

complete your programming course

about resources, doubts and more!

MYEXAM.FK

Amazon

(AWS Certified Machine Learning Engineer - Associate MLA-C01)

AWS Certified Machine Learning Engineer - Associate MLA-C01

Total: **113 Questions**

Question: 1

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: 2

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: 3

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: 4

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: 5

HOTSPOT -

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.

Explanation:

Step 1: Use AWS Glue crawlers to infer the schemas and available columns.

AWS Glue Crawlers scan data in Amazon S3 to infer schema and structure, automatically creating metadata tables in the Glue Data Catalog.

This step is ideal as the first step in a data pipeline when you're preparing raw data.

Automatically discovers schemas, data types, and partitions.

Step 2: Use AWS Glue DataBrew for data cleaning and feature engineering.

AWS Glue DataBrew is a visual data preparation tool that lets users clean and transform data without writing code.

After schema discovery, the next logical step is to transform and clean the data — removing nulls, standardizing formats, etc.

Data cleaning, transformation, feature engineering.

Preview transformations before applying them.

Step 3: Store the resulting data back in Amazon S3.

After cleaning and transforming the data with DataBrew (or any other tool), you typically store the refined dataset back to Amazon S3.

Amazon S3 acts as a central data lake for storage, sharing, or further processing (e.g., using Athena or SageMaker).

Question: 6

HOTSPOT -

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.

Explanation:

Step 1: Create a feature group.

A Feature Group is like a schema for storing features. It defines the structure of the features you want to store, including:

Feature names

Data types

Record identifiers

Event timestamps

Creating a feature group is the first step before you can store any data in the Feature Store.

Think of this like defining a table in a relational database before inserting data into it.

Step 2: **Ingest the records.**

Once the Feature Group is created, you can ingest (insert) the actual data/records into it.

This means populating the Feature Group with rows of data where each row represents a specific entity (e.g., a user, product, or session).

Ingested records are stored in online and/or offline stores, depending on how the Feature Group is configured. Step 3: **Access the store to build datasets for training.**

After data is ingested, you can retrieve the features from the Feature Store (usually the offline store) to build training datasets for machine learning models.

This is typically done before launching training jobs in Amazon SageMaker.

Question: 7

HOTSPOT -

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:

MY EXAM.FX

Step 1:

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:

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:

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.

Explanation:

Step 1: An S3 event notification invokes the pipeline when new data is uploaded.

In AWS, an S3 event notification is often used to trigger a pipeline when new data is uploaded to a bucket. This is a common approach to automate ML workflows by detecting new training data and triggering retraining processes.

Step 2: SageMaker retrains the model by using the data in the S3 bucket.

Once the pipeline is triggered by the S3 event, SageMaker retrieves the new data from the S3 bucket and retrains the ML model. This is a key step in continuous model training workflows.

Step 3: The pipeline deploys the model to a SageMaker endpoint.

After retraining, the final step in an ML pipeline is often deployment. The trained model is deployed to a SageMaker endpoint so it can be used for inference in real-time applications.

Question: 8

HOTSPOT -

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

Explanation:

1. Text representation of basic units of data processed by LLMs

Token.

Tokens are the smallest units of text that LLMs process.

A token can be a word, part of a word, or even punctuation depending on the tokenizer.

LLMs (like GPT) break input text into tokens before performing predictions.

2. High-dimensional vectors that contain the semantic meaning of text

Embedding.

Embeddings are numerical representations of text in a high-dimensional space.

These vectors preserve the semantic meaning, allowing LLMs or other models to understand the context and relationships between words or phrases.

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

Retrieval Augmented Generation (RAG)

RAG is a technique where the model retrieves external documents or data (e.g., from a vector store or a knowledge base) before generating a response.

This helps the model augment its output with factual or contextual information it might not have memorized.

Question: 9

HOTSPOT -

An ML engineer is working on an ML model to predict the prices of similarly sized homes. The model will base predictions on several features. The ML engineer will use the following feature engineering techniques to estimate the prices of the homes:

- Feature splitting
- Logarithmic transformation
- One-hot encoding
- Standardized distribution

Select the correct feature engineering techniques for the following list of features. Each feature engineering technique should be selected one time or not at all (Select three.)

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:

MY EXAM

City (name)

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

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

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

Size of the building (square feet or square meters)

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

Explanation:

1. City (name) : **One-hot encoding.**

The city name is a categorical variable (nominal data) that represents a set of distinct categories without any inherent order.

One-hot encoding is a common technique for handling categorical variables where each category is converted into a binary (0 or 1) vector.

This avoids assigning any numerical relationship between categories, which would be misleading in a model.

Example: If we have three cities—New York, Los Angeles, and Chicago—one-hot encoding would transform them into:

New York → [1, 0, 0]

Los Angeles → [0, 1, 0]

Chicago → [0, 0, 1]

2. Type_year (type of home and year the home was built) : **Feature splitting.**

This feature contains two types of information: (1) the type of home and (2) the year it was built.

Feature splitting means breaking this single column into two separate features: one for home type and one for year.

Example:

"Apartment_2000" → Apartment (categorical feature) and 2000 (numerical feature)

This improves data representation and allows the model to process each type of information correctly.

3. Size of the building (square feet or square meters) : **Logarithmic transformation.**

The size of a building is a continuous numerical variable that can have a wide range of values.

Logarithmic transformation helps normalize skewed distributions where there are a few extremely large values.

It reduces the impact of large outliers and makes the data more normally distributed, improving model performance.

Example:

Raw data: 500, 1000, 5000, 10000, 50000

After log transformation: 2.7, 3.0, 3.7, 4.0, 4.7 (values are compressed in a more manageable range).

Question: 10

Case study -

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

The dataset has a class imbalance that affects the learning of the model's algorithm. Additionally, many of the features have interdependencies. The algorithm is not capturing all the desired underlying patterns in the data. Which AWS service or feature can aggregate the data from the various data sources?

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

Answer: A

Question: 11

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: 12

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: 13

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: 14

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: 15

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: 16

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: 17

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

Question: 18

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: 19

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: 20

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: 21

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: 22

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: 23

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: 24

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: 25

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: 26

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: 27

A company uses a hybrid cloud environment. A model that is deployed on premises uses data in Amazon S3 to provide customers with a live conversational engine.

The model is using sensitive data. An ML engineer needs to implement a solution to identify and remove the sensitive data.

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

- A. Deploy the model on Amazon SageMaker. 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: 28

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: 29

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: 30

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: 31

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: 32

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: 33

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: 34

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

MYEXAM.FE

Thank you

Thank you for being so interested in the premium exam material. I'm glad to hear that you found it informative and helpful.

But Wait

I wanted to let you know that there is more content available in the full version. The full paper contains additional sections and information that you may find helpful, and I encourage you to download it to get a more comprehensive and detailed view of all the subject matter.

[Download Full Version Now](#)



Future is Secured

100% Pass Guarantee



24/7 Customer Support

Mail us - certyiqofficial@gmail.com



Free Updates

Lifetime Free Updates!

Total: **113 Questions**