

Amazon-Web-Services

Exam Questions MLA-C01

AWS Certified Machine Learning Engineer - Associate



NEW QUESTION 1

A company's ML engineer has deployed an ML model for sentiment analysis to an Amazon SageMaker endpoint. The ML engineer needs to explain to company stakeholders how the model makes predictions.

Which solution will provide an explanation for the model's predictions?

- A. Use SageMaker Model Monitor on the deployed model.
- B. Use SageMaker Clarify on the deployed model.
- C. Show the distribution of inferences from A/ testing in Amazon CloudWatch.
- D. Add a shadow endpoint
- E. Analyze prediction differences on samples.

Answer: B

NEW QUESTION 2

An ML engineer is using Amazon SageMaker to train a deep learning model that requires distributed training. After some training attempts, the ML engineer observes that the instances are not performing as expected. The ML engineer identifies communication overhead between the training instances.

What should the ML engineer do to MINIMIZE the communication overhead between the instances?

- A. Place the instances in the same VPC subne
- B. Store the data in a different AWS Region from where the instances are deployed.
- C. Place the instances in the same VPC subnet but in different Availability Zone
- D. Store the data in a different AWS Region from where the instances are deployed.
- E. Place the instances in the same VPC subne
- F. Store the data in the same AWS Region and Availability Zone where the instances are deployed.
- G. Place the instances in the same VPC subne
- H. Store the data in the same AWS Region but in a different Availability Zone from where the instances are deployed.

Answer: C

NEW QUESTION 3

A company is using an Amazon Redshift database as its single data source. Some of the data is sensitive.

A data scientist needs to use some of the sensitive data from the database. An ML engineer must give the data scientist access to the data without transforming the source data and without storing anonymized data in the database.

Which solution will meet these requirements with the LEAST implementation effort?

- A. Configure dynamic data masking policies to control how sensitive data is shared with the data scientist at query time.
- B. Create a materialized view with masking logic on top of the databas
- C. Grant the necessary read permissions to the data scientist.
- D. Unload the Amazon Redshift data to Amazon S3. Use Amazon Athena to create schema-on-read with masking logi
- E. Share the view with the data scientist.
- F. Unload the Amazon Redshift data to Amazon S3. Create an AWS Glue job to anonymize the dat
- G. Share the dataset with the data scientist.

Answer: A

NEW QUESTION 4

A company uses Amazon Athena to query a dataset in Amazon S3. The dataset has a target variable that the company wants to predict.

The company needs to use the dataset in a solution to determine if a model can predict the target variable.

Which solution will provide this information with the LEAST development effort?

- A. Create a new model by using Amazon SageMaker Autopilo
- B. Report the model's achieved performance.
- C. Implement custom scripts to perform data pre-processing, multiple linear regression, and performance evaluatio
- D. Run the scripts on Amazon EC2 instances.
- E. Configure Amazon Macie to analyze the dataset and to create a mode
- F. Report the model's achieved performance.
- G. Select a model from Amazon Bedroc
- H. Tune the model with the dat
- I. Report the model's achieved performance.

Answer: A

NEW QUESTION 5

An advertising company uses AWS Lake Formation to manage a data lake. The data lake contains structured data and unstructured data. The company's ML engineers are assigned to specific advertisement campaigns.

The ML engineers must interact with the data through Amazon Athena and by browsing the data directly in an Amazon S3 bucket. The ML engineers must have access to only the resources that are specific to their assigned advertisement campaigns.

Which solution will meet these requirements in the MOST operationally efficient way?

- A. Configure IAM policies on an AWS Glue Data Catalog to restrict access to Athena based on the ML engineers' campaigns.
- B. Store users and campaign information in an Amazon DynamoDB tabl
- C. Configure DynamoDB Streams to invoke an AWS Lambda function to update S3 bucket policies.
- D. Use Lake Formation to authorize AWS Glue to access the S3 bucke
- E. Configure Lake Formation tags to map ML engineers to their campaigns.
- F. Configure S3 bucket policies to restrict access to the S3 bucket based on the ML engineers' campaigns.

Answer: C

NEW QUESTION 6

A company is running ML models on premises by using custom Python scripts and proprietary datasets. The company is using PyTorch. The model building requires unique domain knowledge. The company needs to move the models to AWS. Which solution will meet these requirements with the LEAST effort?

- A. Use SageMaker built-in algorithms to train the proprietary datasets.
- B. Use SageMaker script mode and premade images for ML frameworks.
- C. Build a container on AWS that includes custom packages and a choice of ML frameworks.
- D. Purchase similar production models through AWS Marketplace.

