Exam   :   **MLA-C01**

Title   :   AWS Certified Machine
Learning Engineer -
Associate

Vendor   :   Amazon

Version   :   V12.65

QUESTION NO: 1

A company plans to deploy an ML model for production inference on an Amazon SageMaker endpoint. The average inference payload size will vary from 100 MB to 300 MB. Inference requests must be processed in 60 minutes or less.

Which SageMaker inference option will meet these requirements?

A. Serverless inference

B. Asynchronous inference

C. Real-time inference

D. Batch transform

Answer: B

QUESTION NO: 2

An ML engineer needs to encrypt all data in transit when an ML training job runs. The ML engineer must ensure that encryption in transit is applied to processes that Amazon SageMaker uses during the training job.

Which solution will meet these requirements?

A. Encrypt communication between nodes for batch processing.

B. Encrypt communication between nodes in a training cluster.

C. Specify an AWS Key Management Service (AWS KMS) key during creation of the training job request.

D. Specify an AWS Key Management Service (AWS KMS) key during creation of the SageMaker domain.

Answer: B

QUESTION NO: 3

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

QUESTION NO: 4

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_dassifier for the predictorjype 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 predictorjype hyperparameter to regressor.

Answer: C

QUESTION NO: 5

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 endpoint. Enable caching.

B. Use an asynchronous inference endpoint. Set the InitialInstanceCount parameter to 0.

C. Use a real-time endpoint. Configure an auto scaling policy to scale the model to 0 when the model is not in use.

D. Use a serverless inference endpoint. Set the MaxConcurrency parameter to 1.

Answer: D

QUESTION NO: 6

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

QUESTION NO: 7

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

QUESTION NO: 8

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

QUESTION NO: 9

A company uses Amazon SageMaker Studio to develop an ML model. The company has a single SageMaker Studio domain. An ML engineer needs to implement a solution that provides an automated alert when SageMaker compute costs reach a specific threshold.

Which solution will meet these requirements?

A. Add resource tagging by editing the SageMaker user profile in the SageMaker domain. Configure AWS Cost Explorer to send an alert when the threshold is reached.

B. Add resource tagging by editing the SageMaker user profile in the SageMaker domain. Configure AWS Budgets to send an alert when the threshold is reached.

C. Add resource tagging by editing each user's IAM profile. Configure AWS Cost Explorer to send an alert when the threshold is reached.

D. Add resource tagging by editing each user's IAM profile. Configure AWS Budgets to send an alert when the threshold is reached.

Answer: B

QUESTION NO: 10

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 database. Grant the necessary read permissions to the data scientist.

C. Unload the Amazon Redshift data to Amazon S3. Use Amazon Athena to create schema-on-read with masking logic. Share the view with the data scientist.

D. Unload the Amazon Redshift data to Amazon S3. Create an AWS Glue job to anonymize the data. Share the dataset with the data scientist.

Answer: A

QUESTION NO: 11

A company wants to build a real-time analytics application that uses streaming data from social media. An ML engineer must implement a solution that ingests and transforms 5 GB of data each minute. The solution also must load the data into a data store that supports fast queries for the real-time analytics. Which solution will meet these requirements?

A. Use Amazon EventBridge to ingest the social media data. Use AWS Glue to transform the data.
Store the transformed data in Amazon ElastiCache (Memcached).

B. Use Amazon Simple Queue Service (Amazon SQS) to ingest the social media data. Use AWS Lambda to transform the data. Store the transformed data in Amazon S3.

C. Use Amazon Simple Notification Service (Amazon SNS) to ingest the social media data. Use Amazon EMR to transform the data. Store the transformed data in Amazon RDS.

D. Use Amazon Kinesis Data Streams to ingest the social media data. Use Amazon Managed Service for Apache Flink to transform the data. Store the transformed data in Amazon DynamoDB.

Answer: D

Explanation:

Amazon Kinesis Data Streams is designed for high-throughput ingestion of streaming data such as social media feeds. Amazon Managed Service for Apache Flink enables real-time transformations on that data. Amazon DynamoDB provides low-latency reads and writes, making it suitable for fast queries in real-time analytics. This combination fully meets the scale and speed requirements.

QUESTION NO: 12

Hotspot Question
An ML engineer must choose the appropriate Amazon SageMaker algorithm to solve specific AI problems.
Select the correct SageMaker built-in algorithm from the following list for each use case.
Each algorithm should be selected one time.
- Random Cut Forest (RCF) algorithm
- Semantic segmentation algorithm
- Sequence-to-Sequence (seq2seq) algorithm

Summarize the text of a research paper.

| Random Cut Forest (RCF) algorithm |
| Semantic segmentation algorithm |
| Sequence-to-Sequence (seq2seq) algorithm |

Scan every pixel of an image to help
self-driving cars identify objects in their path.

| Random Cut Forest (RCF) algorithm |
| Semantic segmentation algorithm |
| Sequence-to-Sequence (seq2seq) algorithm |

Identify abnormal data points in a dataset.

| Random Cut Forest (RCF) algorithm |
| Semantic segmentation algorithm |
| Sequence-to-Sequence (seq2seq) algorithm |

Answer:

Summarize the text of a research paper.

| Random Cut Forest (RCF) algorithm |
| Semantic segmentation algorithm |
| **Sequence-to-Sequence (seq2seq) algorithm** |

Scan every pixel of an image to help
self-driving cars identify objects in their path.

| Random Cut Forest (RCF) algorithm |
| **Semantic segmentation algorithm** |
| Sequence-to-Sequence (seq2seq) algorithm |

Identify abnormal data points in a dataset.

| **Random Cut Forest (RCF) algorithm** |
| Semantic segmentation algorithm |
| Sequence-to-Sequence (seq2seq) algorithm |

QUESTION NO: 13

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

QUESTION NO: 14

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 policies. Attach the policies to IAM users or IAM roles.

Answer: D

QUESTION NO: 15

A company is planning to create an internal-only chat interface to help employees handle customer queries. Currently, the employees need to refer to a massive knowledge base of internal documents to address customer issues. The new solution must be serverless. Which combination of steps will meet these requirements?

A. Set up Amazon Bedrock with the Anthropic Claude foundation model.

B. Set up Amazon SageMaker JumpStart with the Llama foundation model.

C. Use Amazon EC2 instances with Amazon API Gateway to invoke the model API.

D. Use AWS Lambda functions with Amazon API Gateway to invoke the model API.

E. Use an Amazon S3 bucket to store vector database dumps and embeddings.

F. Use Amazon RDS for MySQL to store vector database dumps and embeddings.

Answer: ADE

Explanation:

To build a serverless internal chat interface, you can use Amazon Bedrock with a foundation model like Claude, invoke the model API through AWS Lambda with Amazon API Gateway, and store embeddings in a vector database format using Amazon S3. This avoids server management, ensures scalability, and leverages serverless components end-to-end.

QUESTION NO: 16

A company is planning to use Amazon SageMaker to make classification ratings that are based on images. The company has 6 GB 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 bucket. Use Mountpoint for Amazon S3 to link the S3 bucket to the FSx for ONTAP file system.

C. Create a catalog connection from SageMaker Data Wrangler to the FSx for ONTAP file system.

D. Create a direct connection from SageMaker Data Wrangler to the FSx for ONTAP file system.

Answer: A

QUESTION NO: 17

An ML engineer needs to merge and transform data from two sources to retrain an existing ML model. One data source consists of .csv files that are stored in an Amazon S3 bucket. Each .csv file consists of millions of records. The other data source is an Amazon Aurora DB cluster.

The result of the merge process must be written to a second S3 bucket. The ML engineer needs to perform this merge-and-transform task every week.

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

A. Create a transient Amazon EMR cluster every week. Use the cluster to run an Apache Spark job to merge and transform the data.

B. Create a weekly AWS Glue job that uses the Apache Spark engine. Use DynamicFrame native operations to merge and transform the data.

C. Create an AWS Lambda function that runs Apache Spark code every week to merge and transform the data. Configure the Lambda function to connect to the initial S3 bucket and the DB cluster.

D. Create an AWS Batch job that runs Apache Spark code on Amazon EC2 instances every week.Configure the Spark code to save the data from the EC2 instances to the second S3 bucket.

Answer: B

QUESTION NO: 18

A company is developing a new online application to gather information from customers. An ML engineer has developed a new ML model that will determine a score for each customer. The model will use the score to determine which product to display to the customer. The ML engineer needs to minimize response-time latency for the model. How should the ML engineer deploy the application in Amazon SageMaker to meet these requirements?

A. Configure batch transform.

B. Configure a real-time inference endpoint.

C. Configure a serverless inference endpoint.

D. Configure an asynchronous inference endpoint.

Answer: B

Explanation:

To minimize response-time latency, the ML model should be deployed to a real-time inference endpoint in Amazon SageMaker. This provides low-latency predictions by keeping the model loaded and ready to handle incoming requests, which is critical for an online application serving customers in real time.

QUESTION NO: 19

An ML engineer is developing a classification model. The ML engineer needs to use custom libraries in processing jobs, training jobs, and pipelines in Amazon SageMaker. Which solution will provide this functionality with the LEAST implementation effort?

A. Manually install the libraries in the SageMaker containers.

B. Build a custom Docker container that includes the required libraries. Host the container in Amazon Elastic Container Registry (Amazon ECR). Use the ECR image in the SageMaker jobs and pipelines.

C. Create a SageMaker notebook instance to host the jobs. Create an AWS Lambda function to install the libraries on the notebook instance when the notebook instance starts. Configure the SageMaker jobs and pipelines to run on the notebook instance.

D. Run code for the libraries externally on Amazon EC2 instances. Store the results in Amazon S3.Import the results into the SageMaker jobs and pipelines.

Answer: B

Explanation:

Building a custom Docker container with the required libraries and hosting it in Amazon ECR allows SageMaker jobs, training, and pipelines to consistently use the same environment. This approach minimizes manual setup, ensures portability, and provides the least ongoing implementation effort compared to repeatedly installing or managing libraries separately.

QUESTION NO: 20

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

QUESTION NO: 21

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 needs to run an on-demand workflow to monitor bias drift for models that are deployed to real-time endpoints from the application.

Which action will meet this requirement?

A. Configure the application to invoke an AWS Lambda function that runs a SageMaker Clarify job.

B. Invoke an AWS Lambda function to pull the sagemaker-model-monitor-analyzer built-in SageMaker image.

C. Use AWS Glue Data Quality to monitor bias.

D. Use SageMaker notebooks to compare the bias.

Answer: A

QUESTION NO: 22

A company is training a large language model (LLM) by using on-premises infrastructure. A live conversational engine uses the LLM to help customers find real-time insights in credit card data.

An ML engineer must implement a solution to train and deploy the LLM on Amazon SageMaker.

Which solution will meet these requirements?

A. Use SageMaker Training Compiler to train the LLM. Deploy the LLM by using SageMaker real- time inference.

B. Use SageMaker with deep learning containers for large model inference to train the LLM. Deploy the LLM by using SageMaker real-time inference.

C. Use SageMaker Notebook Jobs to train the LLM. Deploy the LLM by using SageMaker Asynchronous Inference.

D. Use SageMaker Studio to train the LLM. Deploy the LLM by using SageMaker batch transform.

Answer: A

Explanation:

SageMaker Training Compiler accelerates training of large models like LLMs by optimizing GPU utilization, making it suitable for efficient large-scale training. For deployment of a live conversational engine that requires real-time responses, the correct choice is a SageMaker real- time inference endpoint. This combination meets both training and deployment requirements effectively.

QUESTION NO: 23

A data scientist is evaluating different binary classification models. A false positive result is 5 times more expensive (from a business perspective) than a false negative result.

The models should be evaluated based on the following criteria:

1) Must have a recall rate of at least 80%

2) Must have a false positive rate of 10% or less

3) Must minimize business costs

After creating each binary classification model, the data scientist generates the corresponding confusion matrix.

Which confusion matrix represents the model that satisfies the requirements?

A. TN = 91, FP = 9

FN = 22, TP = 78

B. TN = 99, FP = 1

FN = 21, TP = 79

C. TN = 96, FP = 4

FN = 10, TP = 90

D. TN = 98, FP = 2

FN = 18, TP = 82

Answer: D

Explanation:

The following calculations are required:

TP = True Positive

FP = False Positive

FN = False Negative

TN = True Negative

FN = False Negative

Recall = TP / (TP + FN)

False Positive Rate (FPR) = FP / (FP + TN)

Cost = 5 * FP + FN

	A	B	C	D
Recall	78 / (78 + 22) = 0.78	79 / (79 + 21) = 0.79	90 / (90 + 10) = 0.9	82 / (82 + 18) = 0.82
False Positive Rate	9 / (9 + 91) = 0.09	1 / (1 + 99) = 0.01	4 / (4 + 96) = 0.04	2 / (2 + 98) = 0.02
Costs	5 * 9 + 22 = 67	5 * 1 + 21 = 26	5 * 4 + 10 = 30	5 * 2 + 18 = 28

QUESTION NO: 24

A company has historical data that shows whether customers needed long-term support from company staff. The company needs to develop an ML model to predict whether new customers will require long-term support.

Which modeling approach should the company use to meet this requirement?

