, loan amount, and final status (whether the application was approved or denied). What are some data preparation techniques you need to take before training the model?

- Drop the applicants name and perform one-hot encoding on job status, marital status and loan amount.
- Drop the job and marital status and perform one-hot encoding on the customer type.
- Drop both the applicant's name and final status.
- Drop the applicants name and perform one-hot encoding on job status, marital status and final status.

Answer | Flag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 2 of 65 questions.

QUESTION 4

You are a machine learning specialist working on an on-premise Hadoop cluster with thousands of Apache Parquet files. You have successfully loaded these files into Amazon S3 and now need to run SQL queries on these files. Which solution allows you to do this with the least amount of setup?

- EMR and Presto
- AWS Glue Data Catalog and Athena
- Redshift and Redshift Spectrum
- Data Pipeline and RDS

Answer | Flag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 3 of 65 questions.

QUESTION 5

You work for a market research company looking for ways to increase the efficiency of data collection. Presently, they pay students to watch groups of people as they watch commercials. The students record what percentage of the group smiles or laughs during certain moments in the commercial. Which of the following services could you use to improve this process?

- Comprehend
- Rekognition Video
- XGBoost
- Object Detection
- Semantic Segmentation

Answer | Unflag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 5 of 65 questions.

QUESTION 6

Your company wants to design a process where they can group customers into similar groups to create special targeted advertising campaigns. How would you frame this problem?

- Linear Regression Problem
- Supervised Learning Problem
- Reinforcement Learning Problem
- Binomial Classification Problem
- Unsupervised Learning Problem

Answer | Flag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 5 of 65 questions.

QUESTION 7

You have been asked to help design a customer service bot that can help answer the most common customer service questions posed on a public chat service. Which of the following might meet the need and do so with the minimum overhead?

- ○ SageMaker Seq2Seq
- ○ SageMaker BotOps
- ◉ Amazon Lex
- ○ SageMaker Object2Vec
- ○ Amazon Polly
- ○ SageMaker BlazingText

Answer | Flag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 6 of 65 questions.

QUESTION 8

You have a data set with one column missing 30% of its data. You notice that the missing features can be determined from other features in the data set. What can you do to replace the values that will cause least amount of bias?

- ○ removing the items with missing values
- ◉ multiple data imputations
- ○ use mean value
- ○ last observed carried forward

Answer | Flag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

QUESTION 9

You are working with AWS SageMaker through the process of a model based on Linear Learner. What does SageMaker do next after you issue the CreateModel API?

- SageMaker launches an appropriate inference container for the algorithm selected from the global container repository.
- SageMaker provisions an EMR cluster and prepares a Spark script for the training job.
- SageMaker launches an appropriate inference container for the algorithm selected from the regional container repository.
- Sagemaker provisions an EC2 instances using the appropriate AMI for the algorithm selected from the global container registry.
- SageMaker launches an appropriate training container from the algorithm selected from the regional container repository.

Answer | Flag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 8 of 65 questions.

QUESTION 10

You are apply standardization techniques to a feature in your dataset. The column has the following values {5, 20, 15}. The standard deviation is 6.23 and the mean of the feature 13.33. When we apply standardization what will the respective output results be?

- {0, 1, 0.66}
- {0, 0.66, 1}
- {1.33, 1.06, 0.26}
- {-1.33, 1.06, 0.26}
- {1, 0, 1}

Answer | Unflag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 9 of 65 questions.

QUESTION 11

You are working with a machine learning team training an image classification model using MXNet on Amazon SageMaker. The requirements state that the model should be at least 85% accurate. The data appears to be of good quality, but the accuracy is around 48% during training with the test data. Most of the time wrong labels are being predicted. What should be done to help increase the accuracy of the model?

- Use Amazon SageMaker's automatic model tuning. Take the best performing hyperparameters and run multiple training jobs in parallel using Apache Spark and Spark ML
- Use Amazon SageMaker's automatic model tuning. Take the best performing hyperparameters and manually adjust them to meet your requirements.
- Use Amazon SageMaker's automatic model tuning. Use AWS Batch to run multiple batches of the training data with different hyper parameters specified during the autotuning job.
- (selected) Use Amazon SageMaker's automatic model tuning. Specify the object metric and take the best performing parameters suggested by the service to use when training the model

Answer | Unflag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 10 of 65 questions.

← Back Next →

QUESTION 12

After several weeks of working on a model for genome mapping, you believe you have perfected it and now want to deploy it to a platform that will provide the highest performance. Which of the following AWS platforms will provide the highest performance for this compute-intensive model?

- (selected) EC2 P2 Instance
- EC2 X1 Instance
- EC2 F1 instance
- EC2 G3 Instance
- EC2 M2 Instance

Answer | Flag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 11 of 65 questions.