Answer: B

NEW QUESTION 7

A company has trained an ML model in Amazon SageMaker. The company needs to host the model to provide inferences in a production environment. The model must be highly available and must respond with minimum latency. The size of each request will be between 1 KB and 3 MB. The model will receive unpredictable bursts of requests during the day. The inferences must adapt proportionally to the changes in demand. How should the company deploy the model into production to meet these requirements?

- A. Create a SageMaker real-time inference endpoint
- B. Configure auto scaling
- C. Configure the endpoint to present the existing model.
- D. Deploy the model on an Amazon Elastic Container Service (Amazon ECS) cluster
- E. Use ECS scheduled scaling that is based on the CPU of the ECS cluster.
- F. Install SageMaker Operator on an Amazon Elastic Kubernetes Service (Amazon EKS) cluster
- G. Deploy the model in Amazon EK
- H. Set horizontal pod auto scaling to scale replicas based on the memory metric.
- I. Use Spot Instances with a Spot Fleet behind an Application Load Balancer (ALB) for inference
- J. Use the ALBRequestCountPerTarget metric as the metric for auto scaling.

Answer: A

NEW QUESTION 8

A company uses a hybrid cloud environment. A model that is deployed on premises uses data in Amazon S3 to provide customers with a live conversational engine. The model is using sensitive data. An ML engineer needs to implement a solution to identify and remove the sensitive data. Which solution will meet these requirements with the LEAST operational overhead?

- A. Deploy the model on Amazon SageMaker
- B. Create a set of AWS Lambda functions to identify and remove the sensitive data.
- C. Deploy the model on an Amazon Elastic Container Service (Amazon ECS) cluster that uses AWS Fargate
- D. Create an AWS Batch job to identify and remove the sensitive data.
- E. Use Amazon Macie to identify the sensitive data
- F. Create a set of AWS Lambda functions to remove the sensitive data.
- G. Use Amazon Comprehend to identify the sensitive data
- H. Launch Amazon EC2 instances to remove the sensitive data.

Answer: C

NEW QUESTION 9

A company has deployed an XGBoost prediction model in production to predict if a customer is likely to cancel a subscription. The company uses Amazon SageMaker Model Monitor to detect deviations in the F1 score. During a baseline analysis of model quality, the company recorded a threshold for the F1 score. After several months of no change, the model's F1 score decreases significantly. What could be the reason for the reduced F1 score?

- A. Concept drift occurred in the underlying customer data that was used for predictions.
- B. The model was not sufficiently complex to capture all the patterns in the original baseline data.
- C. The original baseline data had a data quality issue of missing values.
- D. Incorrect ground truth labels were provided to Model Monitor during the calculation of the baseline.

Answer: A

NEW QUESTION 10

A company has trained and deployed an ML model by using Amazon SageMaker. The company needs to implement a solution to record and monitor all the API call events for the SageMaker endpoint. The solution also must provide a notification when the number of API call events breaches a threshold. Use SageMaker Debugger to track the inferences and to report metrics. Create a custom rule to provide a notification when the threshold is breached. Which solution will meet these requirements?

- A. Use SageMaker Debugger to track the inferences and to report metric
- B. Create a custom rule to provide a notification when the threshold is breached.
- C. Use SageMaker Debugger to track the inferences and to report metric
- D. Use the tensor_variance built-in rule to provide a notification when the threshold is breached.
- E. Log all the endpoint invocation API events by using AWS CloudTrail
- F. Use an Amazon CloudWatch dashboard for monitoring
- G. Set up a CloudWatch alarm to provide notification when the threshold is breached.
- H. Add the Invocations metric to an Amazon CloudWatch dashboard for monitoring
- I. Set up a CloudWatch alarm to provide notification when the threshold is breached.

Answer: D

NEW QUESTION 10

A company has an ML model that generates text descriptions based on images that customers upload to the company's website. The images can be up to 50 MB in total size.

An ML engineer decides to store the images in an Amazon S3 bucket. The ML engineer must implement a processing solution that can scale to accommodate changes in demand.

Which solution will meet these requirements with the LEAST operational overhead?

- A. Create an Amazon SageMaker batch transform job to process all the images in the S3 bucket.
- B. Create an Amazon SageMaker Asynchronous Inference endpoint and a scaling policy.
- C. Run a script to make an inference request for each image.
- D. Create an Amazon Elastic Kubernetes Service (Amazon EKS) cluster that uses Karpenter for auto scaling.
- E. Host the model on the EKS cluster.
- F. Run a script to make an inference request for each image.
- G. Create an AWS Batch job that uses an Amazon Elastic Container Service (Amazon ECS) cluster.
- H. Specify a list of images to process for each AWS Batch job.

