Amazon

Exam Questions AWS-Certified-Machine-Learning-Specialty

AWS Certified Machine Learning - Specialty

NEW QUESTION 1
A Data Scientist needs to create a serverless ingestion and analytics solution for high-velocity, real-time streaming data.
The ingestion process must buffer and convert incoming records from JSON to a query-optimized, columnar format without data loss. The output datastore must be highly available, and Analysts must be able to run SQL queries against the data and connect to existing business intelligence dashboards.
Which solution should the Data Scientist build to satisfy the requirements?

A. Create a schema in the AWS Glue Data Catalog of the incoming data forma
B. Use an Amazon Kinesis Data Firehose delivery stream to stream the data and transform the data to Apache Parquet or ORC format using the AWS Glue Data Catalog before delivering to Amazon S3. Have the Analysts query the data directly from Amazon S3 using Amazon Athena, and connect to BI tools using the Athena Java Database Connectivity (JDBC) connector.
C. Write each JSON record to a staging location in Amazon S3. Use the S3 Put event to trigger an AWS Lambda function that transforms the data into Apache Parquet or ORC format and writes the data to a processed data location in Amazon S3. Have the Analysts query the data directly from Amazon S3 using Amazon Athena, and connect to BI tools using the Athena Java Database Connectivity (JDBC) connector.
D. Write each JSON record to a staging location in Amazon S3. Use the S3 Put event to trigger an AWS Lambda function that transforms the data into Apache Parquet or ORC format and inserts it into an Amazon RDS PostgreSQL databas
E. Have the Analysts query and run dashboards from the RDS database.
F. Use Amazon Kinesis Data Analytics to ingest the streaming data and perform real-time SQL queries to convert the records to Apache Parquet before delivering to Amazon S3. Have the Analysts query the data directly from Amazon S3 using Amazon Athena and connect to BI tools using the Athena Java Database Connectivity (JDBC) connector.

Answer: A

NEW QUESTION 2
A Machine Learning Specialist observes several performance problems with the training portion of a machine learning solution on Amazon SageMaker The solution uses a large training dataset 2 TB in size and is using the SageMaker k-means algorithm The observed issues include the unacceptable length of time it takes before the training job launches and poor I/O throughput while training the model
What should the Specialist do to address the performance issues with the current solution?

A. Use the SageMaker batch transform feature
B. Compress the training data into Apache Parquet format.
C. Ensure that the input mode for the training job is set to Pipe.
D. Copy the training dataset to an Amazon EFS volume mounted on the SageMaker instance.

Answer: B

NEW QUESTION 3
A Machine Learning Specialist works for a credit card processing company and needs to predict which transactions may be fraudulent in near-real time.
Specifically, the Specialist must train a model that returns the probability that a given transaction may be fraudulent
How should the Specialist frame this business problem'?

A. Streaming classification
B. Binary classification
C. Multi-category classification
D. Regression classification

Answer: A

NEW QUESTION 4
A retail company wants to update its customer support system. The company wants to implement automatic routing of customer claims to different queues to prioritize the claims by category.
Currently, an operator manually performs the category assignment and routing. After the operator classifies and routes the claim, the company stores the claim's record in a central database. The claim's record includes the claim's category.
The company has no data science team or experience in the field of machine learning (ML). The company's small development team needs a solution that requires no ML expertise.
Which solution meets these requirements?

A. Export the database to a .csv file with two columns: claim_label and claim_tex
B. Use the Amazon SageMaker Object2Vec algorithm and the .csv file to train a mode
C. Use SageMaker to deploy the model to an inference endpoin
D. Develop a service in the application to use the inference endpoint to process incoming claims, predict the labels, and route the claims to the appropriate queue.
E. Export the database to a .csv file with one column: claim_tex
F. Use the Amazon SageMaker Latent Dirichlet Allocation (LDA) algorithm and the .csv file to train a mode
G. Use the LDA algorithm to detect labels automaticall
H. Use SageMaker to deploy the model to an inference endpoin
I. Develop a service in the application to use the inference endpoint to process incoming claims, predict the labels, and route the claims to the appropriate queue.
J. Use Amazon Textract to process the database and automatically detect two columns: claim_label and claim_tex
K. Use Amazon Comprehend custom classification and the extracted information to train the custom classifie
L. Develop a service in the application to use the Amazon Comprehend API to process incoming claims, predict the labels, and route the claims to the appropriate queue.
M. Export the database to a .csv file with two columns: claim_label and claim_tex
N. Use Amazon Comprehend custom classification and the .csv file to train the custom classifie
O. Develop a service in the application to use the Amazon Comprehend API to process incoming claims, predict the labels, and route the claims to the appropriate queue.

Answer: C

NEW QUESTION 5

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 syste
B. Run the SageMaker training job with the FSx for Lustre file system as the data source.
C. Launch a transient Amazon EMR cluste
D. Configure steps to mount the EFS file system and copy the data to an Amazon S3 bucket by using S3DistC
E. Run the SageMaker training job with Amazon S3 as the data source.
F. Mount the EFS file system to an Amazon EC2 instance and use the AWS CLI to copy the data to an Amazon S3 bucke
G. Run the SageMaker training job with Amazon S3 as the data source.
H. Run a SageMaker training job with an EFS file system as the data source.

Answer: A

NEW QUESTION 6
An agency collects census information within a country to determine healthcare and social program needs by province and city. The census form collects responses for approximately 500 questions from each citizen
Which combination of algorithms would provide the appropriate insights? (Select TWO)

A. The factorization machines (FM) algorithm
B. The Latent Dirichlet Allocation (LDA) algorithm
C. The principal component analysis (PCA) algorithm
D. The k-means algorithm
E. The Random Cut Forest (RCF) algorithm

Answer: CD

Explanation:
The PCA and K-means algorithms are useful in collection of data using census form.

NEW QUESTION 7
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: C:\Users\Admin\Desktop\Data\Odt data\Untitled.jpgSelect the most accurate hyperparameter configuration form this HPO job.

