Vendor: Amazon

Exam Code: MLS-C01

Exam Name: AWS Certified Machine Learning - Specialty (MLS-C01) Exam

Version: DEMO

QUESTION 1

A company is using Amazon Polly to translate plaintext documents to speech for automated company announcements. However, company acronyms are being mispronounced in the current documents.

How should a Machine Learning Specialist address this issue for future documents?

A. Convert current documents to SSML with pronunciation tags.
B. Create an appropriate pronunciation lexicon.
C. Output speech marks to guide in pronunciation.
D. Use Amazon Lex to preprocess the text files for pronunciation

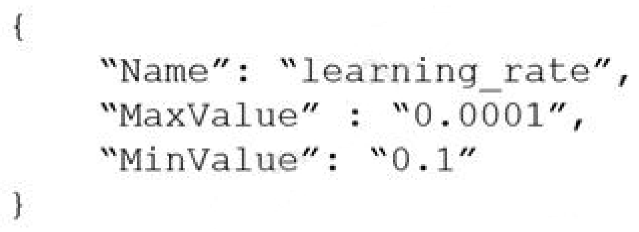
Answer: B

Explanation:

Pronunciation lexicons enable you to customize the pronunciation of words. Amazon Polly provides API operations that you can use to store lexicons in an AWS region. Those lexicons are then specific to that particular region. You can use one or more of the lexicons from that region when synthesizing the text by using the SynthesizeSpeech operation. This applies the specified lexicon to the input text before the synthesis begins.
https://docs.aws.amazon.com/polly/latest/dg/managing-lexicons.html
https://www.smashingmagazine.com/2019/08/text-to-speech-aws/

QUESTION 2

A machine learning (ML) specialist is using Amazon SageMaker hyperparameter optimization (HPO) to improve a model’s accuracy. The learning rate parameter is specified in the following HPO configuration:

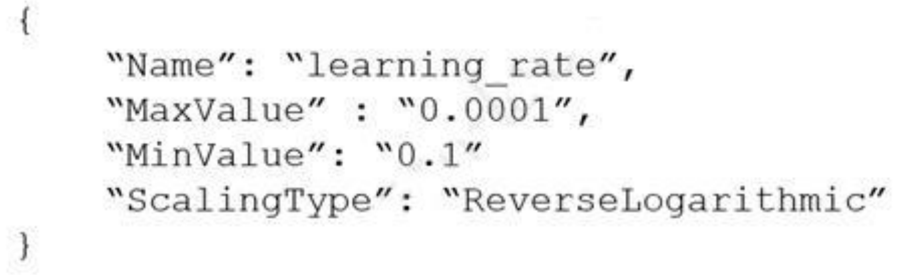


```
{
    “Name”: “learning_rate”,
    “MaxValue” : “0.0001”,
    “MinValue”: “0.1”
}
```

During the results analysis, the ML specialist determines that most of the training jobs had a learning rate between 0.01 and 0.1. The best result had a learning rate of less than 0.01. Training jobs need to run regularly over a changing dataset. The ML specialist needs to find a tuning mechanism that uses different learning rates more evenly from the provided range between MinValue and MaxValue.

Which solution provides the MOST accurate result?

A. Modify the HPO configuration as follows:



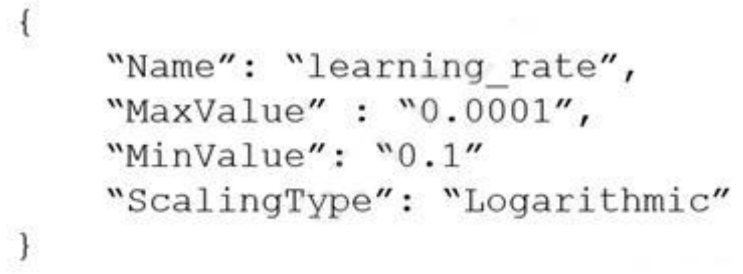
```
{
        "Name": "learning_rate",
        "MaxValue" : "0.0001",
        "MinValue": "0.1"
        "ScalingType": "ReverseLogarithmic"
}
```

Select the most accurate hyperparameter configuration form this HPO job.

