Amazon-Web-Services

Exam Questions MLA-C01

AWS Certified Machine Learning Engineer - Associate

NEW QUESTION 1
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 2
An ML engineer needs to process thousands of existing CSV objects and new CSV objects that are uploaded. The CSV objects are stored in a central Amazon S3 bucket and have the same number of columns. One of the columns is a transaction date. The ML engineer must query the data based on the transaction date.
Which solution will meet these requirements with the LEAST operational overhead?

A. Use an Amazon Athena CREATE TABLE AS SELECT (CTAS) statement to create a table based on the transaction date from data in the central S3 bucke
B. Query the objects from the table.
C. Create a new S3 bucket for processed dat
D. Set up S3 replication from the central S3 bucket to the new S3 bucke
E. Use S3 Object Lambda to query the objects based on transaction date.
F. Create a new S3 bucket for processed dat
G. Use AWS Glue for Apache Spark to create a job to query the CSV objects based on transaction dat
H. Configure the job to store the results in the new S3 bucke
I. Query the objects from the new S3 bucket.
J. Create a new S3 bucket for processed dat
K. Use Amazon Data Firehose to transfer the data from the central S3 bucket to the new S3 bucke
L. Configure Firehose to run an AWS Lambda function to query the data based on transaction date.

Answer: A

NEW QUESTION 3
An ML engineer needs to use data with Amazon SageMaker Canvas to train an ML model. The data is stored in Amazon S3 and is complex in structure. The ML engineer must use a file format that minimizes processing time for the data.
Which file format will meet these requirements?

A. CSV files compressed with Snappy
B. JSON objects in JSONL format
C. JSON files compressed with gzip
D. Apache Parquet files

Answer: D

NEW QUESTION 4
A company is planning to use Amazon SageMaker to make classification ratings that are based on images. The company has 6 of training data that is stored on an Amazon FSx for NetApp ONTAP system virtual machine (SVM). The SVM is in the same VPC as SageMaker.
An ML engineer must make the training data accessible for ML models that are in the SageMaker environment.
Which solution will meet these requirements?

A. Mount the FSx for ONTAP file system as a volume to the SageMaker Instance.
B. Create an Amazon S3 bucke
C. Use Mountpoint for Amazon S3 to link the S3 bucket to the FSx for ONTAP file system.
D. Create a catalog connection from SageMaker Data Wrangler to the FSx for ONTAP file system.
E. Create a direct connection from SageMaker Data Wrangler to the FSx for ONTAP file system.

Answer: A

NEW QUESTION 5
A company has an application that uses different APIs to generate embeddings for input text. The company needs to implement a solution to automatically rotate the API tokens every 3 months.
Which solution will meet this requirement?

A. Store the tokens in AWS Secrets Manage
B. Create an AWS Lambda function to perform the rotation.
C. Store the tokens in AWS Systems Manager Parameter Stor
D. Create an AWS Lambda function to perform the rotation.
E. Store the tokens in AWS Key Management Service (AWS KMS). Use an AWS managed key to perform the rotation.
F. Store the tokens in AWS Key Management Service (AWS KMS). Use an AWS owned key to perform the rotation.

Answer: A

NEW QUESTION 6
A company has an ML model that needs to run one time each night to predict stock values. The model input is 3 MB of data that is collected during the current day. The model produces the predictions for the next day. The prediction process takes less than 1 minute to finish running.
How should the company deploy the model on Amazon SageMaker to meet these requirements?

A. Use a multi-model serverless endpoin
B. Enable caching.
C. Use an asynchronous inference endpoin
D. Set the InitialInstanceCount parameter to 0.
E. Use a real-time endpoin
F. Configure an auto scaling policy to scale the model to 0 when the model is not in use.
G. Use a serverless inference endpoin
H. Set the MaxConcurrency parameter to 1.

Answer: D

NEW QUESTION 7
A company has a team of data scientists who use Amazon SageMaker notebook instances to test ML models. When the data scientists need new permissions, the company attaches the permissions to each individual role that was created during the creation of the SageMaker notebook instance.
The company needs to centralize management of the team's permissions. Which solution will meet this requirement?