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

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: C:\Users\Admin\Desktop\Data\Odt data\Untitled.jpg

```
{
     "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 MaxValu
E. 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

NEW QUESTION 8
A Machine Learning Specialist was given a dataset consisting of unlabeled data The Specialist must create a model that can help the team classify the data into

different buckets What model should be used to complete this work?

A. K-means clustering
B. Random Cut Forest (RCF)
C. XGBoost
D. BlazingText

Answer: A

NEW QUESTION 9
A company will use Amazon SageMaker to train and host a machine learning (ML) model for a marketing campaign. The majority of data is sensitive customer data. The data must be encrypted at rest. The company wants AWS to maintain the root of trust for the master keys and wants encryption key usage to be logged. Which implementation will meet these requirements?

A. Use encryption keys that are stored in AWS Cloud HSM to encrypt the ML data volumes, and to encryptthe model artifacts and data in Amazon S3.
B. Use SageMaker built-in transient keys to encrypt the ML data volume
C. Enable default encryption for new Amazon Elastic Block Store (Amazon EBS) volumes.
D. Use customer managed keys in AWS Key Management Service (AWS KMS) to encrypt the ML data volumes, and to encrypt the model artifacts and data in Amazon S3.
E. Use AWS Security Token Service (AWS STS) to create temporary tokens to encrypt the ML storage volumes, and to encrypt the model artifacts and data in Amazon S3.

Answer: C

NEW QUESTION 10
A manufacturer of car engines collects data from cars as they are being driven The data collected includes timestamp, engine temperature, rotations per minute (RPM), and other sensor readings The company wants to predict when an engine is going to have a problem so it can notify drivers in advance to get engine maintenance The engine data is loaded into a data lake for training
Which is the MOST suitable predictive model that can be deployed into production'?

A. Add labels over time to indicate which engine faults occur at what time in the future to turn this into a supervised learning problem Use a recurrent neural network (RNN) to train the model to recognize when an engine might need maintenance for a certain fault.
B. This data requires an unsupervised learning algorithm Use Amazon SageMaker k-means to cluster the data
C. Add labels over time to indicate which engine faults occur at what time in the future to turn this into a supervised learning problem Use a convolutional neural network (CNN) to train the model to recognize when an engine might need maintenance for a certain fault.
D. This data is already formulated as a time series Use Amazon SageMaker seq2seq to model the time series.

Answer: B

NEW QUESTION 10
A bank wants to launch a low-rate credit promotion. The bank is located in a town that recently experienced economic hardship. Only some of the bank's customers were affected by the crisis, so the bank's credit team must identify which customers to target with the promotion. However, the credit team wants to make sure that loyal customers' full credit history is considered when the decision is made.
The bank's data science team developed a model that classifies account transactions and understands credit eligibility. The data science team used the XGBoost algorithm to train the model. The team used 7 years of bank transaction historical data for training and hyperparameter tuning over the course of several days.
The accuracy of the model is sufficient, but the credit team is struggling to explain accurately why the model denies credit to some customers. The credit team has almost no skill in data science.
What should the data science team do to address this issue in the MOST operationally efficient manner?

A. Use Amazon SageMaker Studio to rebuild the mode
B. Create a notebook that uses the XGBoost training container to perform model trainin
C. Deploy the model at an endpoin
D. Enable Amazon SageMaker Model Monitor to store inference
E. Use the inferences to create Shapley values that help explain model behavio
F. Create a chart that shows features and SHapley Additive explanation (SHAP) values to explain to the credit team how the features affect the model outcomes.
G. Use Amazon SageMaker Studio to rebuild the mode
H. Create a notebook that uses the XGBoost training container to perform model trainin
I. Activate Amazon SageMaker Debugger, and configure it to calculate and collect Shapley value
J. Create a chart that shows features and SHapley Additive explanation (SHAP) values to explain to the credit team how the features affect the model outcomes.
K. Create an Amazon SageMaker notebook instanc
L. Use the notebook instance and the XGBoost library to locally retrain the mode
M. Use the plot_importance() method in the Python XGBoost interface to create a feature importance char
N. Use that chart to explain to the credit team how the features affect the model outcomes.
O. Use Amazon SageMaker Studio to rebuild the mode
P. Create a notebook that uses the XGBoost training container to perform model trainin
Q. Deploy the model at an endpoin
R. Use Amazon SageMakerProcessing to post-analyze the model and create a feature importance explainability chart automatically for the credit team.

Answer: C

NEW QUESTION 13
An e-commerce company needs a customized training model to classify images of its shirts and pants products The company needs a proof of concept in 2 to 3 days with good accuracy Which compute choice should the Machine Learning Specialist select to train and achieve good accuracy on the model quickly?

A. m5 4xlarge (general purpose)
B. r5.2xlarge (memory optimized)
C. p3.2xlarge (GPU accelerated computing)
D. p3 8xlarge (GPU accelerated computing)

Answer: C

NEW QUESTION 15
An online reseller has a large, multi-column dataset with one column missing 30% of its data A Machine Learning Specialist believes that certain columns in the dataset could be used to reconstruct the missing data.
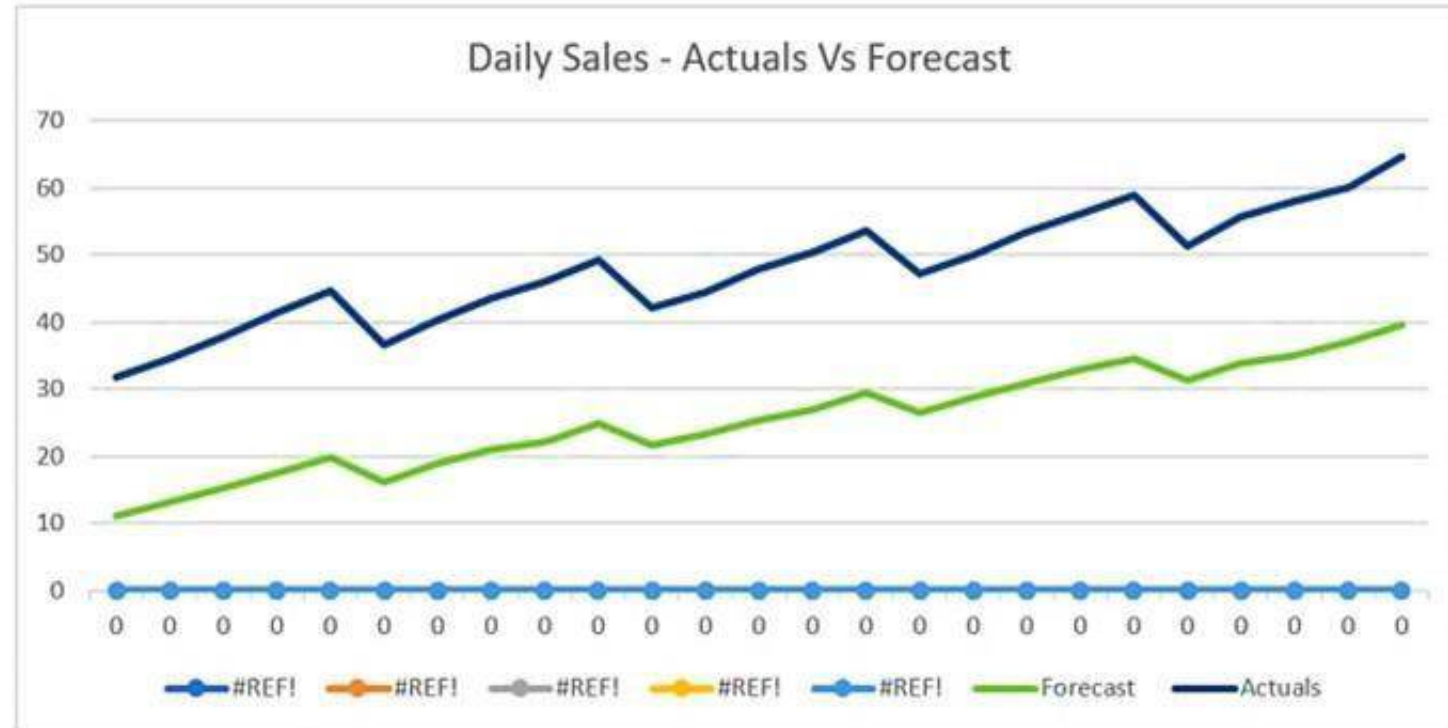Which reconstruction approach should the Specialist use to preserve the integrity of the dataset?

A. Listwise deletion