B. Run three different HPO jobs that use different learning rates form the following intervals for MinValue and MaxValue while using the same number of training jobs for each HPO job:
   - [0.01, 0.1]
   - [0.001, 0.01]
   - [0.0001, 0.001]

   Select the most accurate hyperparameter configuration form these three HPO jobs.

C. Modify the HPO configuration as follows:



```
{
        "Name": "learning_rate",
        "MaxValue" : "0.0001",
        "MinValue": "0.1"
        "ScalingType": "Logarithmic"
}
```

   Select the most accurate hyperparameter configuration form this training job.

D. Run three different HPO jobs that use different learning rates form the following intervals for MinValue and MaxValue. Divide the number of training jobs for each HPO job by three:
   - [0.01, 0.1]
   - [0.001, 0.01]
   - [0.0001, 0.001]

   Select the most accurate hyperparameter configuration form these three HPO jobs.

Answer: C

Explanation:

Choose logarithmic scaling when you are searching a range that spans several orders of magnitude. For example, if you are tuning a Tune a linear learner model model, and you specify a range of values between .0001 and 1.0 for the learning_rate hyperparameter, searching uniformly on a logarithmic scale gives you a better sample of the entire range than searching on a linear scale would, because searching on a linear scale would, on average, devote 90 percent of your training budget to only the values between .1 and 1.0, leaving only 10 percent of your training budget for the values between .0001 and .1.

https://docs.aws.amazon.com/sagemaker/latest/dg/automatic-model-tuning-define-ranges.html

QUESTION 3

A data scientist has a dataset of machine part images stored in Amazon Elastic File System (Amazon EFS). The data scientist needs to use Amazon SageMaker to create and train an image classification machine learning model based on this dataset. Because of budget and time constraints, management wants the data scientist to create and train a model with the least number of steps and integration work required.

How should the data scientist meet these requirements?

A. Mount the EFS file system to a SageMaker notebook and run a script that copies the data to an Amazon FSx for Lustre file system. Run the SageMaker training job with the FSx for Lustre file system as the data source.
B. Launch a transient Amazon EMR cluster. Configure steps to mount the EFS file system and copy the data to an Amazon S3 bucket by using S3DistCp. Run the SageMaker training job with Amazon S3 as the data source.
C. Mount the EFS file system to an Amazon EC2 instance and use the AWS CLI to copy the data to an Amazon S3 bucket. Run the SageMaker training job with Amazon S3 as the data source.
D. Run a SageMaker training job with an EFS file system as the data source.

Answer: D
Explanation:
Amazon SageMaker now supports Amazon Elastic File System (Amazon EFS) and Amazon FSx for Lustre file systems as data sources for training machine learning models on SageMaker.
https://aws.amazon.com/blogs/machine-learning/speed-up-training-on-amazon-sagemaker-using-amazon-efs-or-amazon-fsx-for-lustre-file-systems/

QUESTION 4
A company wants to create a data repository in the AWS Cloud for machine learning (ML) projects. The company wants to use AWS to perform complete ML lifecycles and wants to use Amazon S3 for the data storage. All of the company’s data currently resides on premises and is 40 TB in size.

The company wants a solution that can transfer and automatically update data between the on-premises object storage and Amazon S3. The solution must support encryption, scheduling, monitoring, and data integrity validation.

Which solution meets these requirements?

A. Use the S3 sync command to compare the source S3 bucket and the destination S3 bucket. Determine which source files do not exist in the destination S3 bucket and which source files were modified.
B. Use AWS Transfer for FTPS to transfer the files from the on-premises storage to Amazon S3.
C. Use AWS DataSync to make an initial copy of the entire dataset. Schedule subsequent incremental transfers of changing data until the final cutover from on premises to AWS.
D. Use S3 Batch Operations to pull data periodically from the on-premises storage. Enable S3 Versioning on the S3 bucket to protect against accidental overwrites.

Answer: C
Explanation:
Configure DataSync to make an initial copy of your entire dataset, and schedule subsequent incremental transfers of changing data until the final cut-over from on-premises to AWS.
Reference: https://aws.amazon.com/datasync/faqs/

QUESTION 5
A power company wants to forecast future energy consumption for its customers in residential properties and commercial business properties. Historical power consumption data for the last 10 years is available. A team of data scientists who performed the initial data analysis and feature selection will include the historical power consumption data and data such as weather, number of individuals on the property, and public holidays.