Answer: B

NEW QUESTION 11

An ML engineer has an Amazon Comprehend custom model in Account A in the us-east-1 Region. The ML engineer needs to copy the model to Account B in the same Region.

Which solution will meet this requirement with the LEAST development effort?

- A. Use Amazon S3 to make a copy of the model.
- B. Transfer the copy to Account B.
- C. Create a resource-based IAM policy.
- D. Use the Amazon Comprehend ImportModel API operation to copy the model to Account B.
- E. Use AWS DataSync to replicate the model from Account A to Account B.
- F. Create an AWS Site-to-Site VPN connection between Account A and Account B to transfer the model.

Answer: B

NEW QUESTION 12

An ML engineer is developing a fraud detection model by using the Amazon SageMaker XGBoost algorithm. The model classifies transactions as either fraudulent or legitimate.

During testing, the model excels at identifying fraud in the training dataset. However, the model is inefficient at identifying fraud in new and unseen transactions.

What should the ML engineer do to improve the fraud detection for new transactions?

- A. Increase the learning rate.
- B. Remove some irrelevant features from the training dataset.
- C. Increase the value of the max_depth hyperparameter.
- D. Decrease the value of the max_depth hyperparameter.

Answer: D

NEW QUESTION 13

Case Study

A company is building a web-based AI application by using Amazon SageMaker. The application will provide the following capabilities and features: ML experimentation, training, a central model registry, model deployment, and model monitoring.

The application must ensure secure and isolated use of training data during the ML lifecycle. The training data is stored in Amazon S3.

The company must implement a manual approval-based workflow to ensure that only approved models can be deployed to production endpoints.

Which solution will meet this requirement?

- A. Use SageMaker Experiments to facilitate the approval process during model registration.
- B. Use SageMaker ML Lineage Tracking on the central model registry.
- C. Create tracking entities for the approval process.
- D. Use SageMaker Model Monitor to evaluate the performance of the model and to manage the approval.
- E. Use SageMaker Pipeline.
- F. When a model version is registered, use the AWS SDK to change the approval status to "Approved."

Answer: D

NEW QUESTION 18

An ML engineer needs to use Amazon SageMaker to fine-tune a large language model (LLM) for text summarization. The ML engineer must follow a low-code no-code (LCNC) approach.

Which solution will meet these requirements?

- A. Use SageMaker Studio to fine-tune an LLM that is deployed on Amazon EC2 instances.
- B. Use SageMaker Autopilot to fine-tune an LLM that is deployed by a custom API endpoint.
- C. Use SageMaker Autopilot to fine-tune an LLM that is deployed on Amazon EC2 instances.
- D. Use SageMaker Autopilot to fine-tune an LLM that is deployed by SageMaker JumpStart.

Answer: D

NEW QUESTION 19

An ML engineer has trained a neural network by using stochastic gradient descent (SGD). The neural network performs poorly on the test set. The values for training loss and validation loss remain high and show an oscillating pattern. The values decrease for a few epochs and then increase for a few epochs before repeating the same cycle.

What should the ML engineer do to improve the training process?

- A. Introduce early stopping.
- B. Increase the size of the test set.
- C. Increase the learning rate.
- D. Decrease the learning rate.

Answer: D

NEW QUESTION 21

A company is using ML to predict the presence of a specific weed in a farmer's field. The company is using the Amazon SageMaker linear learner built-in algorithm with a value of multiclass_classifier for the predictor_type hyperparameter.

What should the company do to MINIMIZE false positives?

- A. Set the value of the weight decay hyperparameter to zero.
- B. Increase the number of training epochs.
- C. Increase the value of the target_precision hyperparameter.
- D. Change the value of the predictor_type hyperparameter to regressor.

Answer: C

NEW QUESTION 23

An ML engineer needs to use AWS services to identify and extract meaningful unique keywords from documents. Which solution will meet these requirements with the LEAST operational overhead?

- A. Use the Natural Language Toolkit (NLTK) library on Amazon EC2 instances for text pre-processing
- B. Use the Latent Dirichlet Allocation (LDA) algorithm to identify and extract relevant keywords.
- C. Use Amazon SageMaker and the BlazingText algorithm
- D. Apply custom pre-processing steps for stemming and removal of stop word
- E. Calculate term frequency-inverse document frequency (TF-IDF) scores to identify and extract relevant keywords.
- F. Store the documents in an Amazon S3 bucket
- G. Create AWS Lambda functions to process the documents and to run Python scripts for stemming and removal of stop word
- H. Use bigram and trigram techniques to identify and extract relevant keywords.
- I. Use Amazon Comprehend custom entity recognition and key phrase extraction to identify and extract relevant keywords.