B. Last observation carried forward
C. Multiple imputation
D. Mean substitution

Answer: C

NEW QUESTION 19
The displayed graph is from a foresting model for testing a time series.

Considering the graph only, which conclusion should a Machine Learning Specialist make about the behavior of the model?

A. The model predicts both the trend and the seasonality well.
B. The model predicts the trend well, but not the seasonality.
C. The model predicts the seasonality well, but not the trend.
D. The model does not predict the trend or the seasonality well.

Answer: D

NEW QUESTION 22
A financial services company wants to adopt Amazon SageMaker as its default data science environment. The company's data scientists run machine learning (ML) models on confidential financial data. The company is worried about data egress and wants an ML engineer to secure the environment.
Which mechanisms can the ML engineer use to control data egress from SageMaker? (Choose three.)

A. Connect to SageMaker by using a VPC interface endpoint powered by AWS PrivateLink.
B. Use SCPs to restrict access to SageMaker.
C. Disable root access on the SageMaker notebook instances.
D. Enable network isolation for training jobs and models.
E. Restrict notebook presigned URLs to specific IPs used by the company.
F. Protect data with encryption at rest and in transi
G. Use AWS Key Management Service (AWS KMS) to manage encryption keys.

Answer: BDE

Explanation:
https://aws.amazon.com/blogs/machine-learning/millennium-management-secure-machine-learning-using-amaz

NEW QUESTION 26
A Machine Learning Specialist needs to move and transform data in preparation for training Some of the data needs to be processed in near-real time and other data can be moved hourly There are existing Amazon EMR MapReduce jobs to clean and feature engineering to perform on the data
Which of the following services can feed data to the MapReduce jobs? (Select TWO)

A. AWSDMS
B. Amazon Kinesis
C. AWS Data Pipeline
D. Amazon Athena
E. Amazon ES

Answer: BC

Explanation:

https://aws.amazon.com/jp/emr/?whats-new-cards.sort-by=item.additionalFields.postDateTime&whats-new-car

NEW QUESTION 30
A Data Scientist received a set of insurance records, each consisting of a record ID, the final outcome among 200 categories, and the date of the final outcome. Some partial information on claim contents is also provided, but only for a few of the 200 categories. For each outcome category, there are hundreds of records distributed over the past 3 years. The Data Scientist wants to predict how many claims to expect in each category from month to month, a few months in advance.
What type of machine learning model should be used?

A. Classification month-to-month using supervised learning of the 200 categories based on claim contents.
B. Reinforcement learning using claim IDs and timestamps where the agent will identify how many claims ineach category to expect from month to month.
C. Forecasting using claim IDs and timestamps to identify how many claims in each category to expect frommonth to month.
D. Classification with supervised learning of the categories for which partial information on claim contents isprovided, and forecasting using claim IDs and timestamps for all other categories.

Answer: C

NEW QUESTION 34
A Data Scientist is working on an application that performs sentiment analysis. The validation accuracy is poor and the Data Scientist thinks that the cause may be a rich vocabulary and a low average frequency of words in the dataset
Which tool should be used to improve the validation accuracy?

A. Amazon Comprehend syntax analysts and entity detection
B. Amazon SageMaker BlazingText allow mode
C. Natural Language Toolkit (NLTK) stemming and stop word removal
D. Scikit-learn term frequency-inverse document frequency (TF-IDF) vectorizers

Answer: A

NEW QUESTION 38
A Machine Learning Specialist is configuring Amazon SageMaker so multiple Data Scientists can access notebooks, train models, and deploy endpoints. To ensure the best operational performance, the Specialist needs to be able to track how often the Scientists are deploying models, GPU and CPU utilization on the deployed SageMaker endpoints, and all errors that are generated when an endpoint is invoked.
Which services are integrated with Amazon SageMaker to track this information? (Select TWO.)

A. AWS CloudTrail
B. AWS Health
C. AWS Trusted Advisor
D. Amazon CloudWatch
E. AWS Config

Answer: AD

NEW QUESTION 41
A global financial company is using machine learning to automate its loan approval process. The company has a dataset of customer information. The dataset contains some categorical fields, such as customer location by city and housing status. The dataset also includes financial fields in different units, such as account balances in US dollars and monthly interest in US cents.
The company's data scientists are using a gradient boosting regression model to infer the credit score for each customer. The model has a training accuracy of 99% and a testing accuracy of 75%. The data scientists want to improve the model's testing accuracy.
Which process will improve the testing accuracy the MOST?

A. Use a one-hot encoder for the categorical fields in the datase
B. Perform standardization on the financial fields in the datase
C. Apply L1 regularization to the data.
D. Use tokenization of the categorical fields in the datase
E. Perform binning on the financial fields in the datase
F. Remove the outliers in the data by using the z-score.
G. Use a label encoder for the categorical fields in the datase
H. Perform L1 regularization on the financial fields in the datase
I. Apply L2 regularization to the data.
J. Use a logarithm transformation on the categorical fields in the datase
K. Perform binning on the financial fields in the datase
L. Use imputation to populate missing values in the dataset.

Answer: B

NEW QUESTION 42
A Machine Learning Specialist is building a convolutional neural network (CNN) that will classify 10 types of animals. The Specialist has built a series of layers in a neural network that will take an input image of an animal, pass it through a series of convolutional and pooling layers, and then finally pass it through a dense and fully connected layer with 10 nodes The Specialist would like to get an output from the neural network that is a probability distribution of how likely it is that the input image belongs to each of the 10 classes