The data scientists are using Amazon Forecast to generate the forecasts.

Which algorithm in Forecast should the data scientists use to meet these requirements?

A. Autoregressive Integrated Moving Average (AIRMA)
B. Exponential Smoothing (ETS)
C. Convolutional Neural Network - Quantile Regression (CNN-QR)
D. Prophet

Answer: C
Explanation:
CNN-QR and DeepAR accepts related time series data (weather data, number of people on property, etc.,)
https://docs.aws.amazon.com/forecast/latest/dg/aws-forecast-choosing-recipes.html

QUESTION 6
A company that runs an online library is implementing a chatbot using Amazon Lex to provide book recommendations based on category. This intent is fulfilled by an AWS Lambda function that queries an Amazon DynamoDB table for a list of book titles, given a particular category. For testing, there are only three categories implemented as the custom slot types: "comedy," "adventure," and "documentary."

A machine learning (ML) specialist notices that sometimes the request cannot be fulfilled because Amazon Lex cannot understand the category spoken by users with utterances such as "funny," "fun," and "humor." The ML specialist needs to fix the problem without changing the Lambda code or data in DynamoDB.

How should the ML specialist fix the problem?

A. Add the unrecognized words in the enumeration values list as new values in the slot type.
B. Create a new custom slot type, add the unrecognized words to this slot type as enumeration values, and use this slot type for the slot.
C. Use the AMAZON.SearchQuery built-in slot types for custom searches in the database.
D. Add the unrecognized words as synonyms in the custom slot type.

Answer: D
Explanation:
For each intent, you can specify parameters that indicate the information that the intent needs to fulfill the user's request. These parameters, or slots, have a type. A slot type is a list of values that Amazon Lex uses to train the machine learning model to recognize values for a slot. For example, you can define a slot type called "Genres." Each value in the slot type is the name of a genre, "comedy," "adventure," "documentary," etc. You can define a synonym for a slot type value. For example, you can define the synonyms "funny" and "humorous" for the value "comedy."
https://docs.aws.amazon.com/lex/latest/dg/howitworks-custom-slots.html

QUESTION 7
A retail company wants to combine its customer orders with the product description data from its product catalog. The structure and format of the records in each dataset is different. A data analyst tried to use a spreadsheet to combine the datasets, but the effort resulted in duplicate records and records that were not properly combined. The company needs a solution that it can use to combine similar records from the two datasets and remove any duplicates.

Which solution will meet these requirements?

A. Use an AWS Lambda function to process the data.
Use two arrays to compare equal strings in the fields from the two datasets and remove any duplicates.
B. Create AWS Glue crawlers for reading and populating the AWS Glue Data Catalog.
Call the AWS Glue SearchTables API operation to perform a fuzzy-matching search on the two datasets, and cleanse the data accordingly.
C. Create AWS Glue crawlers for reading and populating the AWS Glue Data Catalog.
Use the FindMatches transform to cleanse the data.
D. Create an AWS Lake Formation custom transform.
Run a transformation for matching products from the Lake Formation console to cleanse the data automatically.

Answer: D
Explanation:
AWS Lake Formation FindMatches is a new machine learning (ML) transform that enables you to match records across different datasets as well as identify and remove duplicate records, with little to no human intervention.
https://aws.amazon.com/lake-formation/features/

QUESTION 8
A telecommunications company is developing a mobile app for its customers. The company is using an Amazon SageMaker hosted endpoint for machine learning model inferences.

Developers want to introduce a new version of the model for a limited number of users who subscribed to a preview feature of the app. After the new version of the model is tested as a preview, developers will evaluate its accuracy. If a new version of the model has better accuracy, developers need to be able to gradually release the new version for all users over a fixed period of time.

How can the company implement the testing model with the LEAST amount of operational overhead?