QUESTION 13

You notice that your hyperparameter tuning job for a model that uses Linear Learning with the objective metric of objective_loss has been running for several hours without improvement. What would you have done to ensure the most efficient use of resources?

- Choose the XGBoost algorithm instead
- Enable more GPU instances for training
- Run the training job on Spot Instances during nighttime hours
- (●) Enable early stopping
- Reform the problem as Logistic Regression

Answer | Flag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 12 of 65 questions.

QUESTION 14

You are a machine learning engineer on a team that is designing a deep sea exploration robot which will explore areas never before visited. The robot will be so deep that it's impractical to send commands for every motion. Rather then robot needs to know how to navigate among obstacles itself while collecting data. How would you frame this problem?

- This problem is not suitable for machine learning.
- Heuristic Learning Problem
- Object Detection Problem
- (●) Reinforcement Learning Problem
- Supervised Learning Problem

Answer | Flag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 13 of 65 questions.

QUESTION 15

You are a data scientist that has been tasked with setting up an Amazon Elastic Map Reduce (EMR) cluster to host your organization's data lake. You also need to setup this cluster for machine learning processes and it has been decided to use Amazon SageMaker libraries as the machine learning platform. What steps do you need to take to start using SageMaker with your EMR cluster data lake?

(Choose 2)

- [x] Download the aws-sagemaker-spark-sdk component along with Spark on your EMR cluster
- [ ] Run your SageMaker Spark application on EMR by submitting your Spark application jar and any additional dependencies your Spark application uses
- [ ] Convert EMR DataFrame to CSV and use that to train and infer your model
- [ ] Use Apache Mahout within an EMR Notebook to train and infer your model
- [x] Ensure the EMR cluster and SageMaker hosted model are in the same region to make successful inferences

Answer | Flag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 14 of 65 questions.

← Back Next →

QUESTION 16

You are working for an online shopping platform that records actions made by its users. This information is captured in multiple JSON files stored in S3. You have been tasked with moving this data into Amazon Redshift database tables as part of a data lake migration process. Which of the following needs to occur to achieve this in the most efficient way?

(Choose 3)

- [x] Launch an Amazon Redshift cluster and create database tables.
- [ ] Setup DynamoDB table and use Data Pipeline to load the S3 data into DynamoDB table.
- [ ] Use the INSERT command to load the tables from the data files on Amazon S3.
- [x] Use multiple concurrent COPY commands to load the table from each JSON file.
- [x] Use COPY commands to load the tables from the data files on Amazon S3.
- [ ] Use COPY commands to load the tables from the data files on DynamoDB.
- [ ] Troubleshoot load errors and modify your COPY commands to correct the errors.

Answer | Flag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 15 of 65 questions.

← Back Next →

QUESTION 17

You are a machine learning specialist designing a regression model to predict the sales for an upcoming summer sale. The data from the past sales consists of 1,000 records containing 20 numeric attributes. As you start to analyze the data, you discovered that 30 records have values that are above the top upper whisker in the box plot upper quartile. You confirm with management that these records are unusual, but certainly valid values. There are also 78 records where another numerical value is blank. What should you do to correct these problems?

- ○ Drop the unusual records and fill in the blank values with 0
- ○ Drop the unusual records and replace the blank values with separate Boolean values
- ○ Use the unusual data and replace the missing values with a separate Boolean variable
- ◉ Drop the unusual records and replace the blank values with the mean value

Answer | Flag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 16 of 65 questions.

QUESTION 18

You want to design a model that will predict the price of a used car based on attributes of the car as a linear regression model. Which of the following algorithms could you use for your car price prediction model?
(Choose 2)

- ☐ BlazingText
- ☑ Linear Learner
- ☐ NTM
- ☐ Seq2Seq
- ☑ XGBoost

Answer | Flag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 17 of 65 questions.

QUESTION 19

You work for an organization that handles highly sensitive information on a daily basis. The company has different compliance rules that require all data be encrypted at rest. When preparing your machine learning models using SageMaker, how can you achieve these requirements?

- Ensure the role associated with the SageMaker Notebook instance is assigned to the customer managed key in KMS
- Create a customer managed key in KMS and use it when creating your SageMaker Notebook instance
- Stop your SageMaker notebook instance, create a customer managed key in KMS and attach it to the stopped SageMaker Notebook instance
- Unmount the EBS volume from SageMaker Notebook instance, encrypt it with a KMS key, and reattach to SageMaker Notebook instance

Answer | Flag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 18 of 65 questions.

QUESTION 20

You are preparing a data repository to host the dataset needed to train a model using Amazon SageMaker's Semantic Segmentation algorithm. You have collected all the data locally on your machine and need to transfer it into your data repository. What must be done to accomplish this?