Which function will produce the desired output?

A. Dropout
B. Smooth L1 loss
C. Softmax
D. Rectified linear units (ReLU)

Answer: C

NEW QUESTION 44
A Machine Learning Specialist is developing a daily ETL workflow containing multiple ETL jobs The workflow consists of the following processes
* Start the workflow as soon as data is uploaded to Amazon S3
* When all the datasets are available in Amazon S3, start an ETL job to join the uploaded datasets with multiple terabyte-sized datasets already stored in Amazon S3
* Store the results of joining datasets in Amazon S3
* If one of the jobs fails, send a notification to the Administrator Which configuration will meet these requirements?

A. Use AWS Lambda to trigger an AWS Step Functions workflow to wait for dataset uploads to complete in Amazon S3. Use AWS Glue to join the datasets Use an Amazon CloudWatch alarm to send an SNS notification to the Administrator in the case of a failure
B. Develop the ETL workflow using AWS Lambda to start an Amazon SageMaker notebook instance Use a lifecycle configuration script to join the datasets and persist the results in Amazon S3 Use an Amazon CloudWatch alarm to send an SNS notification to the Administrator in the case of a failure
C. Develop the ETL workflow using AWS Batch to trigger the start of ETL jobs when data is uploaded to Amazon S3 Use AWS Glue to join the datasets in Amazon S3 Use an Amazon CloudWatch alarm to send an SNS notification to the Administrator in the case of a failure
D. Use AWS Lambda to chain other Lambda functions to read and join the datasets in Amazon S3 as soon as the data is uploaded to Amazon S3 Use an Amazon CloudWatch alarm to send an SNS notification to the Administrator in the case of a failure

Answer: A

NEW QUESTION 47
A Machine Learning Specialist is training a model to identify the make and model of vehicles in images The Specialist wants to use transfer learning and an existing model trained on images of general objects The Specialist collated a large custom dataset of pictures containing different vehicle makes and models

A. Initialize the model with random weights in all layers including the last fully connected layer
B. Initialize the model with pre-trained weights in all layers and replace the last fully connected layer.
C. Initialize the model with random weights in all layers and replace the last fully connected layer
D. Initialize the model with pre-trained weights in all layers including the last fully connected layer

Answer: D

NEW QUESTION 52
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 dat
B. Use two arrays to compare equal strings in the fields from the two datasets and remove any duplicates.
C. Create AWS Glue crawlers for reading and populating the AWS Glue Data Catalo
D. Call the AWS Glue SearchTables API operation to perform a fuzzy-matching search on the two datasets, and cleanse the data accordingly.
E. Create AWS Glue crawlers for reading and populating the AWS Glue Data Catalo
F. Use the FindMatches transform to cleanse the data.
G. Create an AWS Lake Formation custom transfor
H. Run a transformation for matching products from the Lake Formation console to cleanse the data automatically.

Answer: D

NEW QUESTION 55
A machine learning (ML) specialist wants to create a data preparation job that uses a PySpark script with complex window aggregation operations to create data for training and testing. The ML specialist needs to evaluate the impact of the number of features and the sample count on model performance.
Which approach should the ML specialist use to determine the ideal data transformations for the model?

A. Add an Amazon SageMaker Debugger hook to the script to capture key metric
B. Run the script as an AWS Glue job.
C. Add an Amazon SageMaker Experiments tracker to the script to capture key metric
D. Run the script as an AWS Glue job.
E. Add an Amazon SageMaker Debugger hook to the script to capture key parameter
F. Run the script as a SageMaker processing job.
G. Add an Amazon SageMaker Experiments tracker to the script to capture key parameter
H. Run the script as a SageMaker processing job.

Answer: B

NEW QUESTION 56
Amazon Connect has recently been tolled out across a company as a contact call center The solution has been configured to store voice call recordings on Amazon S3
The content of the voice calls are being analyzed for the incidents being discussed by the call operators Amazon Transcribe is being used to convert the audio to text, and the output is stored on Amazon S3
Which approach will provide the information required for further analysis?

A. Use Amazon Comprehend with the transcribed files to build the key topics
B. Use Amazon Translate with the transcribed files to train and build a model for the key topics
C. Use the AWS Deep Learning AMI with Gluon Semantic Segmentation on the transcribed files to train and build a model for the key topics
D. Use the Amazon SageMaker k-Nearest-Neighbors (kNN) algorithm on the transcribed files to generate a word embeddings dictionary for the key topics
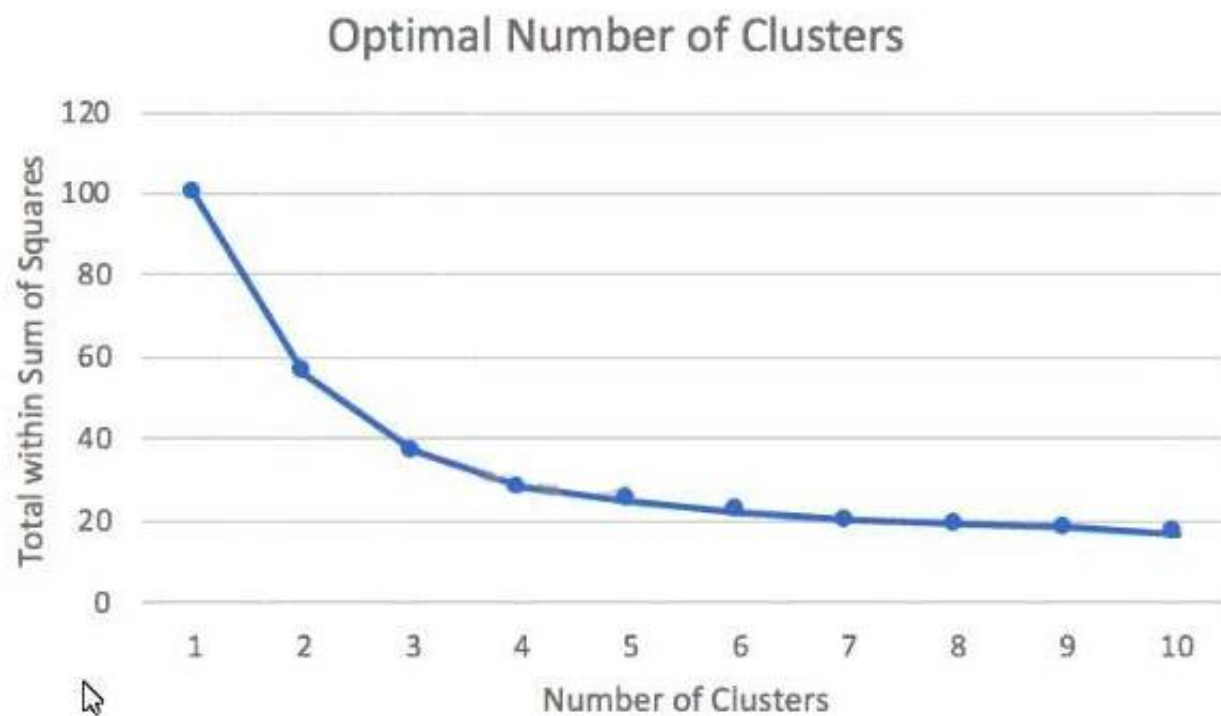
Answer: B

NEW QUESTION 57
A Data Scientist wants to gain real-time insights into a data stream of GZIP files. Which solution would allow the use of SQL to query the stream with the LEAST latency?

A. Amazon Kinesis Data Analytics with an AWS Lambda function to transform the data.
B. AWS Glue with a custom ETL script to transform the data.
C. An Amazon Kinesis Client Library to transform the data and save it to an Amazon ES cluster.
D. Amazon Kinesis Data Firehose to transform the data and put it into an Amazon S3 bucket.

Answer: A

NEW QUESTION 61
A Machine Learning Specialist prepared the following graph displaying the results of k-means for k = [1:10]

Considering the graph, what is a reasonable selection for the optimal choice of k?

A. 1
B. 4
C. 7
D. 10

Answer: C

NEW QUESTION 66