A. Anomaly detection

B. Linear regression

C. Logistic regression

D. Semantic segmentation

Answer: C

QUESTION NO: 25

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

QUESTION NO: 26

A company is planning to use Amazon Redshift ML in its primary AWS account. The source data is in an Amazon S3 bucket in a secondary account.

An ML engineer needs to set up an ML pipeline in the primary account to access the S3

bucket in the secondary account. The solution must not require public IPv4 addresses.

Which solution will meet these requirements?

A. Provision a Redshift cluster and Amazon SageMaker Studio in a VPC with no public access enabled in the primary account. Create a VPC peering connection between the accounts. Update the VPC route tables to remove the route to 0.0.0.0/0.

B. Provision a Redshift cluster and Amazon SageMaker Studio in a VPC with no public access enabled in the primary account. Create an AWS Direct Connect connection and a transit gateway.

Associate the VPCs from both accounts with the transit gateway. Update the VPC route tables to remove the route to 0.0.0.0/0.

C. Provision a Redshift cluster and Amazon SageMaker Studio in a VPC in the primary account.

Create an AWS Site-to-Site VPN connection with two encrypted IPsec tunnels between the accounts. Set up interface VPC endpoints for Amazon S3.

D. Provision a Redshift cluster and Amazon SageMaker Studio in a VPC in the primary account.Create an S3 gateway endpoint. Update the S3 bucket policy to allow IAM principals from the primary account. Set up interface VPC endpoints for SageMaker and Amazon Redshift.

Answer: D

QUESTION NO: 27

An ML engineer is training a simple neural network model. The ML engineer tracks the performance of the model over time on a validation dataset. The model's performance improves substantially at first and then degrades after a specific number of epochs.

Which solutions will mitigate this problem? (Choose two.)

A. Enable early stopping on the model.

B. Increase dropout in the layers.

C. Increase the number of layers.

D. Increase the number of neurons.

E. Investigate and reduce the sources of model bias.

Answer: AB

QUESTION NO: 28

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

QUESTION NO: 29

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. Transfer the copy to Account B.

B. Create a resource-based IAM policy. Use the Amazon Comprehend ImportModel API operation to copy the model to Account B.

C. Use AWS DataSync to replicate the model from Account A to Account B.

D. Create an AWS Site-to-Site VPN connection between Account A and Account B to transfer the model.

Answer: B

QUESTION NO: 30

A machine learning engineer is preparing a data frame for a supervised learning task with the Amazon SageMaker Linear Learner algorithm. The ML engineer notices the target label classes are highly imbalanced and multiple feature columns contain missing values. The proportion of missing values across the entire data frame is less than 5%.

What should the ML engineer do to minimize bias due to missing values?

A. Replace each missing value by the mean or median across non-missing values in same row.

B. Delete observations that contain missing values because these represent less than 5% of the data.

C. Replace each missing value by the mean or median across non-missing values in the same column.

D. For each feature, approximate the missing values using supervised learning based on other features.

Answer: D

Explanation:

Use supervised learning to predict missing values based on the values of other features. Different supervised learning approaches might have different performances, but any properly implemented supervised learning approach should provide the same or better approximation than mean or median approximation, as proposed in responses A and C. Supervised learning applied to the imputation of missing values is an active field of research.

QUESTION NO: 31

A company needs an AWS solution that will automatically create versions of ML models as the models are created.

Which solution will meet this requirement?

A. Amazon Elastic Container Registry (Amazon ECR)

B. Model packages from Amazon SageMaker Marketplace

C. Amazon SageMaker ML Lineage Tracking

D. Amazon SageMaker Model Registry

Answer: D

QUESTION NO: 32

A company uses 10 Reserved Instances of accelerated instance types to serve the current version of an ML model. An ML engineer needs to deploy a new version of the model to an Amazon SageMaker real-time inference endpoint.

The solution must use the original 10 instances to serve both versions of the model. The solution also must include one additional Reserved Instance that is available to use in the deployment process. The transition between versions must occur with no downtime or service interruptions.

Which solution will meet these requirements?

A. Configure a blue/green deployment with all-at-once traffic shifting.

B. Configure a blue/green deployment with canary traffic shifting and a size of 10%.

C. Configure a shadow test with a traffic sampling percentage of 10%.

D. Configure a rolling deployment with a rolling batch size of 1.

Answer: D

QUESTION NO: 33

Hotspot Question

An ML engineer is building a generative AI application on Amazon Bedrock by using large language models (LLMs).

Select the correct generative AI term from the following list for each description. Each term should be selected one time or not at all. (Select three.)

- Embedding
- Retrieval Augmented Generation (RAG)
- Temperature
- Token

Text representation of basic units of data processed by LLMs | Select... ▼

Select...
Embedding
Retrieval Augmented Generation (RAG)
Temperature
Token

High-dimensional vectors that contain the semantic meaning of text | Select... ▼

Select...
Embedding
Retrieval Augmented Generation (RAG)
Temperature
Token

Enrichment of information from additional data sources to improve a generated response | Select... ▼

Select...
Embedding
Retrieval Augmented Generation (RAG)
Temperature
Token

Answer:

Text representation of basic units of data processed by LLMs

```
Select...                                        ▼
Select...
Embedding
Retrieval Augmented Generation (RAG)
Temperature
Token
```

High-dimensional vectors that contain the semantic meaning of text

```
Select...                                        ▼
Select...
Embedding
Retrieval Augmented Generation (RAG)
Temperature
Token
```

Enrichment of information from additional data sources to improve a generated response

```
Select...                                        ▼
Select...
Embedding
Retrieval Augmented Generation (RAG)
Temperature
Token
```

QUESTION NO: 34

A company is using Amazon SageMaker to create ML models. The company's data scientists need fine-grained control of the ML workflows that they orchestrate. The data scientists also need the ability to visualize SageMaker jobs and workflows as a directed acyclic graph (DAG). The data scientists must keep a running history of model discovery experiments and must establish model governance for auditing and compliance verifications.

Which solution will meet these requirements?

A. Use AWS CodePipeline and its integration with SageMaker Studio to manage the entire ML workflows. Use SageMaker ML Lineage Tracking for the running history of experiments and for auditing and compliance verifications.

B. Use AWS CodePipeline and its integration with SageMaker Experiments to manage the entire ML workflows. Use SageMaker Experiments for the running history of experiments and for auditing and compliance verifications.

C. Use SageMaker Pipelines and its integration with SageMaker Studio to manage the entire ML workflows. Use SageMaker ML Lineage Tracking for the running history of experiments and for auditing and compliance verifications.

D. Use SageMaker Pipelines and its integration with SageMaker Experiments to manage the entire ML workflows. Use SageMaker Experiments for the running history of experiments and for auditing and compliance verifications.

Answer: C

QUESTION NO: 35

A company runs an ML model on Amazon SageMaker. The company uses an automatic process that makes API calls to create training jobs for the model. The company has new compliance rules that prohibit the collection of aggregated metadata from training jobs. Which solution will prevent SageMaker from collecting metadata from the training jobs?

A. Opt out of metadata tracking for any training job that is submitted.

B. Ensure that training jobs are running in a private subnet in a custom VPC.

C. Encrypt the training data with an AWS Key Management Service (AWS KMS) customer managed key.

D. Reconfigure the training jobs to use only AWS Nitro instances.

Answer: A

Explanation:

Amazon SageMaker automatically collects training job metadata, but you can opt out of metadata tracking when submitting a training job. This disables collection of aggregated metadata, ensuring compliance with rules that prohibit metadata collection.

QUESTION NO: 36

A company has deployed an ML model that detects fraudulent credit card transactions in real time in a banking application. The model uses Amazon SageMaker Asynchronous Inference. Consumers are reporting delays in receiving the inference results.

An ML engineer needs to implement a solution to improve the inference performance. The solution also must provide a notification when a deviation in model quality occurs.

Which solution will meet these requirements?

A. Use SageMaker real-time inference for inference. Use SageMaker Model Monitor for notifications about model quality.

B. Use SageMaker batch transform for inference. Use SageMaker Model Monitor for notifications about model quality.

C. Use SageMaker Serverless Inference for inference. Use SageMaker Inference Recommender for notifications about model quality.

D. Keep using SageMaker Asynchronous Inference for inference. Use SageMaker Inference Recommender for notifications about model quality.

Answer: A

QUESTION NO: 37

A company has developed a new ML model. The company requires online model validation on

10% of the traffic before the company fully releases the model in production. The company uses an Amazon SageMaker endpoint behind an Application Load Balancer (ALB) to serve the model.

Which solution will set up the required online validation with the LEAST operational overhead?

A. Use production variants to add the new model to the existing SageMaker endpoint. Set the variant weight to 0.1 for the new model. Monitor the number of invocations by using Amazon CloudWatch.

B. Use production variants to add the new model to the existing SageMaker endpoint. Set the variant weight to 1 for the new model. Monitor the number of invocations by using Amazon CloudWatch.

C. Create a new SageMaker endpoint. Use production variants to add the new model to the new endpoint. Monitor the number of invocations by using Amazon CloudWatch.

D. Configure the ALB to route 10% of the traffic to the new model at the existing SageMaker endpoint. Monitor the number of invocations by using AWS CloudTrail.

Answer: A

QUESTION NO: 38

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.

Before the ML engineer trains the model, the ML engineer must resolve the issue of the imbalanced data.

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

A. Use Amazon Athena to identify patterns that contribute to the imbalance. Adjust the dataset accordingly.

B. Use Amazon SageMaker Studio Classic built-in algorithms to process the imbalanced dataset.

C. Use AWS Glue DataBrew built-in features to oversample the minority class.

D. Use the Amazon SageMaker Data Wrangler balance data operation to oversample the minority class.

Answer: D

QUESTION NO: 39

A company needs to develop an ML model. The model must identify an item in an image and must provide the location of the item.

Which Amazon SageMaker algorithm will meet these requirements?

A. Image classification

B. XGBoost

C. Object detection

D. K-nearest neighbors (k-NN)

Answer: C

QUESTION NO: 40

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

QUESTION NO: 41

A machine learning team has several large CSV datasets in Amazon S3. Historically, models

built with the Amazon SageMaker Linear Learner algorithm have taken hours to train on similar-sized datasets. The team's leaders need to accelerate the training process.

What can a machine learning specialist do to address this concern?

A. Use Amazon SageMaker Pipe mode.

B. Use Amazon Machine Learning to train the models.

C. Use Amazon Kinesis to stream the data to Amazon SageMaker.

D. Use AWS Glue to transform the CSV dataset to the JSON format.

Answer: A

Explanation:

Amazon SageMaker Pipe mode streams the data directly to the container, which improves the performance of training jobs. In Pipe mode, your training job streams data directly from Amazon S3. Streaming can provide faster start times for training jobs and better throughput. With Pipe mode, you also reduce the size of the Amazon EBS volumes for your training instances.

QUESTION NO: 42

A company is using an Amazon S3 bucket to collect data that will be used for ML workflows. The company needs to use AWS Glue DataBrew to clean and normalize the data. Which solution will meet these requirements?

A. Create a DataBrew dataset by using the S3 path. Clean and normalize the data by using a DataBrew profile job.

B. Create a DataBrew dataset by using the S3 path. Clean and normalize the data by using a DataBrew recipe job.

C. Create a DataBrew dataset by using a Java Database Connectivity (JDBC) driver to connect to the S3 bucket. Clean and normalize the data by using a DataBrew profile job.

D. Create a DataBrew dataset by using a Java Database Connectivity (JDBC) driver to connect to the S3 bucket. Clean and normalize the data by using a DataBrew recipe job.

Answer: B

Explanation:

The correct solution is to create a DataBrew dataset using the S3 path and then clean and normalize the data with a DataBrew recipe job. Recipes define and apply transformations to the data, while profile jobs are used only for data analysis and profiling, not cleaning.

QUESTION NO: 43

Hotspot Question

A company stores historical data in .csv files in Amazon S3. Only some of the rows and columns in the .csv files are populated. The columns are not labeled. An ML engineer needs to prepare and store the data so that the company can use the data to train ML models. Select and order the correct steps from the following list to perform this task. Each step should be selected one time or not at all. (Select and order three.)

- Create an Amazon SageMaker batch transform job for data cleaning and feature engineering.

- Store the resulting data back in Amazon S3.

- Use Amazon Athena to infer the schemas and available columns.

- Use AWS Glue crawlers to infer the schemas and available columns.

- Use AWS Glue DataBrew for data cleaning and feature engineering.

Step 1: Select...

Select...
Create an Amazon SageMaker batch transform job for data cleaning and feature engineering.
Store the resulting data back in Amazon S3.
Use Amazon Athena to infer the schemas and available columns.
Use AWS Glue crawlers to infer the schemas and available columns.
Use AWS Glue DataBrew for data cleaning and feature engineering.

Step 2: Select...

Select...
Create an Amazon SageMaker batch transform job for data cleaning and feature engineering.
Store the resulting data back in Amazon S3.
Use Amazon Athena to infer the schemas and available columns.
Use AWS Glue crawlers to infer the schemas and available columns.
Use AWS Glue DataBrew for data cleaning and feature engineering.

Step 3: Select...

Select...
Create an Amazon SageMaker batch transform job for data cleaning and feature engineering.
Store the resulting data back in Amazon S3.
Use Amazon Athena to infer the schemas and available columns.
Use AWS Glue crawlers to infer the schemas and available columns.
Use AWS Glue DataBrew for data cleaning and feature engineering.