A. Create a single IAM role that has the necessary permission
B. Attach the role to each notebook instance that the team uses.
C. Create a single IAM grou
D. Add the data scientists to the grou
E. Associate the group with each notebook instance that the team uses.
F. Create a single IAM use
G. Attach the AdministratorAccess AWS managed IAM policy to the use
H. Configure each notebook instance to use the IAM user.
I. Create a single IAM grou
J. Add the data scientists to the grou
K. Create an IAM rol
L. Attach the AdministratorAccess AWS managed IAM policy to the rol
M. Associate the role with the grou
N. Associate the group with each notebook instance that the team uses.

Answer: A

NEW QUESTION 8
A company has a large collection of chat recordings from customer interactions after a product release. An ML engineer needs to create an ML model to analyze the chat data. The ML engineer needs to determine the success of the product by reviewing customer sentiments about the product.
Which action should the ML engineer take to complete the evaluation in the LEAST amount of time?

A. Use Amazon Rekognition to analyze sentiments of the chat conversations.
B. Train a Naive Bayes classifier to analyze sentiments of the chat conversations.
C. Use Amazon Comprehend to analyze sentiments of the chat conversations.
D. Use random forests to classify sentiments of the chat conversations.

Answer: C

NEW QUESTION 9
A company has a large, unstructured dataset. The dataset includes many duplicate records across several key attributes.
Which solution on AWS will detect duplicates in the dataset with the LEAST code development?

A. Use Amazon Mechanical Turk jobs to detect duplicates.
B. Use Amazon QuickSight ML Insights to build a custom deduplication model.
C. Use Amazon SageMaker Data Wrangler to pre-process and detect duplicates.
D. Use the AWS Glue FindMatches transform to detect duplicates.

Answer: D

NEW QUESTION 10
A company uses a hybrid cloud environment. A model that is deployed on premises uses data in Amazon 53 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 SageMake
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 Fargat
D. Create an AWS Batch job to identify and remove the sensitive data.
E. Use Amazon Macie to identify the sensitive dat
F. Create a set of AWS Lambda functions to remove the sensitive data.
G. Use Amazon Comprehend to identify the sensitive dat
H. Launch Amazon EC2 instances to remove the sensitive data.

Answer: C

NEW QUESTION 10
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 12
A company runs an Amazon SageMaker domain in a public subnet of a newly created VPC. The network is configured properly, and ML engineers can access the SageMaker domain.
Recently, the company discovered suspicious traffic to the domain from a specific IP address. The company needs to block traffic from the specific IP address.
Which update to the network configuration will meet this requirement?

A. Create a security group inbound rule to deny traffic from the specific IP addres
B. Assign the security group to the domain.
C. Create a network ACL inbound rule to deny traffic from the specific IP addres
D. Assign the rule to the default network Ad for the subnet where the domain is located.
E. Create a shadow variant for the domai
F. Configure SageMaker Inference Recommender to send traffic from the specific IP address to the shadow endpoint.
G. Create a VPC route table to deny inbound traffic from the specific IP addres
H. Assign the route table to the domain.

Answer: B

NEW QUESTION 14
A company is creating an application that will recommend products for customers to purchase. The application will make API calls to Amazon Q Business. The company must ensure that responses from Amazon Q Business do not include the name of the company's main competitor.
Which solution will meet this requirement?

A. Configure the competitor's name as a blocked phrase in Amazon Q Business.
B. Configure an Amazon Q Business retriever to exclude the competitor??s name.
C. Configure an Amazon Kendra retriever for Amazon Q Business to build indexes that exclude the competitor's name.
D. Configure document attribute boosting in Amazon Q Business to deprioritize the competitor's name.