A company is building a demand forecasting model based on machine learning (ML). In the development stage, an ML specialist uses an Amazon SageMaker notebook to perform feature engineering during work hours that consumes low amounts of CPU and memory resources. A data engineer uses the same notebook to perform data preprocessing once a day on average that requires very high memory and completes in only 2 hours. The data preprocessing is not configured to use GPU. All the processes are running well on an ml.m5.4xlarge notebook instance.
The company receives an AWS Budgets alert that the billing for this month exceeds the allocated budget. Which solution will result in the MOST cost savings?

A. Change the notebook instance type to a memory optimized instance with the same vCPU number as the ml.m5.4xlarge instance ha
B. Stop the notebook when it is not in us
C. Run both data preprocessing and feature engineering development on that instance.
D. Keep the notebook instance type and size the sam
E. Stop the notebook when it is not in us
F. Run data preprocessing on a P3 instance type with the same memory as the ml.m5.4xlarge instance by using Amazon SageMaker Processing.
G. Change the notebook instance type to a smaller general purpose instanc
H. Stop the notebook when it is not in us
I. Run data preprocessing on an ml.r5 instance with the same memory size as the ml.m5.4xlarge instance by using Amazon SageMaker Processing.
J. Change the notebook instance type to a smaller general purpose instanc
K. Stop the notebook when it is not in us
L. Run data preprocessing on an R5 instance with the same memory size as the ml.m5.4xlarge instance by using the Reserved Instance option.

Answer: B

NEW QUESTION 71
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 be 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: C

NEW QUESTION 75

A company is building a line-counting application for use in a quick-service restaurant. The company wants to use video cameras pointed at the line of customers at a given register to measure how many people are in line and deliver notifications to managers if the line grows too long. The restaurant locations have limited bandwidth for connections to external services and cannot accommodate multiple video streams without impacting other operations.
Which solution should a machine learning specialist implement to meet these requirements?

A. Install cameras compatible with Amazon Kinesis Video Streams to stream the data to AWS over the restaurant's existing internet connectio
B. Write an AWS Lambda function to take an image and send it to Amazon Rekognition to count the number of faces in the imag
C. Send an Amazon Simple Notification Service (Amazon SNS) notification if the line is too long.
D. Deploy AWS DeepLens cameras in the restaurant to capture vide
E. Enable Amazon Rekognition on the AWS DeepLens device, and use it to trigger a local AWS Lambda function when a person is recognize
F. Use the Lambda function to send an Amazon Simple Notification Service (Amazon SNS) notification if the line is too long.
G. Build a custom model in Amazon SageMaker to recognize the number of people in an imag
H. Install cameras compatible with Amazon Kinesis Video Streams in the restauran
I. Write an AWS Lambda function to take an imag
J. Use the SageMaker endpoint to call the model to count peopl
K. Send an Amazon Simple Notification Service (Amazon SNS) notification if the line is too long.
L. Build a custom model in Amazon SageMaker to recognize the number of people in an imag
M. Deploy AWS DeepLens cameras in the restauran
N. Deploy the model to the camera
O. Deploy an AWS Lambda function to the cameras to use the model to count people and send an Amazon Simple Notification Service (Amazon SNS) notification if the line is too long.

Answer: A

NEW QUESTION 77

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 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 bucke
B. Determine which source files do not exist in the destination S3 bucket and which source files were modified.
C. Use AWS Transfer for FTPS to transfer the files from the on-premises storage to Amazon S3.
D. Use AWS DataSync to make an initial copy of the entire datase
E. Schedule subsequent incremental transfers of changing data until the final cutover from on premises to AWS.
F. Use S3 Batch Operations to pull data periodically from the on-premises storag
G. 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.

NEW QUESTION 80

A data scientist is working on a public sector project for an urban traffic system. While studying the traffic patterns, it is clear to the data scientist that the traffic behavior at each light is correlated, subject to a small stochastic error term. The data scientist must model the traffic behavior to analyze the traffic patterns and reduce congestion.
How will the data scientist MOST effectively model the problem?

A. The data scientist should obtain a correlated equilibrium policy by formulating this problem as a multi-agent reinforcement learning problem.
B. The data scientist should obtain the optimal equilibrium policy by formulating this problem as a single-agent reinforcement learning problem.
C. Rather than finding an equilibrium policy, the data scientist should obtain accurate predictors of traffic flow by using historical data through a supervised learning approach.
D. Rather than finding an equilibrium policy, the data scientist should obtain accurate predictors of traffic flow by using unlabeled simulated data representing the new traffic patterns in the city and applying an unsupervised learning approach.

Answer: D

NEW QUESTION 83

A Machine Learning Specialist has created a deep learning neural network model that performs well on the training data but performs poorly on the test data.
Which of the following methods should the Specialist consider using to correct this? (Select THREE.)

A. Decrease regularization.
B. Increase regularization.
C. Increase dropout.
D. Decrease dropout.
E. Increase feature combinations.
F. Decrease feature combinations.

Answer: BCD

NEW QUESTION 84

A Machine Learning Specialist is using an Amazon SageMaker notebook instance in a private subnet of a corporate VPC. The ML Specialist has important data stored on the Amazon SageMaker notebook instance's Amazon EBS volume, and needs to take a snapshot of that EBS volume. However the ML Specialist cannot find the Amazon SageMaker notebook instance's EBS volume or Amazon EC2 instance within the VPC.
Why is the ML Specialist not seeing the instance visible in the VPC?