Answer:

Step 1: Select...

Select...
Create an Amazon SageMaker batch transform job for data cleaning and feature engineering.
Store the resulting data back in Amazon S3.
Use Amazon Athena to infer the schemas and available columns.
Use AWS Glue crawlers to infer the schemas and available columns.
Use AWS Glue DataBrew for data cleaning and feature engineering.

Step 2: Select...

Select...
Create an Amazon SageMaker batch transform job for data cleaning and feature engineering.
Store the resulting data back in Amazon S3.
Use Amazon Athena to infer the schemas and available columns.
Use AWS Glue crawlers to infer the schemas and available columns.
Use AWS Glue DataBrew for data cleaning and feature engineering.

Step 3: Select...

Select...
Create an Amazon SageMaker batch transform job for data cleaning and feature engineering.
Store the resulting data back in Amazon S3.
Use Amazon Athena to infer the schemas and available columns.
Use AWS Glue crawlers to infer the schemas and available columns.
Use AWS Glue DataBrew for data cleaning and feature engineering.

QUESTION NO: 44

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.

Which solution will meet these requirements?

A. Use SageMaker Debugger to track the inferences and to report metrics. Create a custom rule to provide a notification when the threshold is breached.

B. Use SageMaker Debugger to track the inferences and to report metrics. Use the tensor_variance built-in rule to provide a notification when the threshold is breached.

C. Log all the endpoint invocation API events by using AWS CloudTrail. Use an Amazon CloudWatch dashboard for monitoring. Set up a CloudWatch alarm to provide notification when the threshold is breached.

D. Add the Invocations metric to an Amazon CloudWatch dashboard for monitoring. Set up a CloudWatch alarm to provide notification when the threshold is breached.

Answer: C

QUESTION NO: 45

An ML engineer needs to deploy a trained model that is based on a genetic algorithm. The algorithm solves a complex problem and can take several minutes to generate predictions. When the model is deployed, the model needs to access large amounts of data to process requests. The requests can involve as much as 100 MB of data.

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

A. Deploy the model to Amazon EC2 instances in an Auto Scaling group behind an Application Load Balancer.

B. Deploy the model to an Amazon SageMaker real-time endpoint.

C. Deploy the model to an Amazon SageMaker Asynchronous Inference endpoint.

D. Package the model as a container. Deploy the model to Amazon Elastic Container Service (Amazon ECS) on Amazon EC2 instances.

Answer: C

Explanation:

SageMaker Asynchronous Inference is designed for models with long processing times and large payloads. It can handle input data up to 1 GB and avoids holding open connections during long inference runs, reducing operational overhead compared to managing EC2 or ECS infrastructure.

This makes it the best fit for the genetic algorithm model that takes minutes and processes large requests.

QUESTION NO: 46

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

QUESTION NO: 47

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. Configure auto scaling. Configure the endpoint to present the existing model.

B. Deploy the model on an Amazon Elastic Container Service (Amazon ECS) cluster. Use ECS scheduled scaling that is based on the CPU of the ECS cluster.

C. Install SageMaker Operator on an Amazon Elastic Kubernetes Service (Amazon EKS) cluster.

Deploy the model in Amazon EKS. Set horizontal pod auto scaling to scale replicas based on the memory metric.

D. Use Spot Instances with a Spot Fleet behind an Application Load Balancer (ALB) for inferences.Use the ALBRequestCountPerTarget metric as the metric for auto scaling.

Answer: A

QUESTION NO: 48

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: D

QUESTION NO: 49

A company is exploring generative AI and wants to add a new product feature. An ML engineer is making API calls from existing Amazon EC2 instances to Amazon Bedrock. The

EC2 instances are in a private subnet and must remain private during the implementation. The EC2 instances have an assigned security group that allows access to all IP addresses in the private subnet.

What should the ML engineer do to establish a connection between the EC2 instances and Amazon Bedrock?

A. Modify the security group to allow inbound and outbound traffic to and from Amazon Bedrock.

B. Use AWS PrivateLink to access Amazon Bedrock through an interface VPC endpoint.

C. Configure Amazon Bedrock to use the private subnet where the EC2 instances are deployed.

D. Link the existing VPC to Amazon Bedrock by using an AWS Direct Connect connection.

Answer: B

Explanation:

Since the EC2 instances are in a private subnet and must not have public internet access, the correct solution is to use AWS PrivateLink with an interface VPC endpoint for Amazon Bedrock.

This allows private connectivity from the VPC to the Bedrock service without exposing traffic to the public internet.

QUESTION NO: 50

An ML engineer wants an Amazon SageMaker notebook to automatically stop running after 1 hour of idle time. How can the ML engineer accomplish this goal?

A. Create a lifecycle configuration in SageMaker. Copy the auto-stop-idle script from GitHub to the Start Notebook section.

B. Create a lifecycle configuration in SageMaker. Copy the auto-stop-idle script from GitHub to the Create Notebook section.

C. Track the notebook's CPU metric by using Amazon CloudWatch Logs. Invoke an AWS Lambda function from CloudWatch Logs to shut down the notebook instance if CPU utilization becomes zero.

D. Track the notebook's memory metric by using Amazon CloudWatch Logs. Invoke an AWS Lambda function from CloudWatch Logs to shut down the notebook instance if memory utilization becomes zero.

Answer: A

Explanation:

The correct approach is to use a SageMaker lifecycle configuration and place the auto-stop-idle script in the Start Notebook section. This ensures the notebook runs the monitoring script on startup, which checks for idle time and automatically stops the notebook after the defined threshold (1 hour in this case).

QUESTION NO: 51

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 dataset. Update Model Monitor to use the new baseline for evaluations.

D. Include additional data in the existing training set for the model. Retrain and redeploy the model.

Answer: C

QUESTION NO: 52

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 subnet. Store the data in a different AWS Region from where the instances are deployed.

B. Place the instances in the same VPC subnet but in different Availability Zones. Store the data in a different AWS Region from where the instances are deployed.

C. Place the instances in the same VPC subnet. Store the data in the same AWS Region and Availability Zone where the instances are deployed.

D. Place the instances in the same VPC subnet. Store the data in the same AWS Region but in a different Availability Zone from where the instances are deployed.

Answer: C

QUESTION NO: 53

A company is using Amazon SageMaker to develop ML models. The company stores sensitive training data in an Amazon S3 bucket. The model training must have network isolation from the internet.

Which solution will meet this requirement?

A. Run the SageMaker training jobs in private subnets. Create a NAT gateway. Route traffic for training through the NAT gateway.

B. Run the SageMaker training jobs in private subnets. Create an S3 gateway VPC endpoint. Route traffic for training through the S3 gateway VPC endpoint.

C. Run the SageMaker training jobs in public subnets that have an attached security group. In the security group, use inbound rules to limit traffic from the internet. Encrypt SageMaker instance storage by using server-side encryption with AWS KMS keys (SSE-KMS).

D. Encrypt traffic to Amazon S3 by using a bucket policy that includes a value of True for the aws:SecureTransport condition key. Use default at-rest encryption for Amazon S3. Encrypt SageMaker instance storage by using server-side encryption with AWS KMS keys (SSE-KMS).

Answer: B

QUESTION NO: 54

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. Use Amazon QuickSight to visualize the result.

C. Use Amazon SageMaker Data Wrangler to automatically detect the anomalies and to visualize the result.

D. Use AWS Batch to automatically detect the anomalies. Use Amazon QuickSight to visualize the result.

Answer: C

QUESTION NO: 55

A term frequency-inverse document frequency (tf-idf) matrix using both unigrams and bigrams is built from a text corpus consisting of the following two sentences:

1. Please call the number below.

2. Please do not call us.

What are the dimensions of the tf-idf matrix?

A. (2, 16)

B. (2, 8)

C. (2, 10)

D. (8, 10)

Answer: A

Explanation:

There are 2 sentences, 8 unique unigrams, and 8 unique bigrams, so the result would be (2,16).

The phrases are "Please call the number below" and "Please do not call us." Each word individually (unigram) is "Please," "call," "the," "number," "below," "do," "not," and "us." The unique bigrams are "Please call," "call the," "the number," "number below," "Please do," "do not,"

"not call," and "call us."

QUESTION NO: 56

An ML engineer has deployed an Amazon SageMaker model to a serverless endpoint in production. The model is invoked by the InvokeEndpoint API operation.

The model's latency in production is higher than the baseline latency in the test environment.

The ML engineer thinks that the increase in latency is because of model startup time.

What should the ML engineer do to confirm or deny this hypothesis?

A. Schedule a SageMaker Model Monitor job. Observe metrics about model quality.

B. Schedule a SageMaker Model Monitor job with Amazon CloudWatch metrics enabled.

C. Enable Amazon CloudWatch metrics. Observe the ModelSetupTime metric in the SageMaker namespace.

D. Enable Amazon CloudWatch metrics. Observe the ModelLoadingWaitTime metric in the SageMaker namespace.

Answer: D

QUESTION NO: 57

A company is using Amazon SageMaker and millions of files to train an ML model. Each file is several megabytes in size. The files are stored in an Amazon S3 bucket. The company needs to improve training performance.

Which solution will meet these requirements in the LEAST amount of time?

A. Transfer the data to a new S3 bucket that provides S3 Express One Zone storage. Adjust the training job to use the new S3 bucket.

B. Create an Amazon FSx for Lustre file system. Link the file system to the existing S3 bucket.

Adjust the training job to read from the file system.

C. Create an Amazon Elastic File System (Amazon EFS) file system. Transfer the existing data to the file system. Adjust the training job to read from the file system.

D. Create an Amazon ElastiCache (Redis OSS) cluster. Link the Redis OSS cluster to the existing S3 bucket. Stream the data from the Redis OSS cluster directly to the training job.

Answer: B

QUESTION NO: 58

An ML engineer needs to train a supervised deep learning model. The available dataset is a large number of unlabeled images that only employees should access. The ML engineer needs to implement a solution that labels the dataset with the highest possible accuracy.

Which combination of steps should the ML engineer take to meet these requirements? (Choose two.)

A. Use Amazon Rekognition to automatically label the dataset.

B. Train the deep learning model directly on the raw data. Let the model infer the labels by itself.

C. Use Amazon SageMaker Ground Truth to create an annotation job that specifies the labeling task and requirements.

D. Set up workforce teams to access a private workforce to run and review the annotation job created by Amazon SageMaker Ground Truth.

E. Use Amazon Mechanical Turk to complete the annotation job created by Amazon SageMaker Ground Truth.

Answer: CD

Explanation:

To achieve the highest labeling accuracy with controlled employee-only access, the ML engineer should use Amazon SageMaker Ground Truth to define the annotation job and then assign it to a private workforce of employees for labeling and review. This ensures high-

quality, secure labeling restricted to authorized personnel.

QUESTION NO: 59

A company needs to use Amazon SageMaker to train a model on more than 300 GB of data. The training data is composed of files that are 200 MB in size. The data is stored in Amazon S3 Standard storage and feeds a dashboard tool. Which SageMaker training ingestion mechanism is the MOST cost-effective solution for this scenario?

A. Amazon Elastic File System (Amazon EFS) file system

B. Amazon FSx for Lustre file system

C. Amazon S3 in fast file mode while using S3 Express One Zone

D. Amazon S3 in fast file mode without using S3 Express One Zone

Answer: D

Explanation:

For large-scale training data already stored in Amazon S3, the most cost-effective solution is to use SageMaker's S3 fast file mode without S3 Express One Zone. Fast file mode enables streaming directly from S3 without duplicating the dataset onto local storage, reducing startup time and storage cost. Using S3 Express One Zone would increase cost, so standard fast file mode is the most economical choice.

QUESTION NO: 60

A medical company is using AWS to build a tool to recommend treatments for patients. The company has obtained health records and self-reported textual information in English from patients. The company needs to use this information to gain insight about the patients.
Which solution will meet this requirement with the LEAST development effort?

A. Use Amazon SageMaker to build a recurrent neural network (RNN) to summarize the data.

B. Use Amazon Comprehend Medical to summarize the data.

C. Use Amazon Kendra to create a quick-search tool to query the data.

D. Use the Amazon SageMaker Sequence-to-Sequence (seq2seq) algorithm to create a text summary from the data.

Answer: B

QUESTION NO: 61

A company has AWS Glue data processing jobs that are orchestrated by an AWS Glue workflow.
The AWS Glue jobs can run on a schedule or can be launched manually.
The company is developing pipelines in Amazon SageMaker Pipelines for ML model development. The pipelines will use the output of the AWS Glue jobs during the data processing phase of model development. An ML engineer needs to implement a solution that integrates the AWS Glue jobs with the pipelines.
Which solution will meet these requirements with the LEAST operational overhead?

A. Use AWS Step Functions for orchestration of the pipelines and the AWS Glue jobs.

B. Use processing steps in SageMaker Pipelines. Configure inputs that point to the Amazon Resource Names (ARNs) of the AWS Glue jobs.

C. Use Callback steps in SageMaker Pipelines to start the AWS Glue workflow and to stop the pipelines until the AWS Glue jobs finish running.