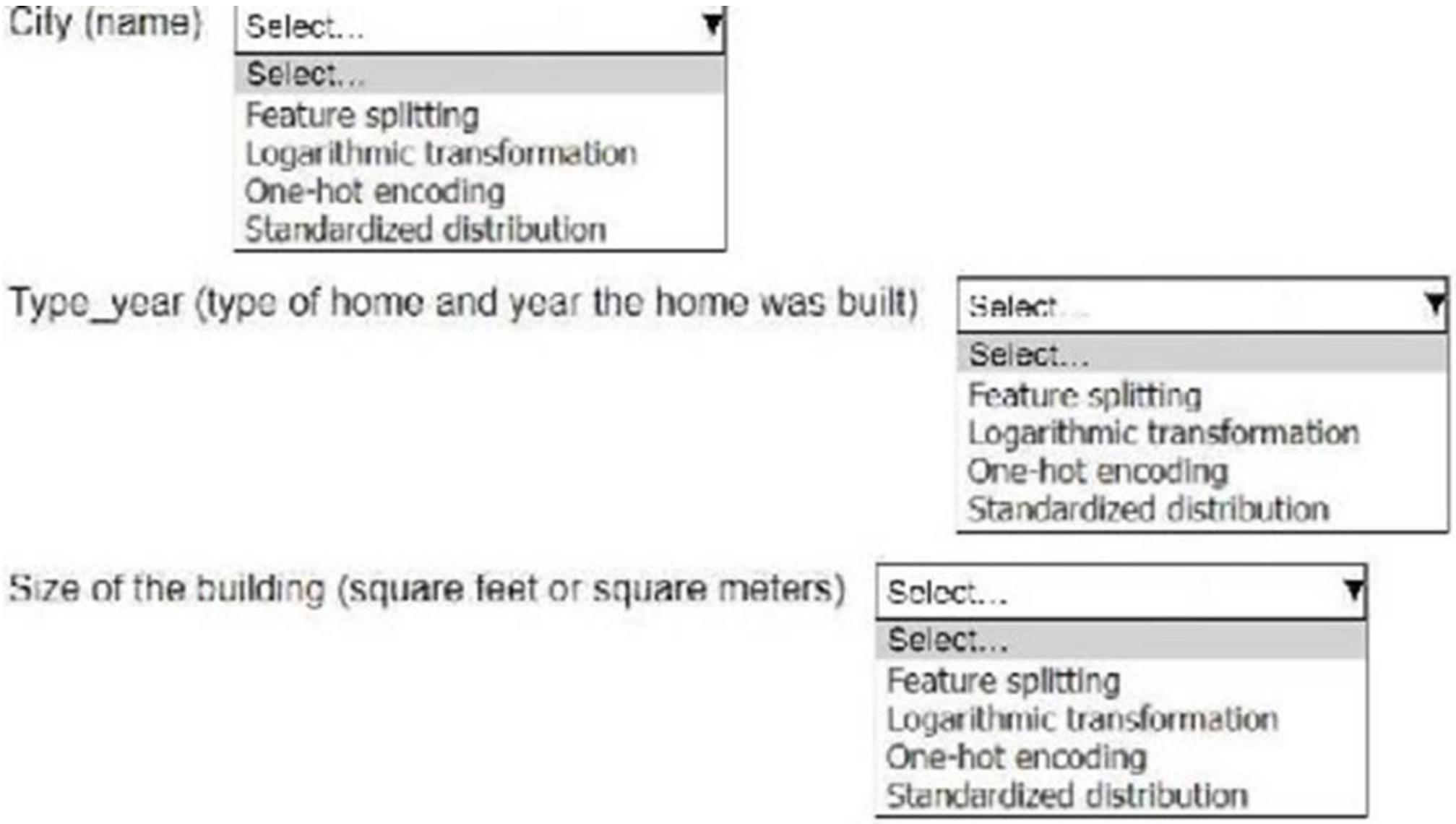
Answer: A

NEW QUESTION 18
A company has a conversational AI assistant that sends requests through Amazon Bedrock to an Anthropic Claude large language model (LLM). Users report that when they ask similar questions multiple times, they sometimes receive different answers. An ML engineer needs to improve the responses to be more consistent and less random.
Which solution will meet these requirements?