A. Amazon SageMaker notebook instances are based on the EC2 instances within the customer account, but they run outside of VPCs.
B. Amazon SageMaker notebook instances are based on the Amazon ECS service within customer accounts.
C. Amazon SageMaker notebook instances are based on EC2 instances running within AWS serviceaccounts.
D. Amazon SageMaker notebook instances are based on AWS ECS instances running within AWS service accounts.

Answer: C

NEW QUESTION 86

A manufacturing company uses machine learning (ML) models to detect quality issues. The models use images that are taken of the company's product at the end of each production step. The company has thousands of machines at the production site that generate one image per second on average.
The company ran a successful pilot with a single manufacturing machine. For the pilot, ML specialists used an industrial PC that ran AWS IoT Greengrass with a long-running AWS Lambda function that uploaded the images to Amazon S3. The uploaded images invoked a Lambda function that was written in Python to perform inference by using an Amazon SageMaker endpoint that ran a custom model. The inference results were forwarded back to a web service that was hosted at the production site to prevent faulty products from being shipped.
The company scaled the solution out to all manufacturing machines by installing similarly configured industrial PCs on each production machine. However, latency for predictions increased beyond acceptable limits. Analysis shows that the internet connection is at its capacity limit.
How can the company resolve this issue MOST cost-effectively?

A. Set up a 10 Gbps AWS Direct Connect connection between the production site and the nearest AWS Regio
B. Use the Direct Connect connection to upload the image
C. Increase the size of the instances and the number of instances that are used by the SageMaker endpoint.
D. Extend the long-running Lambda function that runs on AWS IoT Greengrass to compress the images and upload the compressed files to Amazon S3.
Decompress the files by using a separate Lambda function that invokes the existing Lambda function to run the inference pipeline.
E. Use auto scaling for SageMake
F. Set up an AWS Direct Connect connection between the production site and the nearest AWS Regio
G. Use the Direct Connect connection to upload the images.
H. Deploy the Lambda function and the ML models onto the AWS IoT Greengrass core that is running on the industrial PCs that are installed on each machin
I. Extend the long-running Lambda function that runs on AWS IoT Greengrass to invoke the Lambda function with the captured images and run the inference on the edge component that forwards the results directly to the web service.

Answer: D

NEW QUESTION 91

A company uses camera images of the tops of items displayed on store shelves to determine which items were removed and which ones still remain. After several hours of data labeling, the company has a total of 1,000 hand-labeled images covering 10 distinct items. The training results were poor.
Which machine learning approach fulfills the company's long-term needs?

A. Convert the images to grayscale and retrain the model
B. Reduce the number of distinct items from 10 to 2, build the model, and iterate
C. Attach different colored labels to each item, take the images again, and build the model
D. Augment training data for each item using image variants like inversions and translations, build the model, and iterate.

Answer: A

NEW QUESTION 96

A health care company is planning to use neural networks to classify their X-ray images into normal and abnormal classes. The labeled data is divided into a training set of 1,000 images and a test set of 200 images. The initial training of a neural network model with 50 hidden layers yielded 99% accuracy on the training set, but only 55% accuracy on the test set.
What changes should the Specialist consider to solve this issue? (Choose three.)

A. Choose a higher number of layers
B. Choose a lower number of layers
C. Choose a smaller learning rate
D. Enable dropout
E. Include all the images from the test set in the training set
F. Enable early stopping

Answer: ADE

NEW QUESTION 99

A Machine Learning Specialist needs to be able to ingest streaming data and store it in Apache Parquet files for exploration and analysis. Which of the following services would both ingest and store this data in the correct format?

A. AWSDMS
B. Amazon Kinesis Data Streams
C. Amazon Kinesis Data Firehose
D. Amazon Kinesis Data Analytics

Answer: C

NEW QUESTION 102
An agricultural company is interested in using machine learning to detect specific types of weeds in a 100-acre grassland field. Currently, the company uses tractor-mounted cameras to capture multiple images of the field as 10 × 10 grids. The company also has a large training dataset that consists of annotated images of popular weed classes like broadleaf and non-broadleaf docks.
The company wants to build a weed detection model that will detect specific types of weeds and the location of each type within the field. Once the model is ready, it will be hosted on Amazon SageMaker endpoints. The model will perform real-time inferencing using the images captured by the cameras. Which approach should a Machine Learning Specialist take to obtain accurate predictions?

A. Prepare the images in RecordIO format and upload them to Amazon S3. Use Amazon SageMaker to train, test, and validate the model using an image classification algorithm to categorize images into various weed classes.
B. Prepare the images in Apache Parquet format and upload them to Amazon S3. Use Amazon SageMaker to train, test, and validate the model using an object-detection single-shot multibox detector (SSD) algorithm.
C. Prepare the images in RecordIO format and upload them to Amazon S3. Use Amazon SageMaker to train, test, and validate the model using an object-detection single-shot multibox detector (SSD) algorithm.
D. Prepare the images in Apache Parquet format and upload them to Amazon S3. Use Amazon SageMaker to train, test, and validate the model using an image classification algorithm to categorize images into various weed classes.

Answer: C

NEW QUESTION 107