D. Use Amazon EventBridge to invoke the pipelines and the AWS Glue jobs in the desired order.

Answer: C

QUESTION NO: 62

A company runs training jobs on Amazon SageMaker by using a compute optimized instance.

Demand for training runs will remain constant for the next 55 weeks. The instance needs to run for 35 hours each week. The company needs to reduce its model training costs.

Which solution will meet these requirements?

A. Use a serverless endpoint with a provisioned concurrency of 35 hours for each week. Run the training on the endpoint.

B. Use SageMaker Edge Manager for the training. Specify the instance requirement in the edge device configuration. Run the training.

C. Use the heterogeneous cluster feature of SageMaker Training. Configure the instance_type, instance_count, and instance_groups arguments to run training jobs.

D. Opt in to a SageMaker Savings Plan with a 1-year term and an All Upfront payment. Run a SageMaker Training job on the instance.

Answer: D

QUESTION NO: 63

A medical company needs to store clinical data. The data includes personally identifiable information (PII) and protected health information (PHI).

An ML engineer needs to implement a solution to ensure that the PII and PHI are not used to train ML models.

Which solution will meet these requirements?

A. Store the clinical data in Amazon S3 buckets. Use AWS Glue DataBrew to mask the PII and PHI before the data is used for model training.

B. Upload the clinical data to an Amazon Redshift database. Use built-in SQL stored procedures to automatically classify and mask the PII and PHI before the data is used for model training.

C. Use Amazon Comprehend to detect and mask the PII before the data is used for model training.

Use Amazon Comprehend Medical to detect and mask the PHI before the data is used for model training.

D. Create an AWS Lambda function to encrypt the PII and PHI. Program the Lambda function to save the encrypted data to an Amazon S3 bucket for model training.

Answer: C

QUESTION NO: 64

An ML engineer needs to ensure that a dataset complies with regulations for personally identifiable information (PII). The ML engineer will use the data to train an ML model on

Amazon SageMaker instances. SageMaker must not use any of the PII.

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

A. Use the Amazon Comprehend DetectPiiEntities API call to redact the PII from the data. Store the data in an Amazon S3 bucket. Access the S3 bucket from the SageMaker instances for model training.

B. Use the Amazon Comprehend DetectPiiEntities API call to redact the PII from the data. Store the data in an Amazon Elastic File System (Amazon EFS) file system. Mount the EFS file system to the SageMaker instances for model training.

C. Use AWS Glue DataBrew to cleanse the dataset of PII. Store the data in an Amazon Elastic File System (Amazon EFS) file system. Mount the EFS file system to the SageMaker instances for model training.

D. Use Amazon Macie for automatic discovery of PII in the data. Remove the PII. Store the data in an Amazon S3 bucket. Mount the S3 bucket to the SageMaker instances for model training.

Answer: A

QUESTION NO: 65

A company is building an ML model by using Amazon SageMaker, AWS owned libraries, and open source libraries. The company must ensure that SageMaker does not collect metadata about usage and errors during training. Which solution will meet these requirements?

A. Associate the SageMaker domain with a custom IAM role. Attach the role to a policy that denies Amazon CloudWatch service usage logs.

B. Add an IAM role to the SageMaker domain to deny Amazon CloudWatch the permission to report metadata.

C. Turn off the setting in the SageMaker domain to share metadata for console jobs. Opt out of metadata collection for each training job that is submitted through the AWS CLI or AWS SDKs.

D. Set a parameter to opt out of metadata collection for each training job that is submitted through the AWS CLI, Boto3, or the SageMaker Python SDK.

Answer: C

Explanation:

To prevent metadata collection in Amazon SageMaker, you must disable the metadata sharing setting in the SageMaker domain for console jobs and explicitly opt out of metadata collection for each training job submitted through the AWS CLI or SDKs. This ensures that neither usage nor error metadata is collected during training.

QUESTION NO: 66

Hotspot Question

A company wants to host an ML model on Amazon SageMaker. An ML engineer is configuring a continuous integration and continuous delivery (CI/CD) pipeline in AWS CodePipeline to deploy the model. The pipeline must run automatically when new training data for the model is uploaded to an Amazon S3 bucket.

Select and order the pipeline's correct steps from the following list. Each step should be selected one time or not at all. (Select and order three.)

- An S3 event notification invokes the pipeline when new data is uploaded.
- S3 Lifecycle rule invokes the pipeline when new data is uploaded.
- SageMaker retrains the model by using the data in the S3 bucket.
- The pipeline deploys the model to a SageMaker endpoint.
- The pipeline deploys the model to SageMaker Model Registry.

Step 1:
Select...
Select...
An S3 event notification invokes the pipeline when new data is uploaded.
An S3 Lifecycle rule invokes the pipeline when new data is uploaded.
SageMaker retrains the model by using the data in the S3 bucket.
The pipeline deploys the model to a SageMaker endpoint.
The pipeline deploys the model to SageMaker Model Registry.

Step 2:
Select...
Select...
An S3 event notification invokes the pipeline when new data is uploaded.
An S3 Lifecycle rule invokes the pipeline when new data is uploaded.
SageMaker retrains the model by using the data in the S3 bucket.
The pipeline deploys the model to a SageMaker endpoint.
The pipeline deploys the model to SageMaker Model Registry.

Step 3:
Select...
Select...
An S3 event notification invokes the pipeline when new data is uploaded.
An S3 Lifecycle rule invokes the pipeline when new data is uploaded.
SageMaker retrains the model by using the data in the S3 bucket.
The pipeline deploys the model to a SageMaker endpoint.
The pipeline deploys the model to SageMaker Model Registry.

Answer:

Step 1:

Select...	▼

Select...
An S3 event notification invokes the pipeline when new data is uploaded.
An S3 Lifecycle rule invokes the pipeline when new data is uploaded.
SageMaker retrains the model by using the data in the S3 bucket.
The pipeline deploys the model to a SageMaker endpoint.
The pipeline deploys the model to SageMaker Model Registry.

Step 2:

Select...	▼

Select...
An S3 event notification invokes the pipeline when new data is uploaded.
An S3 Lifecycle rule invokes the pipeline when new data is uploaded.
SageMaker retrains the model by using the data in the S3 bucket.
The pipeline deploys the model to a SageMaker endpoint.
The pipeline deploys the model to SageMaker Model Registry.

Step 3:

Select...	▼

Select...
An S3 event notification invokes the pipeline when new data is uploaded.
An S3 Lifecycle rule invokes the pipeline when new data is uploaded.
SageMaker retrains the model by using the data in the S3 bucket.
The pipeline deploys the model to a SageMaker endpoint.
The pipeline deploys the model to SageMaker Model Registry.

QUESTION NO: 67

A company uses Amazon SageMaker for its ML workloads. The company's ML engineer receives a 50 MB Apache Parquet data file to build a fraud detection model. The file includes several correlated columns that are not required.

What should the ML engineer do to drop the unnecessary columns in the file with the LEAST effort?

A. Download the file to a local workstation. Perform one-hot encoding by using a custom Python script.

B. Create an Apache Spark job that uses a custom processing script on Amazon EMR.

C. Create a SageMaker processing job by calling the SageMaker Python SDK.

D. Create a data flow in SageMaker Data Wrangler. Configure a transform step.

Answer: D

QUESTION NO: 68

An IoT company uses Amazon SageMaker to train and test an XGBoost model for object detection. ML engineers need to monitor performance metrics when they train the model with variants in hyperparameters. The ML engineers also need to send Short Message Service (SMS) text messages after training is complete.

Which solution will meet these requirements?

A. Use Amazon CloudWatch to monitor performance metrics. Use Amazon Simple Queue

Service (Amazon SQS) for message delivery.

B. Use Amazon CloudWatch to monitor performance metrics. Use Amazon Simple Notification Service (Amazon SNS) for message delivery.

C. Use AWS CloudTrail to monitor performance metrics. Use Amazon Simple Queue Service (Amazon SQS) for message delivery.

D. Use AWS CloudTrail to monitor performance metrics. Use Amazon Simple Notification Service (Amazon SNS) for message delivery.

Answer: B

QUESTION NO: 69

An ML engineer needs to deploy four ML models in an Amazon SageMaker inference pipeline.

The models were built with different frameworks. The ML engineer also needs to give clients the ability to use the invoke_endpoint call to perform inference for each model. Which solution will meet these requirements MOST cost-effectively?

A. Create a SageMaker multi-model endpoint.

B. Create a SageMaker multi-container endpoint.

C. Create multiple SageMaker single-model endpoints.

D. Run a SparkML job to generate multiple endpoints.

Answer: B

Explanation:

A SageMaker multi-container endpoint allows deployment of multiple models built with different frameworks in a single endpoint. Each container can host a model with its required framework, and clients can use the same invoke_endpoint call while specifying the target container. This meets the requirement for framework diversity and is more cost-effective than running separate single-model endpoints.

QUESTION NO: 70

A company has collected customer comments on its products, rating them as safe or unsafe, using decision trees. The training dataset has the following features: id, date, full review, full review summary, and a binary safe/unsafe tag. During training, any data sample with missing features was dropped. In a few instances, the test set was found to be missing the full review text field.

For this use case, which is the most effective course of action to address test data samples with missing features?

A. Drop the test samples with missing full review text fields, and then run through the test set.

B. Copy the summary text fields and use them to fill in the missing full review text fields, and then run through the test set.

C. Use an algorithm that handles missing data better than decision trees.

D. Generate synthetic data to fill in the fields that are missing data, and then run through the test set.

Answer: B

Explanation:

In this case, a full review summary usually contains the most descriptive phrases of the entire

review and is a valid stand-in for the missing full review text field.

QUESTION NO: 71

A company receives daily .csv files about customer interactions with its ML model. The company stores the files in Amazon S3 and uses the files to retrain the model. An ML engineer needs to implement a solution to mask credit card numbers in the files before the model is retrained.

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

A. Create a discovery job in Amazon Macie. Configure the job to find and mask sensitive data.

B. Create Apache Spark code to run on an AWS Glue job. Use the Sensitive Data Detection functionality in AWS Glue to find and mask sensitive data.

C. Create Apache Spark code to run on an AWS Glue job. Program the code to perform a regex operation to find and mask sensitive data.

D. Create Apache Spark code to run on an Amazon EC2 instance. Program the code to perform an operation to find and mask sensitive data.

Answer: B

QUESTION NO: 72

A company wants to develop an ML model by using tabular data from its customers. The data contains meaningful ordered features with sensitive information that should not be discarded. 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?

A. Use Amazon Made to categorize the sensitive data.

B. Prepare the data by using AWS Glue DataBrew.

C. Run an AWS Batch job to change the sensitive data to random values.

D. Run an Amazon EMR job to change the sensitive data to random values.

Answer: B

QUESTION NO: 73

A company is building a real-time data processing pipeline for an ecommerce application. The application generates a high volume of clickstream data that must be ingested, processed, and visualized in near real time. The company needs a solution that supports SQL for data processing and Jupyter notebooks for interactive analysis.

Which solution will meet these requirements?

A. Use Amazon Data Firehose to ingest the data. Create an AWS Lambda function to process the data. Store the processed data in Amazon S3. Use Amazon QuickSight to visualize the data.

B. Use Amazon Kinesis Data Streams to ingest the data. Use Amazon Data Firehose to transform the data. Use Amazon Athena to process the data. Use Amazon QuickSight to visualize the data.

C. Use Amazon Managed Streaming for Apache Kafka (Amazon MSK) to ingest the data. Use AWS Glue with PySpark to process the data. Store the processed data in Amazon S3.

Use Amazon QuickSight to visualize the data.

D. Use Amazon Managed Streaming for Apache Kafka (Amazon MSK) to ingest the data. Use Amazon Managed Service for Apache Flink to process the data. Use the built-in Flink dashboard to visualize the data.

Answer: D

QUESTION NO: 74

A company needs to perform feature engineering, aggregation, and data preparation. After the features are produced, the company must implement a solution on AWS to process and store the features. Which solution will meet these requirements?

A. Use Amazon SageMaker Feature Processing to process and ingest the data. Use SageMaker Feature Store to manage and store the features.

B. Use Amazon SageMaker Model Monitor to automatically ingest and transform the data. Create an Amazon S3 bucket to store the features in JSON format.

C. Use Amazon Managed Service for Apache Flink to transform the data and to ingest the data directly into Amazon SageMaker Feature Store. Use Feature Store to manage and store the features.

D. Use an Amazon SageMaker batch transform job to analyze, transform, and ingest the data.Create an Amazon DynamoDB table to store the features.

Answer: A

Explanation:

Amazon SageMaker Feature Processing (via processing jobs) is used to perform feature engineering and data preparation. The engineered features can then be ingested into SageMaker Feature Store, which is a purpose-built service to manage and store ML features for reuse across training and inference. This combination directly addresses the company's requirements.

QUESTION NO: 75

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. Add the tags to the model packages in the SageMaker Model Registry.

B. Create a model group for each category. Move the existing models into these category model groups.

C. Use SageMaker ML Lineage Tracking to automatically identify and tag which model groups should contain the models.

D. Create a Model Registry collection for each of the three categories. Move the existing model groups into the collections.

Answer: D

QUESTION NO: 76

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. Use the cleansing and enrichment functionalities to prepare the data.

B. Use Amazon SageMaker Ground Truth to import the datasets and to consolidate them into a single data frame. Use the human-in-the-loop capability to prepare the data.