A. Increase the temperature parameter and the top_k parameter.
B. Increase the temperature paramete
C. Decrease the top_k parameter.
D. Decrease the temperature paramete
E. Increase the top_k parameter.
F. Decrease the temperature parameter and the top_k parameter.

Answer: D

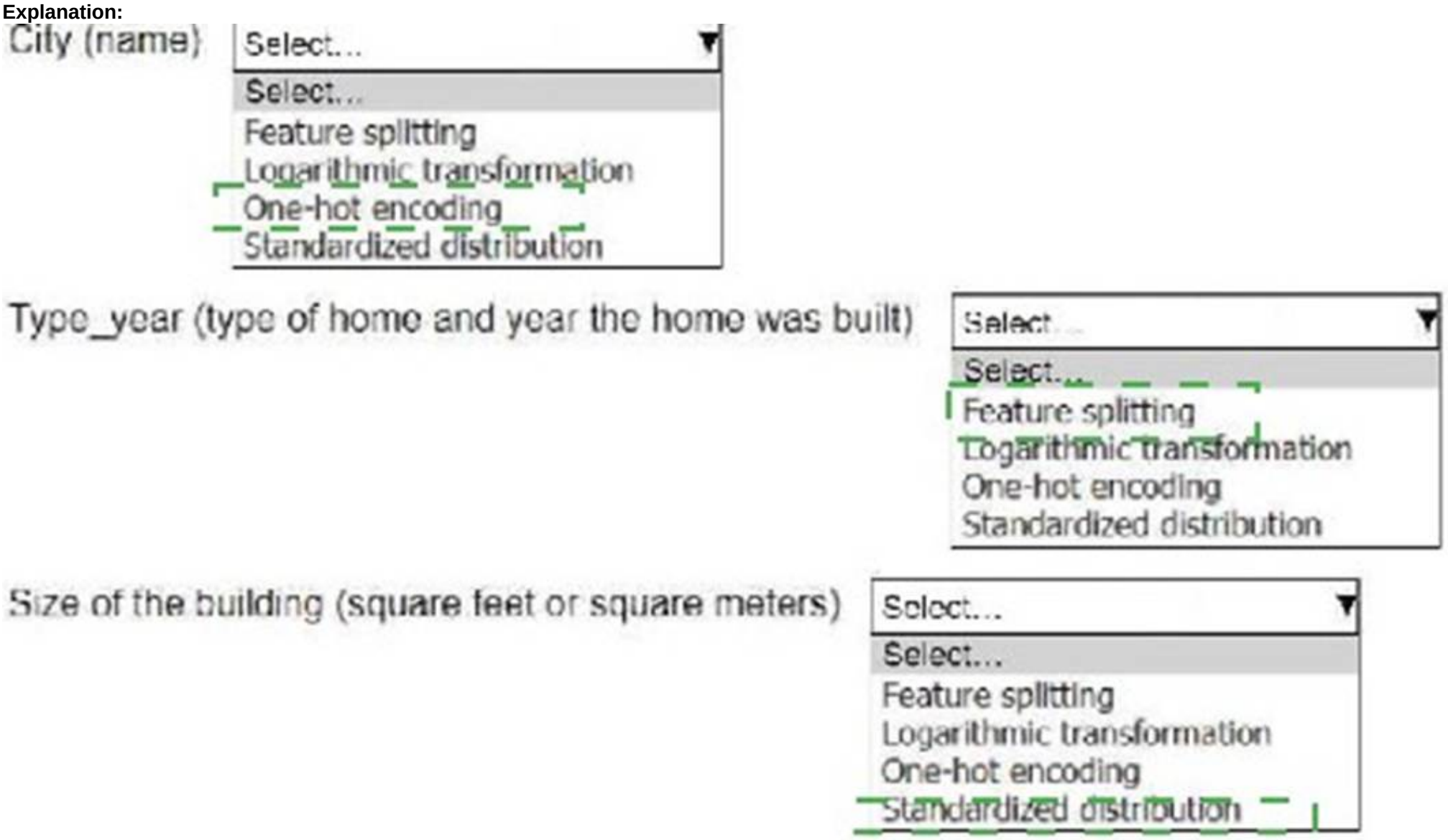
NEW QUESTION 23
HOTSPOT
An ML engineer is working on an ML model to predict the prices of similarly sized homes. The model will base predictions on several features The ML engineer will use the following feature engineering techniques to estimate the prices of the homes:
• Feature splitting
• Logarithmic transformation
• One-hot encoding
• Standardized distribution
Select the correct feature engineering techniques for the following list of features. Each feature engineering technique should be selected one time or not at all (Select three.)