A company has set up and deployed its machine learning (ML) model into production with an endpoint using Amazon SageMaker hosting services. The ML team has configured automatic scaling for its SageMaker instances to support workload changes. During testing, the team notices that additional instances are being launched before the new instances are ready. This behavior needs to change as soon as possible.
How can the ML team solve this issue?

A. Decrease the cooldown period for the scale-in activit
B. Increase the configured maximum capacity of instances.
C. Replace the current endpoint with a multi-model endpoint using SageMaker.
D. Set up Amazon API Gateway and AWS Lambda to trigger the SageMaker inference endpoint.
E. Increase the cooldown period for the scale-out activity.

Answer: A

NEW QUESTION 108
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.

NEW QUESTION 110
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 featur
B. When the new version of the model is ready for release, gradually increase InitialVariantWeight until all users have the updated version.
C. Configure two SageMaker hosted endpoints that serve the different versions of the mode
D. Create an Application Load Balancer (ALB) to route traffic to both endpoints based on the TargetVariant query string paramete
E. Reconfigure the app to send the TargetVariant query string parameter for users who subscribed to the preview featur
F. 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.
G. 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 featur
H. When the new version of the model is ready for release, gradually increase DesiredWeight until all users have the updated version.
I. Configure two SageMaker hosted endpoints that serve the different versions of the model
J. Create an Amazon Route 53 record that is configured with a simple routing policy and that points to the current version of the mode
K. 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 user
L. 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: D

NEW QUESTION 112
An employee found a video clip with audio on a company's social media feed. The language used in the video is Spanish. English is the employee's first language, and they do not understand Spanish. The employee wants to do a sentiment analysis.
What combination of services is the MOST efficient to accomplish the task?

A. Amazon Transcribe, Amazon Translate, and Amazon Comprehend
B. Amazon Transcribe, Amazon Comprehend, and Amazon SageMaker seq2seq
C. Amazon Transcribe, Amazon Translate, and Amazon SageMaker Neural Topic Model (NTM)
D. Amazon Transcribe, Amazon Translate, and Amazon SageMaker BlazingText

Answer: A

NEW QUESTION 113
A retail company uses a machine learning (ML) model for daily sales forecasting. The company's brand manager reports that the model has provided inaccurate results for the past 3 weeks.
At the end of each day, an AWS Glue job consolidates the input data that is used for the forecasting with the actual daily sales data and the predictions of the model. The AWS Glue job stores the data in Amazon S3. The company's ML team is using an Amazon SageMaker Studio notebook to gain an understanding about the source of the model's inaccuracies.
What should the ML team do on the SageMaker Studio notebook to visualize the model's degradation MOST accurately?

A. Create a histogram of the daily sales over the last 3 week
B. In addition, create a histogram of the daily sales from before that period.
C. Create a histogram of the model errors over the last 3 week
D. In addition, create a histogram of the model errors from before that period.
E. Create a line chart with the weekly mean absolute error (MAE) of the model.
F. Create a scatter plot of daily sales versus model error for the last 3 week
G. In addition, create a scatter plot of daily sales versus model error from before that period.

Answer: C

NEW QUESTION 114
A Machine Learning Specialist is designing a system for improving sales for a company. The objective is to use the large amount of information the company has on users' behavior and product preferences to predict which products users would like based on the users' similarity to other users.
What should the Specialist do to meet this objective?

A. Build a content-based filtering recommendation engine with Apache Spark ML on Amazon EMR.
B. Build a collaborative filtering recommendation engine with Apache Spark ML on Amazon EMR.
C. Build a model-based filtering recommendation engine with Apache Spark ML on Amazon EMR.
D. Build a combinative filtering recommendation engine with Apache Spark ML on Amazon EMR.

Answer: B

Explanation:
Many developers want to implement the famous Amazon model that was used to power the "People who bought this also bought these items" feature on Amazon.com. This model is based on a method called Collaborative Filtering. It takes items such as movies, books, and products that were rated highly by a set of users and recommending them to other users who also gave them high ratings. This method works well in domains where explicit ratings or implicit user actions can be gathered and analyzed.

NEW QUESTION 115
A company's Machine Learning Specialist needs to improve the training speed of a time-series forecasting model using TensorFlow. The training is currently implemented on a single-GPU machine and takes approximately 23 hours to complete. The training needs to be run daily.
The model accuracy js acceptable, but the company anticipates a continuous increase in the size of the training data and a need to update the model on an hourly, rather than a daily, basis. The company also wants to minimize coding effort and infrastructure changes
What should the Machine Learning Specialist do to the training solution to allow it to scale for future demand?

A. Do not change the TensorFlow cod
B. Change the machine to one with a more powerful GPU to speed up the training.
C. Change the TensorFlow code to implement a Horovod distributed framework supported by Amazon SageMake
D. Parallelize the training to as many machines as needed to achieve the business goals.
E. Switch to using a built-in AWS SageMaker DeepAR mode
F. Parallelize the training to as many machines as needed to achieve the business goals.
G. Move the training to Amazon EMR and distribute the workload to as many machines as needed to achieve the business goals.