A. Update the ProductionVariant data type with the new version of the model by using the CreateEndpointConfig operation with the InitialVariantWeight parameter set to 0.
Specify the TargetVariant parameter for InvokeEndpoint calls for users who subscribed to the preview feature.
When the new version of the model is ready for release, gradually increase InitialVariantWeight until all users have the updated version.
B. Configure two SageMaker hosted endpoints that serve the different versions of the model.
Create an Application Load Balancer (ALB) to route traffic to both endpoints based on the TargetVariant query string parameter.
Reconfigure the app to send the TargetVariant query string parameter for users who subscribed to the preview feature.
When the new version of the model is ready for release, change the ALB's routing algorithm to

weighted until all users have the updated version.

C. Update the DesiredWeightsAndCapacity data type with the new version of the model by using the UpdateEndpointWeightsAndCapacities operation with the DesiredWeight parameter set to 0.
Specify the TargetVariant parameter for InvokeEndpoint calls for users who subscribed to the preview feature.
When the new version of the model is ready for release, gradually increase DesiredWeight until all users have the updated version.

D. Configure two SageMaker hosted endpoints that serve the different versions of the model.
Create an Amazon Route 53 record that is configured with a simple routing policy and that points to the current version of the model.
Configure the mobile app to use the endpoint URL for users who subscribed to the preview feature and to use the Route 53 record for other users.
When the new version of the model is ready for release, add a new model version endpoint to Route 53, and switch the policy to weighted until all users have the updated version.

Answer: C
Explanation:
Step 4: Increase traffic to the best model
Now that we have determined that Variant2 performs better than Variant1, we shift more traffic to it. We can continue to use TargetVariant to invoke a specific model variant, but a simpler approach is to update the weights assigned to each variant by calling UpdateEndpointWeightsAndCapacities.
https://docs.aws.amazon.com/sagemaker/latest/dg/model-ab-testing.html

QUESTION 9
A library is developing an automatic book-borrowing system that uses Amazon Rekognition. Images of library members' faces are stored in an Amazon S3 bucket. When members borrow books, the Amazon Rekognition CompareFaces API operation compares real faces against the stored faces in Amazon S3.

The library needs to improve security by making sure that images are encrypted at rest. Also, when the images are used with Amazon Rekognition. they need to be encrypted in transit. The library also must ensure that the images are not used to improve Amazon Rekognition as a service.

How should a machine learning specialist architect the solution to satisfy these requirements?

A. Enable server-side encryption on the S3 bucket.
Submit an AWS Support ticket to opt out of allowing images to be used for improving the service, and follow the process provided by AWS Support.

B. Switch to using an Amazon Rekognition collection to store the images.
Use the IndexFaces and SearchFacesByImage API operations instead of the CompareFaces API operation.

C. Switch to using the AWS GovCloud (US) Region for Amazon S3 to store images and for Amazon Rekognition to compare faces.
Set up a VPN connection and only call the Amazon Rekognition API operations through the VPN.

D. Enable client-side encryption on the S3 bucket.
Set up a VPN connection and only call the Amazon Rekognition API operations through the VPN.

Answer: A
Explanation:
A Images passed to Amazon Rekognition API operations may be stored and used to improve the service unless you unless you have opted out by visiting the AI services opt-out policy page and following the process explained there

https://docs.aws.amazon.com/rekognition/latest/dg/security-data-encryption.html

QUESTION 10
A company is using Amazon Textract to extract textual data from thousands of scanned text-heavy legal documents daily. The company uses this information to process loan applications automatically. Some of the documents fail business validation and are returned to human reviewers, who investigate the errors. This activity increases the time to process the loan applications.

What should the company do to reduce the processing time of loan applications?

A. Configure Amazon Textract to route low-confidence predictions to Amazon SageMaker Ground Truth.
Perform a manual review on those words before performing a business validation.
B. Use an Amazon Textract synchronous operation instead of an asynchronous operation.
C. Configure Amazon Textract to route low-confidence predictions to Amazon Augmented AI (Amazon A2I).
Perform a manual review on those words before performing a business validation.
D. Use Amazon Rekognition's feature to detect text in an image to extract the data from scanned images.
Use this information to process the loan applications.