Answer: D

NEW QUESTION 26

An ML engineer is evaluating several ML models and must choose one model to use in production. The cost of false negative predictions by the models is much higher than the cost of false positive predictions.

Which metric finding should the ML engineer prioritize the MOST when choosing the model?

- A. Low precision
- B. High precision
- C. Low recall
- D. High recall

Answer: D

NEW QUESTION 28

A company needs to create a central catalog for all the company's ML models. The models are in AWS accounts where the company developed the models initially. The models are hosted in Amazon Elastic Container Registry (Amazon ECR) repositories.

Which solution will meet these requirements?

- A. Configure ECR cross-account replication for each existing ECR repository
- B. Ensure that each model is visible in each AWS account.
- C. Create a new AWS account with a new ECR repository as the central catalog
- D. Configure ECR cross-account replication between the initial ECR repositories and the central catalog.
- E. Use the Amazon SageMaker Model Registry to create a model group for models hosted in Amazon EC
- F. Create a new AWS account
- G. In the new account, use the SageMaker Model Registry as the central catalog
- H. Attach a cross-account resource policy to each model group in the initial AWS accounts.
- I. Use an AWS Glue Data Catalog to store the model
- J. Run an AWS Glue crawler to migrate the models from the ECR repositories to the Data Catalog
- K. Configure cross-account access to the Data Catalog.

Answer: C

NEW QUESTION 33

A company regularly receives new training data from the vendor of an ML model. The vendor delivers cleaned and prepared data to the company's Amazon S3 bucket every 3-4 days.

The company has an Amazon SageMaker pipeline to retrain the model. An ML engineer needs to implement a solution to run the pipeline when new data is uploaded to the S3 bucket.

Which solution will meet these requirements with the LEAST operational effort?

- A. Create an S3 Lifecycle rule to transfer the data to the SageMaker training instance and to initiate training.
 - B. Create an AWS Lambda function that scans the S3 bucket.
 - C. Program the Lambda function to initiate the pipeline when new data is uploaded.
 - D. Create an Amazon EventBridge rule that has an event pattern that matches the S3 upload.
 - E. Configure the pipeline as the target of the rule.
 - F. Use Amazon Managed Workflows for Apache Airflow (Amazon MWAA) to orchestrate the pipeline when new data is uploaded.
- A company wants to develop an ML model by using tabular data from its customer.
- G. The data contains meaningful ordered features with sensitive information that should not be discarded.
 - H. An ML engineer must ensure that the sensitive data is masked before another team starts to build the model. Which solution will meet these requirements?
 - I. Use Amazon Macie to categorize the sensitive data.
 - J. Prepare the data by using AWS Glue DataBrew.
 - K. Run an AWS Batch job to change the sensitive data to random values.
 - L. Run an Amazon EMR job to change the sensitive data to random values.

Answer: B

NEW QUESTION 35

Case study

An ML engineer is developing a fraud detection model on AWS. The training dataset includes transaction logs, customer profiles, and tables from an on-premises MySQL database. The transaction logs and customer profiles are stored in Amazon S3.

The dataset has a class imbalance that affects the learning of the model's algorithm. Additionally, many of the features have interdependencies. The algorithm is not capturing all the desired underlying patterns in the data.

Which AWS service or feature can aggregate the data from the various data sources?

- A. Amazon EMR Spark jobs
- B. Amazon Kinesis Data Streams
- C. Amazon DynamoDB
- D. AWS Lake Formation

Answer: A

NEW QUESTION 37

A company has implemented a data ingestion pipeline for sales transactions from its ecommerce website. The company uses Amazon Data Firehose to ingest data into Amazon OpenSearch Service. The buffer interval of the Firehose stream is set for 60 seconds. An OpenSearch linear model generates real-time sales forecasts based on the data and presents the data in an OpenSearch dashboard.

The company needs to optimize the data ingestion pipeline to support sub-second latency for the real-time dashboard.

Which change to the architecture will meet these requirements?

- A. Use zero buffering in the Firehose stream.
- B. Tune the batch size that is used in the PutRecordBatch operation.
- C. Replace the Firehose stream with an AWS DataSync task.
- D. Configure the task with enhanced fan-out consumers.
- E. Increase the buffer interval of the Firehose stream from 60 seconds to 120 seconds.
- F. Replace the Firehose stream with an Amazon Simple Queue Service (Amazon SQS) queue.

Answer: A

NEW QUESTION 40

Case study

An ML engineer is developing a fraud detection model on AWS. The training dataset includes transaction logs, customer profiles, and tables from an on-premises MySQL database. The transaction logs and customer profiles are stored in Amazon S3.