A. Mastered
B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 26
A company has used Amazon SageMaker to deploy a predictive ML model in production. The company is using SageMaker Model Monitor on the model. After a model update, an ML engineer notices data quality issues in the Model Monitor checks.
What should the ML engineer do to mitigate the data quality issues that Model Monitor has identified?

A. Adjust the model's parameters and hyperparameters.

B. Initiate a manual Model Monitor job that uses the most recent production data.
C. Create a new baseline from the latest datase
D. Update Model Monitor to use the new baseline for evaluations.
E. Include additional data in the existing training set for the mode
F. Retrain and redeploy the model.

Answer: C

NEW QUESTION 30
A company needs to give its ML engineers appropriate access to training data. The ML engineers must access training data from only their own business group. The ML engineers must not be allowed to access training data from other business groups.
The company uses a single AWS account and stores all the training data in Amazon S3 buckets. All ML model training occurs in Amazon SageMaker.
Which solution will provide the ML engineers with the appropriate access?

A. Enable S3 bucket versioning.
B. Configure S3 Object Lock settings for each user.
C. Add cross-origin resource sharing (CORS) policies to the S3 buckets.
D. Create IAM policie
E. Attach the policies to IAM users or IAM roles.

Answer: D

NEW QUESTION 35
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 36
FILL IN THE BLANK
A company stores time-series data about user clicks in an Amazon S3 bucket. The raw data consists of millions of rows of user activity every day. ML engineers access the data to develop their ML models.
The ML engineers need to generate daily reports and analyze click trends over the past 3 days by using Amazon Athena. The company must retain the data for 30 days before archiving the data.
Which solution will provide the HIGHEST performance for data retrieval?

A. Keep all the time-series data without partitioning in the S3 bucke
B. Manually move data that is older than 30 days to separate S3 buckets.
C. Create AWS Lambda functions to copy the time-series data into separate S3 bucket
D. Apply S3 Lifecycle policies to archive data that is older than 30 days to S3 Glacier Flexible Retrieval.
E. Organize the time-series data into partitions by date prefix in the S3 bucke
F. Apply S3 Lifecycle policies to archive partitions that are older than 30 days to S3 Glacier Flexible Retrieval.
G. Put each day's time-series data into its own S3 bucke
H. Use S3 Lifecycle policies to archive S3 buckets that hold data that is older than 30 days to S3 Glacier Flexible Retrieval.

Answer: C

NEW QUESTION 37
A company needs to run a batch data-processing job on Amazon EC2 instances. The job will run during the weekend and will take 90 minutes to finish running. The processing can
handle interruptions. The company will run the job every weekend for the next 6 months.
Which EC2 instance purchasing option will meet these requirements MOST cost- effectively?

A. Spot Instances
B. Reserved Instances
C. On-Demand Instances
D. Dedicated Instances

Answer: A

NEW QUESTION 40
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- processin
B. Use the Latent Dirichlet Allocation (LDA) algorithm to identify and extract relevant keywords.
C. Use Amazon SageMaker and the BlazingText algorith
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 bucke
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 41
An ML engineer normalized training data by using min-max normalization in AWS Glue DataBrew. The ML engineer must normalize the production inference data in the same way as the training data before passing the production inference data to the model for predictions.
Which solution will meet this requirement?