Answer: C
Explanation:
Loan or mortgage applications, tax forms, and many other financial documents contain millions of data points which need to be processed and extracted quickly and effectively. Using Amazon Textract and Amazon A2I you can extract critical data from these forms.
https://aws.amazon.com/augmented-ai/
https://aws.amazon.com/blogs/machine-learning/automated-monitoring-of-your-machine-learning-models-with-amazon-sagemaker-model-monitor-and-sending-predictions-to-human-review-workflows-using-amazon-a2i/

QUESTION 11
A data scientist has developed a machine learning translation model for English to Japanese by using Amazon SageMaker's built-in seq2seq algorithm with 500,000 aligned sentence pairs. While testing with sample sentences, the data scientist finds that the translation quality is reasonable for an example as short as five words. However, the quality becomes unacceptable if the sentence is 100 words long.

Which action will resolve the problem?

A. Change preprocessing to use n-grams.
B. Add more nodes to the recurrent neural network (RNN) than the largest sentence's word count.
C. Adjust hyperparameters related to the attention mechanism.
D. Choose a different weight initialization type.

Answer: C
Explanation:
Attention mechanism. The disadvantage of an encoder-decoder framework is that model performance decreases as and when the length of the source sequence increases because of the limit of how much information the fixed-length encoded feature vector can contain. To tackle this problem, in 2015, Bahdanau et al. proposed the attention mechanism. In an attention mechanism, the decoder tries to find the location in the encoder sequence where the most important

information could be located and uses that information and previously decoded words to predict the next token in the sequence.
https://docs.aws.amazon.com/sagemaker/latest/dg/seq-2-seq-howitworks.html

QUESTION 12
A data scientist is developing a pipeline to ingest streaming web traffic data. The data scientist needs to implement a process to identify unusual web traffic patterns as part of the pipeline. The patterns will be used downstream for alerting and incident response. The data scientist has access to unlabeled historic data to use, if needed.

The solution needs to do the following:

- Calculate an anomaly score for each web traffic entry.
- Adapt unusual event identification to changing web patterns over time.

Which approach should the data scientist implement to meet these requirements?

A. Use historic web traffic data to train an anomaly detection model using the Amazon SageMaker Random Cut Forest (RCF) built-in model.
Use an Amazon Kinesis Data Stream to process the incoming web traffic data. Attach a preprocessing AWS Lambda function to perform data enrichment by calling the RCF model to calculate the anomaly score for each record.
B. Use historic web traffic data to train an anomaly detection model using the Amazon SageMaker built-in XGBoost model.
Use an Amazon Kinesis Data Stream to process the incoming web traffic data. Attach a preprocessing AWS Lambda function to perform data enrichment by calling the XGBoost model to calculate the anomaly score for each record.
C. Collect the streaming data using Amazon Kinesis Data Firehose. Map the delivery stream as an input source for Amazon Kinesis Data Analytics.
Write a SQL query to run in real time against the streaming data with the k-Nearest Neighbors (kNN) SQL extension to calculate anomaly scores for each record using a tumbling window.
D. Collect the streaming data using Amazon Kinesis Data Firehose. Map the delivery stream as an input source for Amazon Kinesis Data Analytics.
Write a SQL query to run in real time against the streaming data with the Amazon Random Cut Forest (RCF) SQL extension to calculate anomaly scores for each record using a sliding window.

Answer: D
Explanation:
The algorithm starts developing the machine learning model using current records in the stream when you start the application. The algorithm does not use older records in the stream for machine learning, nor does it use statistics from previous executions of the application.
https://docs.aws.amazon.com/kinesisanalytics/latest/sqlref/sqlrf-random-cut-forest.html

QUESTION 13
A technology startup is using complex deep neural networks and GPU compute to recommend the company's products to its existing customers based upon each customer's habits and interactions. The solution currently pulls each dataset from an Amazon S3 bucket before loading the data into a TensorFlow model pulled from the company's Git repository that runs locally. This job then runs for several hours while continually outputting its progress to the same S3 bucket. The job can be paused, restarted, and continued at any time in the event of a failure, and is run from a central queue.

Senior managers are concerned about the complexity of the solution's resource management and the costs involved in repeating the process regularly. They ask for the workload to the automated so it runs once a week, starting Monday and completing by the close of business Friday.