- Host the dataset on a SageMaker Jupyter Notebook on a ml.c4.8xlarge instance, storing it in one train channel. Store the images in one directory and annotations in another directory. Use a label map that describes how the annotation mappings are established.
- Host the dataset in Amazon S3 and storing it in two channels. One for train and one for validation, in four directories, two for images and two for annotations. Use a label map that describes how the annotation mappings are established.
- Host the dataset on an Provisioned IOPS EBS volume optimized to process IO for image data storing it in two channels. One for train and one for validation, in four directories, two for images and two for annotations. Use a label map that describes how the annotation mappings are established.
- Host the dataset in Amazon S3, storing it in one train channel. Store the images in one directory and annotations in another directory. Use a label map that describes how the annotation mappings are established.
- Host the dataset on a SageMaker Jupyter Notebook on a ml.p2.8xlarge instance, storing it in one train channel. Store the images in one directory and annotations in another directory. Use a label map that describes how the annotation mappings are established.

Answer | Unflag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 19 of 65 questions.

← Back | Next →

QUESTION 21

You are a machine learning specialist who has millions of rows of data spread across thousands of files stored in S3. The metadata about these files are stored in the S3 bucket as well. First you need to search and analyze the data before running it through your model. You will most likely do multiple searches depending on results found throughout your research. Which solution meets the requirements with the least amount of setup?

- (●) Use Amazon Athena to analyze and query your S3 data
- ( ) First, enable S3 analytics then use the metastore files to analyze your data
- ( ) Create a Redshift cluster that uses S3 as the input data source and use Redshift Spectrum to analyze and query your S3 data
- ( ) Create an EMR cluster with Apache Hive to analyze and query your data

Answer | Flag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 20 of 65 questions.

QUESTION 22

You have build two slightly different models for performing multi-class classification. What metric would be the most holistic for evaluating the models against each other?

- ( ) k-Fold Cross-validation
- ( ) Recall
- ( ) Precision
- (●) Macro F1 Score
- ( ) Retention

Answer | Flag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 21 of 65 questions.

QUESTION 23

You want to build a recommendation engine that can predict if people might like new jelly bean flavors based on their shared preferences for other jellybeans they have sampled. You have data from hundreds of thousands of customers, but out of the thousand flavors in your database, the large majority of customers have only sampled a small percentage of flavors. What algorithm would you choose to provide the best results for your recommendation engine?

- ○ DeepAR
- ◉ Factorization Machines
- ○ Random Cut Forest
- ○ Linear Learner

Answer | Flag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 22 of 65 questions.

QUESTION 24

Creating an S3 VPC Endpoint in your VPC will have which of the following impacts?
(Choose 2)

- ☐ Increase egress costs.
- ☑ Reduce egress costs.
- ☐ Reduce security.
- ☐ Increase latency.
- ☑ Improve security.

Answer | Unflag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 23 of 65 questions.

QUESTION 25

You are designing a security approach for ensuring only members of certain projects can access notebook instances for their own projects. DevTeam1 should only access Project1 notebooks while DevTeam2 should only access Project2 notebooks. What is a valid way to implement this restriction?

- Implement Federation using LDAP and SAML over SSH. Ensure that all DEV team members use MFA upon each attempt to access their respective project notebooks.
- Use the ResourceTag condition and add a Project tag to each notebook.
- Created shared accounts for the DEV teams and record all activities using CloudTrail. Use CloudTrail alarms to notify if a team access a notebook they are not authorized to access.
- Create a VPC Gateway Endpoint and route all traffic from each team member to only the S3 buckets containing their respective project models.

Answer | Unflag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 24 of 65 questions.

QUESTION 26

How might you frame an ML approach for predicting if a random number generator will return a 1 or 0 given that the number generator is perfectly random and you are provided with 1000 prior number selections?

- Forecasting
- Logical Regression
- Binary Classification
- Machine Learning not needed
- Linear Regression

Answer | Flag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 25 of 65 questions.

QUESTION 27

You are designing a binary classification model using the XGBoost algorithm. Which of the following would you most likely use as an objective for evaluating the model?

- Root Square Mean Error approaches 1.
- Area Under the Curve approaches 0.
- Macro F1 score approaches 0.
- Root Square Mean Error approaches 0.
- Area Under the Curve approaches 1.

Answer | Unflag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 26 of 65 questions.

QUESTION 28

You work for a company that is spinning up a new machine learning team. You need to setup a machine learning environment that will be accessed by multiple Data Scientists. Each new Data Scientist on the team needs to have their own Jupyter Notebook instance in Amazon SageMaker. How should you manage access to SageMaker Notebook instances?