A. Apply statistics from a well-known dataset to normalize the production samples.
B. Keep the min-max normalization statistics from the training se
C. Use these values to normalize the production samples.
D. Calculate a new set of min-max normalization statistics from a batch of production sample
E. Use these values to normalize all the production samples.
F. Calculate a new set of min-max normalization statistics from each production sampl
G. Use these values to normalize all the production samples.

Answer: B

NEW QUESTION 45
An ML engineer needs to use AWS CloudFormation to create an ML model that an Amazon SageMaker endpoint will host.
Which resource should the ML engineer declare in the CloudFormation template to meet this requirement?

A. AWS::SageMaker::Model
B. AWS::SageMaker::Endpoint
C. AWS::SageMaker::NotebookInstance
D. AWS::SageMaker::Pipeline

Answer: A

NEW QUESTION 49
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 bucke
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 uploa
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 discarde
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 Made 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 54
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 strea
B. Tune the batch size that is used in the PutRecordBatch operation.
C. Replace the Firehose stream with an AWS DataSync tas
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 59
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 ML engineer needs to use an Amazon SageMaker built-in algorithm to train the model. Which algorithm should the ML engineer use to meet this requirement?

A. LightGBM

B. Linear learner
C. -means clustering
D. Neural Topic Model (NTM)

Answer: B

NEW QUESTION 62
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 anomalie
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 anomalie
F. Use Amazon QuickSight to visualize the result.

Answer: C

NEW QUESTION 67
A company wants to predict the success of advertising campaigns by considering the color scheme of each advertisement. An ML engineer is preparing data for a neural network model. The dataset includes color information as categorical data.
Which technique for feature engineering should the ML engineer use for the model?

A. Apply label encoding to the color categorie
B. Automatically assign each color a unique integer.
C. Implement padding to ensure that all color feature vectors have the same length.
D. Perform dimensionality reduction on the color categories.
E. One-hot encode the color categories to transform the color scheme feature into a binary matrix.

Answer: D

NEW QUESTION 71
An ML engineer needs to use an Amazon EMR cluster to process large volumes of data in batches. Any data loss is unacceptable.
Which instance purchasing option will meet these requirements MOST cost-effectively?

A. Run the primary node, core nodes, and task nodes on On-Demand Instances.
B. Run the primary node, core nodes, and task nodes on Spot Instances.
C. Run the primary node on an On-Demand Instanc
D. Run the core nodes and task nodes on Spot Instances.
E. Run the primary node and core nodes on On-Demand Instance
F. Run the task nodes on Spot Instances.

Answer: D

NEW QUESTION 76
A company wants to improve the sustainability of its ML operations.
Which actions will reduce the energy usage and computational resources that are associated with the company's training jobs? (Choose two.)

A. Use Amazon SageMaker Debugger to stop training jobs when non-converging conditions are detected.
B. Use Amazon SageMaker Ground Truth for data labeling.
C. Deploy models by using AWS Lambda functions.
D. Use AWS Trainium instances for training.
E. Use PyTorch or TensorFlow with the distributed training option.

Answer: AD

NEW QUESTION 77
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 categorie
B. Add the tags to the model packages in the SageMaker Model Registry.
C. Create a model group for each categor
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 categorie
G. Move the existing model groups into the collections.

Answer: A

NEW QUESTION 79
A credit card company has a fraud detection model in production on an Amazon SageMaker endpoint. The company develops a new version of the model. The company needs to assess the new model's performance by using live data and without affecting production end users.
Which solution will meet these requirements?

A. Set up SageMaker Debugger and create a custom rule.
B. Set up blue/green deployments with all-at-once traffic shifting.
C. Set up blue/green deployments with canary traffic shifting.
D. Set up shadow testing with a shadow variant of the new model.

Answer: D

NEW QUESTION 83
.....

Thank You for Trying Our Product

We offer two products:

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

2nd - Questons 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