Which architecture should be used to scale the solution at the lowest cost?

A. Implement the solution using AWS Deep Learning Containers and run the container as a job using AWS Batch on a GPU-compatible Spot Instance
B. Implement the solution using a low-cost GPU-compatible Amazon EC2 instance and use the AWS Instance Scheduler to schedule the task
C. Implement the solution using AWS Deep Learning Containers, run the workload using AWS Fargate running on Spot Instances, and then schedule the task using the built-in task scheduler
D. Implement the solution using Amazon ECS running on Spot Instances and schedule the task using the ECS service scheduler.

Answer: A
Explanation:
You can set up compute environments that use a particular type of EC2 instance, a particular model such as c5.2xlarge or m5.10xlarge, or simply specify that you want to use the newest instance types. You can also specify the minimum, desired, and maximum number of vCPUs for the environment, along with the amount you are willing to pay for a Spot Instance as a percentage of the On-Demand Instance price and a target set of VPC subnets. AWS Batch will efficiently launch, manage, and terminate compute types as needed. You can also manage your own compute environments. In this case you are responsible for setting up and scaling the instances in an Amazon ECS cluster that AWS Batch creates for you.
https://docs.aws.amazon.com/batch/latest/userguide/what-is-batch.html

QUESTION 14
A Machine Learning Specialist wants to bring a custom algorithm to Amazon SageMaker. The Specialist implements the algorithm in a Docker container supported by Amazon SageMaker.

How should the Specialist package the Docker container so that Amazon SageMaker can launch the training correctly?

A. Modify the bash_profile file in the container and add a bash command to start the training program
B. Use CMD config in the Dockerfile to add the training program as a CMD of the image
C. Configure the training program as an ENTRYPOINT named train
D. Copy the training program to directory /opt/ml/train

Answer: C
Explanation:
To configure a Docker container to run as an executable, use an ENTRYPOINT instruction in a Dockerfile.
SageMaker overrides any default CMD statement in a container by specifying the train argument after the image name.
https://docs.aws.amazon.com/sagemaker/latest/dg/your-algorithms-training-algo-dockerfile.html

QUESTION 15
A Machine Learning team runs its own training algorithm on Amazon SageMaker. The training algorithm requires external assets. The team needs to submit both its own algorithm code and algorithm-specific parameters to Amazon SageMaker.

What combination of services should the team use to build a custom algorithm in Amazon

SageMaker? (Choose two.)

A. AWS Secrets Manager
B. AWS CodeStar
C. Amazon ECR
D. Amazon ECS
E. Amazon S3

Answer: CE
Explanation:
For Location of inference image, type the path to the image that contains your inference code. The image must be stored as a Docker container in Amazon ECR.
For Location of model data artifacts, type the location in S3 where your model artifacts are stored.
https://docs.aws.amazon.com/sagemaker/latest/dg/sagemaker-mkt-create-model-package.html

QUESTION 16
A Machine Learning Specialist at a company sensitive to security is preparing a dataset for model training. The dataset is stored in Amazon S3 and contains Personally Identifiable Information (PII).

The dataset:

```
- Must be accessible from a VPC only.
- Must not traverse the public internet.
```

How can these requirements be satisfied?

A. Create a VPC endpoint and apply a bucket access policy that restricts access to the given VPC endpoint and the VPC.
B. Create a VPC endpoint and apply a bucket access policy that allows access from the given VPC endpoint and an Amazon EC2 instance.
C. Create a VPC endpoint and use Network Access Control Lists (NACLs) to allow traffic between only the given VPC endpoint and an Amazon EC2 instance.
D. Create a VPC endpoint and use security groups to restrict access to the given VPC endpoint and an Amazon EC2 instance

Answer: A
Explanation:
You can control which VPCs or VPC endpoints have access to your buckets by using Amazon S3 bucket policies. For examples of this type of bucket policy access control, see the following topics on restricting access.
https://docs.aws.amazon.com/AmazonS3/latest/dev/example-bucket-policies-vpc-endpoint.html

QUESTION 17
A Machine Learning Specialist is preparing data for training on Amazon SageMaker. The Specialist is using one of the SageMaker built-in algorithms for the training. The dataset is stored in .CSV format and is transformed into a numpy.array, which appears to be negatively affecting the speed of the training.