- Set Up an ACL (Access Control List) for each notebook instance. Attach each Data Scientist role to the ACL associated with their personal notebook instance
- Setup a VPC for each notebook instance with an egress rule with IP addresses associated with the respective Data Scientists who are allowed access
- Ensure that each IAM policy associated with the Data Scientists role and their respective notebook instance has the iam:PassRole denied
- Attach an IAM policy to the Data Scientists' IAM users that allows access to their personal notebook instances only
- Use Amazon CloudWatch to trigger a Lambda function that restricts unauthorized access

Answer | Flag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 27 of 65 questions.

← Back Next →

QUESTION 30

You have recently started a training job for a machine learning model in an Amazon SageMaker Jupyter notebook. What is the easiest way to visualize memory utilization, CPU, and training metrics?

- (●) Setup CloudWatch dashboard
- ( ) Use CloudWatch logs and Kafka
- ( ) Stream Kinesis Delivery Stream to stream instance and training metrics to S3. Use QuickSight to visualize the metrics.
- ( ) Push CloudWatch logs to S3 and use QuickSight to visualize the metrics

Answer | Flag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 29 of 65 questions.

QUESTION 31

What method does Amazon SageMaker uses to facilitate hyperparameter tuning?

- ( ) Stochastic Search
- (●) Bayesian Optimization
- ( ) Random Search
- ( ) Gaussian Optimization
- ( ) Matrix Search

Answer | Unflag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 30 of 65 questions.

QUESTION 32

After evaluating a training job, you realize that your model is overfitting. You suspect that its due to the very high number of features in the dataset. What might be a way you can reduce dimensionality for the dataset?

- ○ LDA
- ○ NTM
- ◉ PCA
- ○ One-hot Encoding
- ○ Softmax function
- ○ Object2Vec

Answer | Flag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 31 of 65 questions.

QUESTION 33

You have been tasked with using Polly to translate text to speech in the company announcements that launch weekly. The problem you are encountering is how Polly is incorrectly translating the companies acronyms. What can be done for future tasks to help prevent this?

(Choose 2)

- ☐ Use Amazon Comprehend to pull parts of speech and use to help pronounce acronyms
- ☑ Use speech marks for input text documents
- ☐ Use Amazon Transcribe to first map the acronyms to pronunciations then include them in the Amazon polly pipeline
- ☑ Create dictionary lexicon
- ☐ Use SSML tags in documents

Answer | Flag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 32 of 65 questions.

QUESTION 34

You have been working on a machine learning model for several iterations and feel that it is ready for production and allow real users to begin making inferences to it. You want to ensure that the models are ran on multiple instances in different availability zones. What steps can you take to ensure this occurs?

- Use Amazon SageMaker hosting services and specify a single instance. Use Route53 with failover routing policy to ensure users are routed to different availability zone if the instance becomes unreachable
- Use Amazon SageMaker hosting services, specify two or more instances and specify multiple availability zones you want to launch models in
- Use Amazon SageMaker hosting services, deploy two different variants of the model routing 50% of the traffic to one availability zone and the other 50% to the other availability zone
- Use Amazon SageMaker hosting services and specify two or more instances. Amazon SageMaker launches them in multiple availability zones automatically

Answer | Flag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 33 of 65 questions.

← Back Next →

QUESTION 34

You have been working on a machine learning model for several iterations and feel that it is ready for production and allow real users to begin making inferences to it. You want to ensure that the models are ran on multiple instances in different availability zones. What steps can you take to ensure this occurs?

- Use Amazon SageMaker hosting services and specify a single instance. Use Route53 with failover routing policy to ensure users are routed to different availability zone if the instance becomes unreachable
- Use Amazon SageMaker hosting services, specify two or more instances and specify multiple availability zones you want to launch models in
- Use Amazon SageMaker hosting services, deploy two different variants of the model routing 50% of the traffic to one availability zone and the other 50% to the other availability zone
- Use Amazon SageMaker hosting services and specify two or more instances. Amazon SageMaker launches them in multiple availability zones automatically

Answer | Flag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 33 of 65 questions.

QUESTION 35

You are a Data Scientist working on a model that predicts fraudulent and non fraudulent transactions. You notice that 90% of the samples are non fraudulent which makes up the majority of the dataset. What are some methods you can use to address this issue in the data?

(Choose 3)

- [ ] Resample the dataset to correct imbalances of each transaction type
- [ ] Apply Principal Component Analysis (PCA) to undersample non fraudulent transactions
- [ ] Drop all fraudulent transaction before training the model
- [x] Combine dataset with a public dataset that have a majority of fraudulent transactions
- [ ] Use K-means cluster to find outliers for non fraudulent transactions and use those as fraudulent samples
- [x] Collect more data to even the imbalances in the dataset
- [x] Try a different algorithm or a combination of algorithms on your dataset

Answer | Unflag for Review

NAVIGATE

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	

Submit Exam

You've answered 34 of 65 questions.

← Back Next →