C. Manually import and merge the datasets. Consolidate the datasets into a single data frame. Use Amazon Q Developer to generate code snippets that will prepare the data.

D. Manually import and merge the datasets. Consolidate the datasets into a single data frame. Use Amazon SageMaker data labeling to prepare the data.

Answer: A

QUESTION NO: 77

Hotspot Question

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.)

City (name)

Select... ▼
Select...
Feature splitting
Logarithmic transformation
One-hot encoding
Standardized distribution

Type_year (type of home and year the home was built)

Select... ▼
Select...
Feature splitting
Logarithmic transformation
One-hot encoding
Standardized distribution

Size of the building (square feet or square meters)

Select... ▼
Select...
Feature splitting
Logarithmic transformation
One-hot encoding
Standardized distribution

Answer:

City (name)

Select... ▼
Select...
Feature splitting
Logarithmic transformation
One-hot encoding
Standardized distribution

Type_year (type of home and year the home was built)

Select... ▼
Select...
Feature splitting
Logarithmic transformation
One-hot encoding
Standardized distribution

Size of the building (square feet or square meters)

Select... ▼
Select...
Feature splitting
Logarithmic transformation
One-hot encoding
Standardized distribution

QUESTION NO: 78

An ML engineer trained an ML model on Amazon SageMaker to detect automobile accidents from dosed-circuit TV footage. The ML engineer used SageMaker Data Wrangler to create a training dataset of images of accidents and non-accidents.

The model performed well during training and validation. However, the model is underperforming in production because of variations in the quality of the images from various

cameras.

Which solution will improve the model's accuracy in the LEAST amount of time?

A. Collect more images from all the cameras. Use Data Wrangler to prepare a new training dataset.

B. Recreate the training dataset by using the Data Wrangler corrupt image transform. Specify the impulse noise option.

C. Recreate the training dataset by using the Data Wrangler enhance image contrast transform.

Specify the Gamma contrast option.

D. Recreate the training dataset by using the Data Wrangler resize image transform. Crop all images to the same size.

Answer: C

QUESTION NO: 79

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. Use Amazon Managed Workflows for Apache Airflow (Amazon MWAA) to orchestrate the pipeline when new data is uploaded.

B. Create an Amazon EventBridge rule that has an event pattern that matches the S3 upload

.

Configure the pipeline as the target of the rule.

C. Create an S3 Lifecycle rule to transfer the data to the SageMaker training instance and to initiate training.

D. Create an AWS Lambda function that scans the S3 bucket. Program the Lambda function to initiate the pipeline when new data is uploaded.

Answer: B

QUESTION NO: 80

A company has a binary classification model in production. An ML engineer needs to develop a new version of the model.

The new model version must maximize correct predictions of positive labels and negative labels.

The ML engineer must use a metric to recalibrate the model to meet these requirements.

Which metric should the ML engineer use for the model recalibration?

A. Accuracy

B. Precision

C. Recall

D. Specificity

Answer: A

QUESTION NO: 81

A company wants to launch a new internal generative AI interface to answer user questions. The interface will be based on a popular open source large language model (LLM). Which combination of steps will deploy the interface with the LEAST operational overhead? (Choose two.)

A. Use Amazon SageMaker JumpStart to deploy the LLM.

B. Download the LLM as a .zip file. Deploy the LLM on a GPU-based Amazon EC2 instance.

C. Create a frontend HTML interface that uses an Amazon API Gateway WebSocket API with AWS Lambda functions to handle the user interaction.

D. Use Amazon QuickSight to create a UI to handle the user interaction.

E. Use Amazon Lex to create a UI to handle the user interaction.

Answer: AC

Explanation:

The least operational overhead comes from using Amazon SageMaker JumpStart to quickly deploy the open source LLM without needing to manage infrastructure, and building a lightweight frontend HTML interface with API Gateway WebSocket API and Lambda to handle user interactions efficiently. This avoids the manual setup of EC2 or unrelated services like QuickSight or Lex.

QUESTION NO: 82

An ML engineer is training an ML model to identify people's health risk based on 20 features and

1 target. The target class has two values:

- Likely to have health risk (positive class)

- Unlikely to have health risk (negative class)

The age range of people in the dataset is 30 years old to 60 years old. Age is one of the features.

The ML engineer analyzes the features. For the positive class, the difference in proportions of labels (DPL) value is (+0.9) for the age range of 40 to 45 compared with all other age ranges.

What should the ML engineer do to correct this data imbalance?

A. Oversample the positive class for the age range of 40 to 45.

B. Undersample the positive class for the age range of 40 to 45.

C. Undersample the positive class for all age ranges except 40 to 45.

D. Oversample the negative class for all age ranges except 40 to 45.

Answer: B

Explanation:

A DPL of +0.9 indicates that the positive class is heavily overrepresented in the 40-45 age range compared to other age ranges. To correct this imbalance, the solution is to undersample the positive class within the 40-45 range, reducing its dominance and improving fairness in the dataset.

QUESTION NO: 83

A data scientist is working on optimizing a model during the training process by varying

multiple parameters. The data scientist observes that, during multiple runs with identical parameters, the loss function converges to different, yet stable, values.

What should the data scientist do to improve the training process?

A. Increase the learning rate. Keep the batch size the same.

B. Decrease the learning rate. Reduce the batch size.

C. Decrease the learning rate. Keep the batch size the same.

D. Do not change the learning rate. Increase the batch size.

Answer: B

Explanation:

It is most likely that the loss function is very curvy and has multiple local minima where the training is getting stuck. Decreasing the batch size would help the data scientist stochastically get out of the local minima saddles. Decreasing the learning rate would prevent overshooting the global loss function minimum.

QUESTION NO: 84

A company wants to use Amazon SageMaker to host an ML model that runs on CPU for real-time predictions. The model will have intermittent traffic during business hours and will have periods of no traffic after business hours. The company needs a solution that will serve inference requests in the most cost-effective manner. Which hosting option will meet these requirements?

A. Deploy the model to a SageMaker real-time endpoint. Add a schedule-based auto scaling policy to handle traffic surges during business hours.

B. Deploy the model to a SageMaker Serverless Inference endpoint. Configure increased provisioned concurrency during business hours.

C. Deploy the model to a SageMaker Asynchronous Inference endpoint. Configure an auto scaling policy that scales in to zero outside business hours.

D. Deploy the model to a SageMaker real-time endpoint. Create a scheduled AWS Lambda function that activates the endpoint during business hours only.

Answer: B

Explanation:

SageMaker Serverless Inference is the most cost-effective option for models with intermittent traffic. It automatically scales down to zero when idle, so no cost is incurred outside business hours. Configuring provisioned concurrency during business hours ensures low-latency responses when traffic is expected.

QUESTION NO: 85

A company has several teams that have developed separate prediction models on their own laptops. The teams developed the models by using Python with scikit-learn and TensorFlow frameworks.

The company must rebuild the models and must integrate the models into an ML infrastructure that the company manages by using Amazon SageMaker. The company also must incorporate the models into a model registry.

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

A. Export the models from the laptops to an Amazon S3 bucket. Use an Amazon API

Gateway REST API and AWS Lambda functions with SageMaker endpoints to access the models. Register the models in the SageMaker Model Registry.

B. Import the models into the SageMaker Model Registry. Use SageMaker to run the imported models.

C. Use code from the laptops to create containers for the models. Use the bring your own container (BYOC) functionality of SageMaker to import and use the models. Register the models in the SageMaker Model Registry.

D. Import the Python-based models into SageMaker. Rebuild the scikit-learn and TensorFlow models in SageMaker. Register all the models in the SageMaker Model Registry.

Answer: D

Explanation:

The least operational overhead comes from directly importing the scikit-learn and TensorFlow models into SageMaker, rebuilding them using the respective prebuilt SageMaker frameworks, and then registering them in the SageMaker Model Registry. This leverages managed framework containers provided by SageMaker, avoids custom container management, and integrates seamlessly with the registry.

QUESTION NO: 86

A company runs Amazon SageMaker ML models that use accelerated instances. The models require real-time responses. Each model has different scaling requirements. The company must not allow a cold start for the models.

Which solution will meet these requirements?

A. Create a SageMaker Serverless Inference endpoint for each model. Use provisioned concurrency for the endpoints.

B. Create a SageMaker Asynchronous Inference endpoint for each model. Create an auto scaling policy for each endpoint.

C. Create a SageMaker endpoint. Create an inference component for each model. In the inference component settings, specify the newly created endpoint. Create an auto scaling policy for each inference component. Set the parameter for the minimum number of copies to at least 1.

D. Create an Amazon S3 bucket. Store all the model artifacts in the S3 bucket. Create a SageMaker multi-model endpoint. Point the endpoint to the S3 bucket. Create an auto scaling policy for the endpoint. Set the parameter for the minimum number of copies to at least 1.

Answer: C

QUESTION NO: 87

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 parameter. Decrease the top_k parameter.

C. Decrease the temperature parameter. Increase the top_k parameter.

D. Decrease the temperature parameter and the top_k parameter.

Answer: D

QUESTION NO: 88

An ML engineer wants to use a set of survey responses as training data for an ML classifier. All the survey responses are either "yes" or "no." The ML engineer needs to convert the responses into a feature that will produce better model training results. The ML engineer must not increase the dimensionality of the dataset.

Which methods will meet these requirements? (Choose two.)

A. Binary encoding

B. Label encoding

C. One-hot encoding

D. Statistical imputation

E. Tokenization

Answer: AB

Explanation:

Both binary encoding and label encoding convert categorical yes/no responses into numerical values without increasing dimensionality. For example, mapping yes → 1 and no → 0. Unlike one-hot encoding, which would add extra dimensions, these methods keep the dataset compact and effective for training.

QUESTION NO: 89

A financial company receives a high volume of real-time market data streams from an external provider. The streams consist of thousands of JSON records every second.

The company needs to implement a scalable solution on AWS to identify anomalous data points.

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

A. Ingest real-time data into Amazon Kinesis data streams. Use the built-in RANDOM_CUT_FOREST function in Amazon Managed Service for Apache Flink to process the data streams and to detect data anomalies.

B. Ingest real-time data into Amazon Kinesis data streams. Deploy an Amazon SageMaker endpoint for real-time outlier detection. Create an AWS Lambda function to detect anomalies. Use the data streams to invoke the Lambda function.

C. Ingest real-time data into Apache Kafka on Amazon EC2 instances. Deploy an Amazon SageMaker endpoint for real-time outlier detection. Create an AWS Lambda function to detect anomalies. Use the data streams to invoke the Lambda function.

D. Send real-time data to an Amazon Simple Queue Service (Amazon SQS) FIFO queue. Create an AWS Lambda function to consume the queue messages. Program the Lambda function to start an AWS Glue extract, transform, and load (ETL) job for batch processing and anomaly detection.

Answer: A

QUESTION NO: 90

An ML engineer notices class imbalance in an image classification training job.

What should the ML engineer do to resolve this issue?

A. Reduce the size of the dataset.

B. Transform some of the images in the dataset.

C. Apply random oversampling on the dataset.

D. Apply random data splitting on the dataset.

Answer: C

QUESTION NO: 91

A company has an Amazon S3 bucket that contains 1 TB of files from different sources. The S3 bucket contains the following file types in the same S3 folder: CSV, JSON, XLSX, and Apache Parquet.

An ML engineer must implement a solution that uses AWS Glue DataBrew to process the data.

The ML engineer also must store the final output in Amazon S3 so that AWS Glue can consume the output in the future.

Which solution will meet these requirements?

A. Use DataBrew to process the existing S3 folder. Store the output in Apache Parquet format.

B. Use DataBrew to process the existing S3 folder. Store the output in AWS Glue Parquet format.

C. Separate the data into a different folder for each file type. Use DataBrew to process each folder individually. Store the output in Apache Parquet format.

D. Separate the data into a different folder for each file type. Use DataBrew to process each folder individually. Store the output in AWS Glue Parquet format.

Answer: C

QUESTION NO: 92

A company is setting up a system to manage all of the datasets it stores in Amazon S3. The company would like to automate running transformation jobs on the data and maintaining a catalog of the metadata concerning the datasets. The solution should require the least amount of setup and maintenance.

Which solution will allow the company to achieve its goals?

A. Create an Amazon EMR cluster with Apache Hive installed. Then, create a Hive metastore and a script to run transformation jobs on a schedule.

B. Create an AWS Glue crawler to populate the AWS Glue Data Catalog. Then, author an AWS Glue ETL job, and set up a schedule for data transformation jobs.

C. Create an Amazon EMR cluster with Apache Spark installed. Then, create an Apache Hive metastore and a script to run transformation jobs on a schedule.

D. Create an Amazon SageMaker Jupyter notebook instance that transforms the data. Then, create an Apache Hive metastore and a script to run transformation jobs on a schedule.

Answer: B

Explanation:

AWS Glue is the correct answer because this option requires the least amount of setup and

maintenance since it is serverless, and it does not require management of the infrastructure.