What should the Specialist do to optimize the data for training on SageMaker?

A. Use the SageMaker batch transform feature to transform the training data into a DataFrame.

B. Use AWS Glue to compress the data into the Apache Parquet format.
C. Transform the dataset into the RecordIO protobuf format.
D. Use the SageMaker hyperparameter optimization feature to automatically optimize the data.

Answer: C
Explanation:
Most Amazon SageMaker algorithms work best when you use the optimized protobuf recordIO format for the training data.
https://docs.aws.amazon.com/sagemaker/latest/dg/cdf-training.html

QUESTION 18
A large consumer goods manufacturer has the following products on sale:

```
- 34 different toothpaste variants
- 48 different toothbrush variants
- 43 different mouthwash variants
```

The entire sales history of all these products is available in Amazon S3. Currently, the company is using custom-built autoregressive integrated moving average (ARIMA) models to forecast demand for these products. The company wants to predict the demand for a new product that will soon be launched.

Which solution should a Machine Learning Specialist apply?

A. Train a custom ARIMA model to forecast demand for the new product.
B. Train an Amazon SageMaker DeepAR algorithm to forecast demand for the new product.
C. Train an Amazon SageMaker k-means clustering algorithm to forecast demand for the new product.
D. Train a custom XGBoost model to forecast demand for the new product.

Answer: B
Explanation:
The Amazon SageMaker DeepAR forecasting algorithm is a supervised learning algorithm for forecasting scalar (one-dimensional) time series using recurrent neural networks (RNN). Classical forecasting methods, such as autoregressive integrated moving average (ARIMA) or exponential smoothing (ETS), fit a single model to each individual time series. They then use that model to extrapolate the time series into the future.
https://docs.aws.amazon.com/sagemaker/latest/dg/deepar.html

QUESTION 19
A Machine Learning Specialist uploads a dataset to an Amazon S3 bucket protected with server-side encryption using AWS KMS.

How should the ML Specialist define the Amazon SageMaker notebook instance so it can read the same dataset from Amazon S3?

A. Define security group(s) to allow all HTTP inbound/outbound traffic and assign those security group(s) to the Amazon SageMaker notebook instance.
B. onfigure the Amazon SageMaker notebook instance to have access to the VPC. Grant permission in the KMS key policy to the notebook's KMS role.
C. Assign an IAM role to the Amazon SageMaker notebook with S3 read access to the dataset. Grant permission in the KMS key policy to that role.

D. Assign the same KMS key used to encrypt data in Amazon S3 to the Amazon SageMaker notebook instance.

Answer: C
Explanation:
You don’t need to specify the AWS KMS key ID when you download an SSE-KMS-encrypted object from an S3 bucket. Instead, you need the permission to decrypt the AWS KMS key. When a user sends a GET request, Amazon S3 checks if the AWS Identity and Access Management (IAM) user or role that sent the request is authorized to decrypt the key associated with the object. If the IAM user or role belongs to the same AWS account as the key, then the permission to decrypt must be granted on the AWS KMS key’s policy.
https://aws.amazon.com/premiumsupport/knowledge-center/decrypt-kms-encrypted-objects-s3/?nc1=h_ls

QUESTION 20
A Machine Learning Specialist must build out a process to query a dataset on Amazon S3 using Amazon Athena. The dataset contains more than 800,000 records stored as plaintext CSV files. Each record contains 200 columns and is approximately 1.5 MB in size. Most queries will span 5 to 10 columns only.

How should the Machine Learning Specialist transform the dataset to minimize query runtime?

A. Convert the records to Apache Parquet format.
B. Convert the records to JSON format.
C. Convert the records to GZIP CSV format.
D. Convert the records to XML format.

Answer: A
Explanation:
Using compressions will reduce the amount of data scanned by Amazon Athena, and also reduce your S3 bucket storage. It's a Win-Win for your AWS bill. Supported formats: GZIP, LZO, SNAPPY (Parquet) and ZLIB.
https://www.cloudforecast.io/blog/using-parquet-on-athena-to-save-money-on-aws/