The dataset has a class imbalance that affects the learning of the model's algorithm. Additionally, many of the features have interdependencies. The algorithm is not capturing all the desired underlying patterns in the data.

After the data is aggregated, the ML engineer must implement a solution to automatically detect anomalies in the data and to visualize the result.

Which solution will meet these requirements?

- A. Use Amazon Athena to automatically detect the anomalies and to visualize the result.
- B. Use Amazon Redshift Spectrum to automatically detect the anomalies.
- C. Use Amazon QuickSight to visualize the result.
- D. Use Amazon SageMaker Data Wrangler to automatically detect the anomalies and to visualize the result.
- E. Use AWS Batch to automatically detect the anomalies.
- F. Use Amazon QuickSight to visualize the result.

Answer: C

NEW QUESTION 43

Case study

An ML engineer is developing a fraud detection model on AWS. The training dataset includes transaction logs, customer profiles, and tables from an on-premises MySQL database. The transaction logs and customer profiles are stored in Amazon S3.

The dataset has a class imbalance that affects the learning of the model's algorithm. Additionally, many of the features have interdependencies. The algorithm is not capturing all the desired underlying patterns in the data.

The training dataset includes categorical data and numerical data. The ML engineer must prepare the training dataset to maximize the accuracy of the model.

Which action will meet this requirement with the LEAST operational overhead?

- A. Use AWS Glue to transform the categorical data into numerical data.
- B. Use AWS Glue to transform the numerical data into categorical data.
- C. Use Amazon SageMaker Data Wrangler to transform the categorical data into numerical data.
- D. Use Amazon SageMaker Data Wrangler to transform the numerical data into categorical data.

Answer: C

NEW QUESTION 48

An ML engineer has developed a binary classification model outside of Amazon SageMaker. The ML engineer needs to make the model accessible to a SageMaker Canvas user for additional tuning.

The model artifacts are stored in an Amazon S3 bucket. The ML engineer and the Canvas user are part of the same SageMaker domain.

Which combination of requirements must be met so that the ML engineer can share the model with the Canvas user? (Choose two.)

- A. The ML engineer and the Canvas user must be in separate SageMaker domains.
- B. The Canvas user must have permissions to access the S3 bucket where the model artifacts are stored.
- C. The model must be registered in the SageMaker Model Registry.
- D. The ML engineer must host the model on AWS Marketplace.
- E. The ML engineer must deploy the model to a SageMaker endpoint.

Answer: BC

NEW QUESTION 51

An ML engineer receives datasets that contain missing values, duplicates, and extreme outliers. The ML engineer must consolidate these datasets into a single data frame and must prepare the data for ML.

Which solution will meet these requirements?

- A. Use Amazon SageMaker Data Wrangler to import the datasets and to consolidate them into a single data frame
- B. Use the cleansing and enrichment functionalities to prepare the data.
- C. Use Amazon SageMaker Ground Truth to import the datasets and to consolidate them into a single data frame
- D. Use the human-in-the-loop capability to prepare the data.
- E. Manually import and merge the dataset
- F. Consolidate the datasets into a single data frame
- G. Use Amazon Q Developer to generate code snippets that will prepare the data.
- H. Manually import and merge the dataset
- I. Consolidate the datasets into a single data frame
- J. Use Amazon SageMaker data labeling to prepare the data.

Answer: A

NEW QUESTION 56

A company that has hundreds of data scientists is using Amazon SageMaker to create ML models. The models are in model groups in the SageMaker Model Registry.

The data scientists are grouped into three categories: computer vision, natural language processing (NLP), and speech recognition. An ML engineer needs to implement a solution to organize the existing models into these groups to improve model discoverability at scale. The solution must not affect the integrity of the model artifacts and their existing groupings.

Which solution will meet these requirements?

- A. Create a custom tag for each of the three categories
- B. Add the tags to the model packages in the SageMaker Model Registry.
- C. Create a model group for each category
- D. Move the existing models into these category model groups.
- E. Use SageMaker ML Lineage Tracking to automatically identify and tag which model groups should contain the models.
- F. Create a Model Registry collection for each of the three categories
- G. Move the existing model groups into the collections.

Answer: A

NEW QUESTION 57

.....

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questions and Answers in PDF Format

MLA-C01 Practice Exam Features:

- * MLA-C01 Questions and Answers Updated Frequently
- * MLA-C01 Practice Questions Verified by Expert Senior Certified Staff
- * MLA-C01 Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- * MLA-C01 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

100% Actual & Verified — Instant Download, Please Click
[Order The MLA-C01 Practice Test Here](#)