QUESTION NO: 93
Hotspot Question
A company needs to train an ML model that will use historical transaction data to predict customer behavior.
Select the correct AWS service from the following list to perform each task on the data. Each service should be selected one time or not at all. (Select three.)
- Amazon Athena
- AWS Glue
- Amazon Kinesis Data Streams
- Amazon S3

Query the data for exploration and analysis.
Select...
- Select...
- Amazon Athena
- AWS Glue
- Amazon Kinesis Data Streams
- Amazon S3

Store the data.
Select...
- Select...
- Amazon Athena
- AWS Glue
- Amazon Kinesis Data Streams
- Amazon S3

Transform the data.
Select...
- Select...
- Amazon Athena
- AWS Glue
- Amazon Kinesis Data Streams
- Amazon S3

Answer:

Query the data for exploration and analysis.

Select...
- Select...
- **Amazon Athena**
- AWS Glue
- Amazon Kinesis Data Streams
- Amazon S3

Store the data.

Select...
- Select...
- Amazon Athena
- AWS Glue
- Amazon Kinesis Data Streams
- **Amazon S3**

Transform the data.

Select...
- Select...
- Amazon Athena
- **AWS Glue**
- Amazon Kinesis Data Streams
- Amazon S3

QUESTION NO: 94

An ML engineer needs to implement a solution to host a trained ML model. The rate of requests to the model will be inconsistent throughout the day.

The ML engineer needs a scalable solution that minimizes costs when the model is not in use.

The solution also must maintain the model's capacity to respond to requests during times of peak usage.

Which solution will meet these requirements?

A. Create AWS Lambda functions that have fixed concurrency to host the model. Configure the Lambda functions to automatically scale based on the number of requests to the model.

B. Deploy the model on an Amazon Elastic Container Service (Amazon ECS) cluster that uses AWS Fargate. Set a static number of tasks to handle requests during times of peak usage.

C. Deploy the model to an Amazon SageMaker endpoint. Deploy multiple copies of the model to the endpoint. Create an Application Load Balancer to route traffic between the different copies of the model at the endpoint.

D. Deploy the model to an Amazon SageMaker endpoint. Create SageMaker endpoint auto scaling policies that are based on Amazon CloudWatch metrics to adjust the number of instances dynamically.

Answer: D

QUESTION NO: 95

A company is working on an ML project that will include Amazon SageMaker notebook instances.

An ML engineer must ensure that the SageMaker notebook instances do not allow root access.

Which solution will prevent the deployment of notebook instances that allow root access?

A. Use IAM condition keys to stop deployments of SageMaker notebook instances that allow root access.

B. Use AWS Key Management Service (AWS KMS) keys to stop deployments of SageMaker notebook instances that allow root access.

C. Monitor resource creation by using Amazon EventBridge events. Create an AWS Lambda function that deletes all deployed SageMaker notebook instances that allow root access.

D. Monitor resource creation by using AWS CloudFormation events. Create an AWS Lambda function that deletes all deployed SageMaker notebook instances that allow root access.

Answer: A

QUESTION NO: 96

A company wants to provide services to help other businesses label images. The company wants its labeling specialists to complete human labeling tasks on AWS. How should the company register the labeling specialists to receive tasks on AWS?

A. Use AWS Data Exchange.

B. Create and use an internal workforce in Amazon SageMaker Ground Truth.

C. Create and use Amazon Mechanical Turk entities in an Amazon SageMaker human loop.

D. Use the Amazon Mechanical Turk website.

Answer: B

Explanation:

To enable labeling specialists within the company to perform tasks, the correct solution is to create and use an internal workforce in Amazon SageMaker Ground Truth. This allows the company to securely register and manage its own labeling team to receive and complete human labeling tasks.

QUESTION NO: 97

A company is using an AWS Lambda function to monitor the metrics from an ML model. An ML engineer needs to implement a solution to send an email message when the metrics breach a threshold.

Which solution will meet this requirement?

A. Log the metrics from the Lambda function to AWS CloudTrail. Configure a CloudTrail trail to send the email message.

B. Log the metrics from the Lambda function to Amazon CloudFront. Configure an Amazon CloudWatch alarm to send the email message.

C. Log the metrics from the Lambda function to Amazon CloudWatch. Configure a CloudWatch alarm to send the email message.

D. Log the metrics from the Lambda function to Amazon CloudWatch. Configure an Amazon CloudFront rule to send the email message.

Answer: C

QUESTION NO: 98
An ML engineer is deploying a trained model to an Amazon SageMaker endpoint. The ML engineer needs to receive alerts when data quality issues occur in production. Which solution will meet this requirement?

A. Configure an Amazon CloudWatch metric alarm and a corresponding action to send an Amazon Simple Notification Service (Amazon SNS) notification.

B. Integrate the SageMaker endpoint with a SageMaker Clarify processing job. Configure an Amazon CloudWatch alarm to provide alerts.

C. Configure a monitoring job in SageMaker Model Monitor. Integrate Model Monitor with Amazon CloudWatch to provide alerts.

D. Configure a data flow in SageMaker Data Wrangler. Integrate Data Wrangler with Amazon CloudWatch to provide alerts.

Answer: C

Explanation:

SageMaker Model Monitor is designed to continuously monitor deployed models for data quality issues such as data drift or violations of data constraints. By integrating Model Monitor with Amazon CloudWatch, alerts can be automatically triggered when data quality issues occur in production.

QUESTION NO: 99
A company deployed an ML model that uses the XGBoost algorithm to predict product failures.

The model is hosted on an Amazon SageMaker endpoint and is trained on normal operating data.

An AWS Lambda function provides the predictions to the company's application.

An ML engineer must implement a solution that uses incoming live data to detect decreased model accuracy over time.

Which solution will meet these requirements?

A. Use Amazon CloudWatch to create a dashboard that monitors real-time inference data and model predictions. Use the dashboard to detect drift.

B. Modify the Lambda function to calculate model drift by using real-time inference data and model predictions. Program the Lambda function to send alerts.

C. Schedule a monitoring job in SageMaker Model Monitor. Use the job to detect drift by analyzing the live data against a baseline of the training data statistics and constraints.

D. Schedule a monitoring job in SageMaker Debugger. Use the job to detect drift by analyzing the live data against a baseline of the training data statistics and constraints.

Answer: C

QUESTION NO: 100
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. Analyze prediction differences on samples.

Answer: B

QUESTION NO: 101

An ML engineer needs to create data ingestion pipelines and ML model deployment pipelines on AWS. All the raw data is stored in Amazon S3 buckets.

Which solution will meet these requirements?

A. Use Amazon Data Firehose to create the data ingestion pipelines. Use Amazon SageMaker Studio Classic to create the model deployment pipelines.

B. Use AWS Glue to create the data ingestion pipelines. Use Amazon SageMaker Studio Classic to create the model deployment pipelines.

C. Use Amazon Redshift ML to create the data ingestion pipelines. Use Amazon SageMaker Studio Classic to create the model deployment pipelines.

D. Use Amazon Athena to create the data ingestion pipelines. Use an Amazon SageMaker notebook to create the model deployment pipelines.

Answer: B

QUESTION NO: 102

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. Run a script to make an inference request for each image.

C. Create an Amazon Elastic Kubernetes Service (Amazon EKS) cluster that uses Karpenter for auto scaling. Host the model on the EKS cluster. Run a script to make an inference request for each image.

D. Create an AWS Batch job that uses an Amazon Elastic Container Service (Amazon ECS) cluster.Specify a list of images to process for each AWS Batch job.

Answer: B

QUESTION NO: 103

An ML engineer wants to use a set of survey responses as training data for an ML classifier. All the survey responses are either "yes" or "no." The ML engineer needs to convert the responses into a feature that will produce better model training results. The ML engineer must not increase the dimensionality of the dataset.

Which methods will meet these requirements? (Choose two.)

A. One-hot encoding

B. Binary encoding

C. Label encoding

D. Statistical imputation

E. Tokenization

Answer: B,C

Explanation:

Both binary encoding and label encoding convert categorical yes/no responses into numerical values without increasing dimensionality. For example, mapping yes → 1 and no → 0. Unlike one-hot encoding, which would add extra dimensions, these methods keep the dataset compact and effective for training.

QUESTION NO: 104

A data scientist uses logistic regression to build a fraud detection model. While the model accuracy is 99%, 90% of the fraud cases are not detected by the model.

What action will definitively help the model detect more than 10% of fraud cases?

A. Using undersampling to balance the dataset

B. Decreasing the class probability threshold

C. Using regularization to reduce overfitting

D. Using oversampling to balance the dataset

Answer: B

Explanation:

Decreasing the class probability threshold makes the model more sensitive and, therefore, marks more cases as the positive class, which is fraud in this case. This will increase the likelihood of fraud detection. However, it comes at the price of lowering precision.

QUESTION NO: 105

Hotspot Question

An ecommerce company is using Amazon SageMaker Clarify Foundation Model Evaluations (FMEval) to evaluate ML models.

Select the correct model evaluation task from the following list for each ecommerce use case

.

Each model evaluation task should be selected one time.

- Classification evaluation
- Open-ended generation
- Question answering
- Text summarization

Categorize customer reviews as positive, neutral, or negative sentiment.

▼
Classification evaluation
Open-ended generation
Question answering
Text summarization

Create concise product descriptions from complete manufacturer details.

▼
Classification evaluation
Open-ended generation
Question answering
Text summarization

Recommend products based on a user's browsing history.

▼
Classification evaluation
Open-ended generation
Question answering
Text summarization

Respond to specific customer queries about product features.

▼
Classification evaluation
Open-ended generation
Question answering
Text summarization

Answer:

Categorize customer reviews as positive, neutral, or negative sentiment.

▼
Classification evaluation
Open-ended generation
Question answering
Text summarization

Create concise product descriptions from complete manufacturer details.

▼
Classification evaluation
Open-ended generation
Question answering
Text summarization

Recommend products based on a user's browsing history.

▼
Classification evaluation
Open-ended generation
Question answering
Text summarization

Respond to specific customer queries about product features.

▼
Classification evaluation
Open-ended generation
Question answering
Text summarization

QUESTION NO: 106

An insurance company needs to automate claim compliance reviews because human reviews are expensive and error-prone. The company has a large set of claims and a compliance label for each. Each claim consists of a few sentences in English, many of which

contain complex related information. Management would like to use Amazon SageMaker built-in algorithms to design a machine learning supervised model that can be trained to read each claim and predict if the claim is compliant or not.

Which approach should be used to extract features from the claims to be used as inputs for the downstream supervised task?

A. Derive a dictionary of tokens from claims in the entire dataset. Apply one-hot encoding to tokens found in each claim of the training set. Send the derived features space as inputs to an Amazon SageMaker built- in supervised learning algorithm.

B. Apply Amazon SageMaker BlazingText in Word2Vec mode to claims in the training set. Send the derived features space as inputs for the downstream supervised task.

C. Apply Amazon SageMaker BlazingText in classification mode to labeled claims in the training set to derive features for the claims that correspond to the compliant and non-compliant labels, respectively.

D. Apply Amazon SageMaker Object2Vec to claims in the training set. Send the derived features space as inputs for the downstream supervised task.

Answer: D

Explanation:

Amazon SageMaker Object2Vec generalizes the Word2Vec embedding technique for words to more complex objects, such as sentences and paragraphs. Since the supervised learning task is at the level of whole claims, for which there are labels, and no labels are available at the word level, Object2Vec needs be used instead of Word2Vec.

QUESTION NO: 107

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. Tune the batch size that is used in the PutRecordBatch operation.

B. Replace the Firehose stream with an AWS DataSync task. Configure the task with enhanced fan- out consumers.

C. Increase the buffer interval of the Firehose stream from 60 seconds to 120 seconds.

D. Replace the Firehose stream with an Amazon Simple Queue Service (Amazon SQS) queue.

Answer: A

QUESTION NO: 108

An ML engineer needs to deploy ML models to get inferences from large datasets in an asynchronous manner. The ML engineer also needs to implement scheduled monitoring of the data quality of the models. The ML engineer must receive alerts when changes in data quality occur.

Which solution will meet these requirements?

A. Deploy the models by using scheduled AWS Glue jobs. Use Amazon CloudWatch alarms to monitor the data quality and to send alerts.

B. Deploy the models by using scheduled AWS Batch jobs. Use AWS CloudTrail to monitor the data quality and to send alerts.

C. Deploy the models by using Amazon Elastic Container Service (Amazon ECS) on AWS Fargate.

Use Amazon EventBridge to monitor the data quality and to send alerts.

D. Deploy the models by using Amazon SageMaker batch transform. Use SageMaker Model Monitor to monitor the data quality and to send alerts.

Answer: D

QUESTION NO: 109

A company uses Amazon SageMaker for its ML process. A compliance audit discovers that an Amazon S3 bucket for training data uses server-side encryption with S3 managed keys (SSE- S3).

The company requires customer managed keys. An ML engineer changes the S3 bucket to use server-side encryption with AWS KMS keys (SSE-KMS). The ML engineer makes no other configuration changes.

After the change to the encryption settings, SageMaker training jobs start to fail with AccessDenied errors.

What should the ML engineer do to resolve this problem?

A. Update the IAM policy that is attached to the execution role for the training jobs. Include the s3:ListBucket and s3:GetObject permissions.

B. Update the S3 bucket policy that is attached to the S3 bucket. Set the value of the aws:SecureTransport condition key to True.

C. Update the IAM policy that is attached to the execution role for the training jobs. Include the kms:Encrypt and kms:Decrypt permissions.

D. Update the IAM policy that is attached to the user that created the training jobs. Include the kms:CreateGrant permission.

Answer: C

QUESTION NO: 110

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 bucket. Manually move data that is older than 30 days to separate S3 buckets.

B. Create AWS Lambda functions to copy the time-series data into separate S3 buckets. Apply S3 Lifecycle policies to archive data that is older than 30 days to S3 Glacier Flexible

Retrieval.

C. Organize the time-series data into partitions by date prefix in the S3 bucket. Apply S3 Lifecycle policies to archive partitions that are older than 30 days to S3 Glacier Flexible Retrieval.

D. Put each day's time-series data into its own S3 bucket. Use S3 Lifecycle policies to archive S3 buckets that hold data that is older than 30 days to S3 Glacier Flexible Retrieval.

Answer: C

QUESTION NO: 111

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

QUESTION NO: 112

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

QUESTION NO: 113

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. Ensure that each model is visible in each AWS account.

B. Create a new AWS account with a new ECR repository as the central catalog. Configure ECR cross-account replication between the initial ECR repositories and the central catalog.

C. Use the Amazon SageMaker Model Registry to create a model group for models hosted in Amazon ECR. Create a new AWS account. In the new account, use the SageMaker Model Registry as the central catalog. Attach a cross-account resource policy to each model group in the initial AWS accounts.

D. Use an AWS Glue Data Catalog to store the models. Run an AWS Glue crawler to migrate the models from the ECR repositories to the Data Catalog. Configure cross-account access to the Data Catalog.

Answer: C

QUESTION NO: 114

A company has an ML model that is deployed to an Amazon SageMaker endpoint for real-time inference. The company needs to deploy a new model. The company must compare the new model's performance to the currently deployed model's performance before shifting all traffic to the new model. Which solution will meet these requirements with the LEAST operational effort?

A. Deploy the new model to a separate endpoint. Manually split traffic between the two endpoints.

B. Deploy the new model to a separate endpoint. Use Amazon CloudFront to distribute traffic between the two endpoints.

C. Deploy the new model as a shadow variant on the same endpoint as the current model. Route a portion of live traffic to the shadow model for evaluation.

D. Use AWS Lambda functions with custom logic to route traffic between the current model and the new model.

Answer: C

Explanation:

SageMaker supports shadow variant deployments, which allow a new model to run alongside the current one on the same endpoint. A portion of live traffic is mirrored to the shadow model for evaluation, while only the current model's output is returned to users. This provides the required comparison with minimal operational effort, avoiding the need for custom traffic-splitting solutions.

QUESTION NO: 115

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 address. Assign the security group to the domain.

B. Create a network ACL inbound rule to deny traffic from the specific IP address. Assign the rule to the default network Ad for the subnet where the domain is located.

C. Create a shadow variant for the domain. Configure SageMaker Inference Recommender to send traffic from the specific IP address to the shadow endpoint.

D. Create a VPC route table to deny inbound traffic from the specific IP address. Assign the route table to the domain.

Answer: B

QUESTION NO: 116

A medical company ingests streams of data from devices that monitor patients' vital signs. The company uses Amazon SageMaker and plans to prepare ML models to predict adverse events for patients. The dataset is large with thousands of features.

An ML engineer needs to run several hundred training iterations with different sets of features, different algorithms, and many potential parameters. The ML engineer must implement a solution to log the characteristics and results of each training iteration. Which solution will meet these requirements with the LEAST implementation effort?

A. Use Amazon CloudWatch to create custom metrics for the characteristics of each iteration.

B. Write the characteristics of each iteration to logs in Amazon S3. Use AWS Glue and Amazon Athena to search the logs.

C. Use the SageMaker Model Registry to track the characteristics and results of each iteration.

D. Use SageMaker Experiments to track the characteristics and results of each iteration.

Answer: D

Explanation:

SageMaker Experiments is specifically designed to track and organize ML experiments, including characteristics such as features, algorithms, parameters, and results. It provides experiment tracking with minimal implementation effort, making it the best fit for logging and comparing multiple training iterations.

QUESTION NO: 117

Hotspot Question

An ML engineer needs to use Amazon SageMaker Feature Store to create and manage features to train a model.

Select and order the steps from the following list to create and use the features in Feature Store.

Each step should be selected one time. (Select and order three.)

- Access the store to build datasets for training.
- Create a feature group.
- Ingest the records.

Step 1:
Select... ▼
Select...
Access the store to build datasets for training.
Create a feature group.
Ingest the records.

Step 2:
Select... ▼
Select...
Access the store to build datasets for training.
Create a feature group.
Ingest the records.

Step 3:
Select... ▼
Select...
Access the store to build datasets for training.
Create a feature group.
Ingest the records.

Answer:

Step 1:
Select...
Select...
Access the store to build datasets for training.
Create a feature group.
Ingest the records.

Step 2:
Select...
Select...
Access the store to build datasets for training.
Create a feature group.
Ingest the records.

Step 3:
Select...
Select...
Access the store to build datasets for training.
Create a feature group.
Ingest the records.

QUESTION NO: 118

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. Use the Latent Dirichlet Allocation (LDA) algorithm to identify and extract relevant keywords.

B. Use Amazon SageMaker and the BlazingText algorithm. Apply custom pre-processing steps for stemming and removal of stop words. Calculate term frequency-inverse document frequency (TF- IDF) scores to identify and extract relevant keywords.

C. Store the documents in an Amazon S3 bucket. Create AWS Lambda functions to process the documents and to run Python scripts for stemming and removal of stop words. Use bigram and trigram techniques to identify and extract relevant keywords.

D. Use Amazon Comprehend custom entity recognition and key phrase extraction to identify and extract relevant keywords.

Answer: D

QUESTION NO: 119

A company is interested in building a fraud detection model. Currently, the data scientist does not have a sufficient amount of information due to the low number of fraud cases.

Which method is MOST likely to detect the GREATEST number of valid fraud cases?

A. Oversampling using bootstrapping

B. Undersampling

C. Oversampling using SMOTE

D. Class weight adjustment

Answer: C

Explanation:

With datasets that are not fully populated, the Synthetic Minority Over-sampling Technique (SMOTE. adds new information by adding synthetic data points to the minority class. This technique would be the most effective in this scenario.

QUESTION NO: 120

A manufacturing company uses an ML model to determine whether products meet a standard for quality. The model produces an output of "Passed" or "Failed." Robots separate the products into the two categories by using the model to analyze photos on the assembly line.

Which metrics should the company use to evaluate the model's performance? (Choose two.)

A. Precision and recall

B. Root mean square error (RMSE) and mean absolute percentage error (MAPE)

C. Accuracy and F1 score

D. Bilingual Evaluation Understudy (BLEU) score

E. Perplexity

Answer: AC

QUESTION NO: 121

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

QUESTION NO: 122

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 categories. Automatically assign each color a unique integer.

B. Implement padding to ensure that all color feature vectors have the same length.

C. Perform dimensionality reduction on the color categories.

D. One-hot encode the color categories to transform the color scheme feature into a binary

matrix.

Answer: D

QUESTION NO: 123

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

QUESTION NO: 124

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 Instance. Run the core nodes and task nodes on Spot Instances.

D. Run the primary node and core nodes on On-Demand Instances. Run the task nodes on Spot Instances.

Answer: D

QUESTION NO: 125

A company is developing a new ML model that uses the XGBoost algorithm. The company will train the model on data that is stored in an Amazon S3 bucket. The data is in a nested JSON format.

An ML engineer needs to convert the JSON files into a tabular format.

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

A. Create an AWS Glue PySpark job that uses the Relationalize transform to convert the files.

B. Write custom Scala code to convert the files. Use Amazon EMR Serverless to run the Scala code.

C. Create an AWS Lambda function that uses a Python runtime and invokes the reduce() function to convert the files. Invoke the Lambda function.

D. Create an Amazon Athena database that is based on the JSON files. Use the Athena flatten function to convert the data.

Answer: A

Explanation:

The AWS Glue PySpark Relationalize transform is purpose-built to convert nested JSON into tabular format with minimal operational overhead. It automates the flattening process without requiring custom code or complex infrastructure, making it the most efficient solution for preparing the data for XGBoost training.

QUESTION NO: 126

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 bucket. Query the objects from the table.

B. Create a new S3 bucket for processed data. Set up S3 replication from the central S3 bucket to the new S3 bucket. Use S3 Object Lambda to query the objects based on transaction date.

C. Create a new S3 bucket for processed data. Use AWS Glue for Apache Spark to create a job to query the CSV objects based on transaction date. Configure the job to store the results in the new S3 bucket. Query the objects from the new S3 bucket.

D. Create a new S3 bucket for processed data. Use Amazon Data Firehose to transfer the data from the central S3 bucket to the new S3 bucket. Configure Firehose to run an AWS Lambda function to query the data based on transaction date.

Answer: A

QUESTION NO: 127

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 permissions. Attach the role to each notebook instance that the team uses.

B. Create a single IAM group. Add the data scientists to the group. Associate the group with each notebook instance that the team uses.

C. Create a single IAM user. Attach the AdministratorAccess AWS managed IAM policy to the user.

Configure each notebook instance to use the IAM user.

D. Create a single IAM group. Add the data scientists to the group. Create an IAM role. Attach the AdministratorAccess AWS managed IAM policy to the role. Associate the role with the group.Associate the group with each notebook instance that the team uses.

Answer: A

QUESTION NO: 128

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 table. Configure DynamoDB Streams to invoke an AWS Lambda function to update S3 bucket policies.

C. Use Lake Formation to authorize AWS Glue to access the S3 bucket. Configure Lake Formation tags to map ML engineers to their campaigns.

D. Configure S3 bucket policies to restrict access to the S3 bucket based on the ML engineers' campaigns.

Answer: C

QUESTION NO: 129

A company wants to reduce the cost of its containerized ML applications. The applications use ML models that run on Amazon EC2 instances, AWS Lambda functions, and an Amazon Elastic Container Service (Amazon ECS) cluster. The EC2 workloads and ECS workloads use Amazon Elastic Block Store (Amazon EBS) volumes to save predictions and artifacts.

An ML engineer must identify resources that are being used inefficiently. The ML engineer also must generate recommendations to reduce the cost of these resources.

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

A. Create code to evaluate each instance's memory and compute usage.

B. Add cost allocation tags to the resources. Activate the tags in AWS Billing and Cost Management.

C. Check AWS CloudTrail event history for the creation of the resources.

D. Run AWS Compute Optimizer.

Answer: D

QUESTION NO: 130

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. Create tracking entities for the approval process.

C. Use SageMaker Model Monitor to evaluate the performance of the model and to manage the approval.

D. Use SageMaker Pipelines. When a model version is registered, use the AWS SDK to change the approval status to "Approved."

Answer: D

QUESTION NO: 131

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. K-means clustering

D. Neural Topic Model (NTM)

Answer: A

QUESTION NO: 132

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

QUESTION NO: 133

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 SageMaker. Create a set of AWS Lambda functions to identify and remove the sensitive data.

B. Deploy the model on an Amazon Elastic Container Service (Amazon ECS) cluster that uses AWS Fargate. Create an AWS Batch job to identify and remove the sensitive data.

C. Use Amazon Macie to identify the sensitive data. Create a set of AWS Lambda functions to remove the sensitive data.

D. Use Amazon Comprehend to identify the sensitive data. Launch Amazon EC2 instances to remove the sensitive data.

Answer: C

QUESTION NO: 134

A company must install a custom script on any newly created Amazon SageMaker notebook instances.

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

A. Create a lifecycle configuration script to install the custom script when a new SageMaker notebook is created. Attach the lifecycle configuration to every new SageMaker notebook as part of the creation steps.

B. Create a custom Amazon Elastic Container Registry (Amazon ECR) image that contains the custom script. Push the ECR image to a Docker registry. Attach the Docker image to a SageMaker Studio domain. Select the kernel to run as part of the SageMaker notebook.

C. Create a custom package index repository. Use AWS CodeArtifact to manage the installation of the custom script. Set up AWS PrivateLink endpoints to connect CodeArtifact to the SageMaker instance. Install the script.

D. Store the custom script in Amazon S3. Create an AWS Lambda function to install the custom script on new SageMaker notebooks. Configure Amazon EventBridge to invoke the Lambda function when a new SageMaker notebook is initialized.

Answer: A

QUESTION NO: 135

A company stores training data as a .csv file in an Amazon S3 bucket. The company must encrypt the data and must control which applications have access to the encryption key.

Which solution will meet these requirements?

A. Create a new SSH access key. Use the AWS Encryption CLI with a reference to the new access key to encrypt the file.

B. Create a new API key by using the Amazon API Gateway CreateApiKey API operation. Use the AWS CLI with a reference to the new API key to encrypt the file.

C. Create a new IAM role. Attach a policy that allows the AWS Key Management Service (AWS KMS) GenerateDataKey action. Use the role to encrypt the file.

D. Create a new AWS Key Management Service (AWS KMS) key. Use the AWS Encryption CLI with a reference to the new KMS key to encrypt the file.

Answer: D

Explanation:

The correct approach is to create a new AWS KMS key and use the AWS Encryption CLI to encrypt the file with that key. This ensures the data in S3 is encrypted, and access to the encryption key can be controlled through KMS key policies and IAM permissions, meeting both encryption and access control requirements.

QUESTION NO: 136

A company is building a deep learning model on Amazon SageMaker. The company uses a large amount of data as the training dataset. The company needs to optimize the model's hyperparameters to minimize the loss function on the validation dataset.

Which hyperparameter tuning strategy will accomplish this goal with the LEAST computation time?

A. Hyperband

B. Grid search

C. Bayesian optimization

D. Random search

Answer: A

QUESTION NO: 137

A company has an ML model that uses historical transaction data to predict customer behavior.

An ML engineer is optimizing the model in Amazon SageMaker to enhance the model's predictive accuracy. The ML engineer must examine the input data and the resulting predictions to identify trends that could skew the model's performance across different demographics.

Which solution will provide this level of analysis?

A. Use Amazon CloudWatch to monitor network metrics and CPU metrics for resource optimization during model training.

B. Create AWS Glue DataBrew recipes to correct the data based on statistics from the model output.

C. Use SageMaker Clarify to evaluate the model and training data for underlying patterns that might affect accuracy.

D. Create AWS Lambda functions to automate data pre-processing and to ensure consistent quality of input data for the model.

Answer: C

QUESTION NO: 138

A company needs to use Retrieval Augmented Generation (RAG) to supplement an open source large language model (LLM) that runs on Amazon Bedrock. The company's data for RAG is a set of documents in an Amazon S3 bucket. The documents consist of .csv files and .docx files.

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

A. Create a pipeline in Amazon SageMaker Pipelines to generate a new model. Call the new model from Amazon Bedrock to perform RAG queries.

B. Convert the data into vectors. Store the data in an Amazon Neptune database. Connect the database to Amazon Bedrock. Call the Amazon Bedrock API to perform RAG queries.

C. Fine-tune an existing LLM by using an AutoML job in Amazon SageMaker. Configure the S3 bucket as a data source for the AutoML job. Deploy the LLM to a SageMaker endpoint. Use the endpoint to perform RAG queries.

D. Create a knowledge base for Amazon Bedrock. Configure a data source that references the S3 bucket. Use the Amazon Bedrock API to perform RAG queries.

Answer: D

QUESTION NO: 139

A company shares Amazon SageMaker Studio notebooks that are accessible through a VPN.

The company must enforce access controls to prevent malicious actors from exploiting presigned URLs to access the notebooks.

Which solution will meet these requirements?

A. Set up Studio client IP validation by using the aws:sourceIp IAM policy condition.

B. Set up Studio client VPC validation by using the aws:sourceVpc IAM policy condition.

C. Set up Studio client role endpoint validation by using the aws:PrimaryTag IAM policy condition.

D. Set up Studio client user endpoint validation by using the aws:PrincipalTag IAM policy condition.

Answer: A

QUESTION NO: 140

A company is gathering audio, video, and text data in various languages. The company needs to use a large language model (LLM) to summarize the gathered data that is in Spanish.

Which solution will meet these requirements in the LEAST amount of time?

A. Train and deploy a model in Amazon SageMaker to convert the data into English text. Train and deploy an LLM in SageMaker to summarize the text.

B. Use Amazon Transcribe and Amazon Translate to convert the data into English text. Use Amazon Bedrock with the Jurassic model to summarize the text.

C. Use Amazon Rekognition and Amazon Translate to convert the data into English text. Use Amazon Bedrock with the Anthropic Claude model to summarize the text.

D. Use Amazon Comprehend and Amazon Translate to convert the data into English text. Use Amazon Bedrock with the Stable Diffusion model to summarize the text.

Answer: B

QUESTION NO: 141

A company needs to extract entities from a PDF document to build a classifier model.

Which solution will extract and store the entities in the LEAST amount of time?

A. Use Amazon Comprehend to extract the entities. Store the output in Amazon S3.

B. Use an open source AI optical character recognition (OCR) tool on Amazon SageMaker to extract the entities. Store the output in Amazon S3.

C. Use Amazon Textract to extract the entities. Use Amazon Comprehend to convert the entities to text. Store the output in Amazon S3.

D. Use Amazon Textract integrated with Amazon Augmented AI (Amazon A2I) to extract the entities.Store the output in Amazon S3.

Answer: C

QUESTION NO: 142

A company is using Amazon EMR. The company has a large dataset in Amazon S3 that needs to be ingested into Amazon SageMaker Feature Store. The dataset contains historical data and real-time streaming data.

The company must ensure that the Feature Store online store is updated with the most recent data as soon as the data becomes available. The company also must maintain a complete Feature Store offline store for batch processing.

Which solution will meet these requirements?

A. Use the PutRecord API in Feature Store Runtime to ingest all the data into the online store.

B. Use the PutRecord API in Feature Store Runtime to ingest all the data into the offline store.

C. Use the Feature Store Spark connector to ingest the data as Spark DataFrames with the online store and offline store enabled.

D. Use the Feature Store Spark connector to ingest the data as Spark DataFrames with only the online store enabled.

Answer: C

Explanation:

The SageMaker Feature Store Spark connector allows ingestion of large-scale data from Amazon EMR into Feature Store as Spark DataFrames. Enabling both the online store ensures real-time updates for the latest data, while the offline store maintains the full historical dataset for batch analytics. This setup meets both low-latency and historical processing requirements.

QUESTION NO: 143

A company is planning to create several ML prediction models. The training data is stored in Amazon S3. The entire dataset is more than 5 TB in size and consists of CSV, JSON, Apache Parquet, and simple text files.

The data must be processed in several consecutive steps. The steps include complex manipulations that can take hours to finish running. Some of the processing involves natural language processing (NLP) transformations. The entire process must be automated.

Which solution will meet these requirements?

A. Process data at each step by using Amazon SageMaker Data Wrangler. Automate the process by using Data Wrangler jobs.

B. Use Amazon SageMaker notebooks for each data processing step. Automate the process by using Amazon EventBridge.

C. Process data at each step by using AWS Lambda functions. Automate the process by using AWS Step Functions and Amazon EventBridge.

D. Use Amazon SageMaker Pipelines to create a pipeline of data processing steps. Automate the pipeline by using Amazon EventBridge.

Answer: D

QUESTION NO: 144

A company has an existing Amazon SageMaker model (v1) on a production endpoint. The company develops a new model version (v2) and needs to test v2 in production before substituting v2 for v1.

The company needs to implement a solution to minimize the risk of v2 generating incorrect output in production. The solution must prevent any disruption of production traffic during the change to v2.

Which solution will meet these requirements?

A. Create a second production variant for v2. Assign 1% of the traffic to v2 and 99% of the traffic to v1. Collect all the output of v2 in an Amazon S3 bucket. If v2 performs as expected, switch all the traffic to v2.

B. Create a second production variant for v2. Assign 10% of the traffic to v2 and 90% of the traffic to v1. Collect all the output of v2 in an Amazon S3 bucket. If v2 performs as expected, switch all the traffic to v2.

C. Deploy v2 to a new endpoint. Turn on data capturing for the production endpoint. Write a script to pass 100% of input data to v2. If v2 performs as expected, deactivate the v1 endpoint and direct the traffic to v2.

D. Deploy v2 into a shadow variant that samples 100% of the inference requests. Collect all the output in an Amazon S3 bucket. If v2 performs as expected, promote v2 to production.

Answer: D

Explanation:

A shadow variant allows the new model (v2) to receive a copy of 100% of production traffic while only v1's outputs are returned to users. This enables safe side-by-side evaluation of v2 without impacting production responses, minimizing risk and ensuring no disruption of live traffic until v2 is validated and promoted.

QUESTION NO: 145

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 is experimenting with consecutive training jobs.

How can the company MINIMIZE infrastructure startup times for these jobs?

A. Use Managed Spot Training.

B. Use SageMaker managed warm pools.

C. Use SageMaker Training Compiler.

D. Use the SageMaker distributed data parallelism (SMDDP) library.

Answer: B

QUESTION NO: 146

A company needs to host a custom ML model to perform forecast analysis. The forecast analysis will occur with predictable and sustained load during the same 2-hour period every day.

Multiple invocations during the analysis period will require quick responses. The company needs AWS to manage the underlying infrastructure and any auto scaling activities.

Which solution will meet these requirements?

A. Schedule an Amazon SageMaker batch transform job by using AWS Lambda.

B. Configure an Auto Scaling group of Amazon EC2 instances to use scheduled scaling.

C. Use Amazon SageMaker Serverless Inference with provisioned concurrency.

D. Run the model on an Amazon Elastic Kubernetes Service (Amazon EKS) cluster on Amazon EC2 with pod auto scaling.

Answer: C

QUESTION NO: 147

A company has a Retrieval Augmented Generation (RAG) application that uses a vector database to store embeddings of documents. The company must migrate the application to AWS and must implement a solution that provides semantic search of text files. The company has already migrated the text repository to an Amazon S3 bucket.

Which solution will meet these requirements?

A. Use an AWS Batch job to process the files and generate embeddings. Use AWS Glue to store the embeddings. Use SQL queries to perform the semantic searches.

B. Use a custom Amazon SageMaker notebook to run a custom script to generate embeddings. Use SageMaker Feature Store to store the embeddings. Use SQL queries to perform the semantic searches.

C. Use the Amazon Kendra S3 connector to ingest the documents from the S3 bucket into Amazon Kendra. Query Amazon Kendra to perform the semantic searches.

D. Use an Amazon Textract asynchronous job to ingest the documents from the S3 bucket. Query Amazon Textract to perform the semantic searches.

Answer: C

QUESTION NO: 148

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 needs to use the central model registry to manage different versions of models in the application.

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

A. Create a separate Amazon Elastic Container Registry (Amazon ECR) repository for each model.

B. Use Amazon Elastic Container Registry (Amazon ECR) and unique tags for each model

version.

C. Use the SageMaker Model Registry and model groups to catalog the models.

D. Use the SageMaker Model Registry and unique tags for each model version.

Answer: C

QUESTION NO: 149

A company is planning to use an Amazon SageMaker prebuilt algorithm to create a recommendation model. The algorithm must be able to make predictions on high-dimensional sparse data. Which SageMaker algorithm should the company choose for the recommendation model?

A. K-nearest neighbors (k-NN)

B. Factorization Machines

C. Principal component analysis (PCA)

D. Sequence-to-Sequence (seq2seq)

Answer: B

Explanation:

The Factorization Machines algorithm in SageMaker is specifically designed for recommendation systems and works well with high-dimensional sparse data such as user-item interactions. It efficiently models variable interactions and is the best choice for building a recommendation model in this scenario.

QUESTION NO: 150

An ML engineer needs to use metrics to assess the quality of a time-series forecasting model.

Which metrics apply to this model? (Choose two.)

A. Recall

B. LogLoss

C. Root mean square error (RMSE)

D. InferenceLatency

E. Average weighted quantile loss (wQL)

Answer: CE

QUESTION NO: 151

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

QUESTION NO: 152

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. Program the Lambda function to initiate the pipeline when new data is uploaded.

C. Create an Amazon EventBridge rule that has an event pattern that matches the S3 upload. Configure the pipeline as the target of the rule.

D. Use Amazon Managed Workflows for Apache Airflow (Amazon MWAA) to orchestrate the pipeline when new data is uploaded.

Answer: C

QUESTION NO: 153

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

QUESTION NO: 154

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 set. Use these values to normalize the production samples.

C. Calculate a new set of min-max normalization statistics from a batch of production samples. Use these values to normalize all the production samples.

D. Calculate a new set of min-max normalization statistics from each production sample. Use these values to normalize all the production samples.

Answer: B

QUESTION NO: 155

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 Autopilot. Report the model's achieved performance.

B. Implement custom scripts to perform data pre-processing, multiple linear regression, and performance evaluation. Run the scripts on Amazon EC2 instances.

C. Configure Amazon Macie to analyze the dataset and to create a model. Report the model's achieved performance.

D. Select a model from Amazon Bedrock. Tune the model with the data. Report the model's achieved performance.

Answer: A

QUESTION NO: 156

An ML engineer is using a training job to fine-tune a deep learning model in Amazon SageMaker Studio. The ML engineer previously used the same pre-trained model with a similar dataset. The ML engineer expects vanishing gradient, underutilized GPU, and overfitting problems.

The ML engineer needs to implement a solution to detect these issues and to react in predefined ways when the issues occur. The solution also must provide comprehensive real-time metrics during the training.

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

A. Use TensorBoard to monitor the training job. Publish the findings to an Amazon Simple Notification Service (Amazon SNS) topic. Create an AWS Lambda function to consume the findings and to initiate the predefined actions.

B. Use Amazon CloudWatch default metrics to gain insights about the training job. Use the metrics to invoke an AWS Lambda function to initiate the predefined actions.

C. Expand the metrics in Amazon CloudWatch to include the gradients in each training step. Use the metrics to invoke an AWS Lambda function to initiate the predefined actions.

D. Use SageMaker Debugger built-in rules to monitor the training job. Configure the rules to initiate the predefined actions.

Answer: D

QUESTION NO: 157

An ML engineer needs to use an ML model to predict the price of apartments in a specific location.

Which metric should the ML engineer use to evaluate the model's performance?

A. Accuracy

B. Area Under the ROC Curve (AUC)

C. F1 score

D. Mean absolute error (MAE)

Answer: D