Answer: B

NEW QUESTION 119
A machine learning specialist is developing a proof of concept for government users whose primary concern is security. The specialist is using Amazon SageMaker to train a convolutional neural network (CNN) model for a photo classifier application. The specialist wants to protect the data so that it cannot be accessed and transferred to a remote host by malicious code accidentally installed on the training container.
Which action will provide the MOST secure protection?

A. Remove Amazon S3 access permissions from the SageMaker execution role.
B. Encrypt the weights of the CNN model.
C. Encrypt the training and validation dataset.
D. Enable network isolation for training jobs.

Answer: D

NEW QUESTION 122
A Machine Learning Specialist is building a supervised model that will evaluate customers' satisfaction with their mobile phone service based on recent usage The model's output should infer whether or not a customer is likely to switch to a competitor in the next 30 days
Which of the following modeling techniques should the Specialist use1?

A. Time-series prediction
B. Anomaly detection
C. Binary classification
D. Regression

Answer: D

NEW QUESTION 123
An interactive online dictionary wants to add a widget that displays words used in similar contexts. A Machine Learning Specialist is asked to provide word features for the downstream nearest neighbor model powering the widget.
What should the Specialist do to meet these requirements?

A. Create one-hot word encoding vectors.
B. Produce a set of synonyms for every word using Amazon Mechanical Turk.
C. Create word embedding factors that store edit distance with every other word.
D. Download word embedding's pre-trained on a large corpus.

Answer: D

NEW QUESTION 126
A Machine Learning Specialist is applying a linear least squares regression model to a dataset with 1 000 records and 50 features Prior to training, the ML Specialist notices that two features are perfectly linearly dependent
Why could this be an issue for the linear least squares regression model?

A. It could cause the backpropagation algorithm to fail during training
B. It could create a singular matrix during optimization which fails to define a unique solution
C. It could modify the loss function during optimization causing it to fail during training
D. It could introduce non-linear dependencies within the data which could invalidate the linear assumptions of the model

Answer: C

NEW QUESTION 130
A Machine Learning Specialist is building a model to predict future employment rates based on a wide range of economic factors While exploring the data, the Specialist notices that the magnitude of the input features vary greatly The Specialist does not want variables with a larger magnitude to dominate the model
What should the Specialist do to prepare the data for model training'?

A. Apply quantile binning to group the data into categorical bins to keep any relationships in the data by replacing the magnitude with distribution
B. Apply the Cartesian product transformation to create new combinations of fields that are independent of the magnitude
C. Apply normalization to ensure each field will have a mean of 0 and a variance of 1 to remove any significant magnitude
D. Apply the orthogonal sparse Diagram (OSB) transformation to apply a fixed-size sliding window to generate new features of a similar magnitude.

Answer: C

NEW QUESTION 132
A retail company is selling products through a global online marketplace. The company wants to use machine learning (ML) to analyze customer feedback and identify specific areas for improvement. A developer has built a tool that collects customer reviews from the online marketplace and stores them in an Amazon S3 bucket. This process yields a dataset of 40 reviews. A data scientist building the ML models must identify additional sources of data to increase the size of the dataset.
Which data sources should the data scientist use to augment the dataset of reviews? (Choose three.)

A. Emails exchanged by customers and the company's customer service agents
B. Social media posts containing the name of the company or its products
C. A publicly available collection of news articles
D. A publicly available collection of customer reviews
E. Product sales revenue figures for the company
F. Instruction manuals for the company's products

Answer: BDF

NEW QUESTION 136
During mini-batch training of a neural network for a classification problem, a Data Scientist notices that training accuracy oscillates What is the MOST likely cause of this issue?

A. The class distribution in the dataset is imbalanced
B. Dataset shuffling is disabled
C. The batch size is too big
D. The learning rate is very high

Answer: B

NEW QUESTION 139
A data scientist uses an Amazon SageMaker notebook instance to conduct data exploration and analysis. This requires certain Python packages that are not

natively available on Amazon SageMaker to be installed on the notebook instance.
How can a machine learning specialist ensure that required packages are automatically available on the notebook instance for the data scientist to use?

A. Install AWS Systems Manager Agent on the underlying Amazon EC2 instance and use Systems Manager Automation to execute the package installation commands.
B. Create a Jupyter notebook file (.ipynb) with cells containing the package installation commands to execute and place the file under the /etc/init directory of each Amazon SageMaker notebook instance.
C. Use the conda package manager from within the Jupyter notebook console to apply the necessary conda packages to the default kernel of the notebook.
D. Create an Amazon SageMaker lifecycle configuration with package installation commands and assign the lifecycle configuration to the notebook instance.

Answer: D

Explanation:
https://docs.aws.amazon.com/sagemaker/latest/dg/nbi-add-external.html

NEW QUESTION 144
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.

NEW QUESTION 147
A Machine Learning Specialist is assigned to a Fraud Detection team and must tune an XGBoost model, which is working appropriately for test data. However, with unknown data, it is not working as expected. The existing parameters are provided as follows.

```
param = {
    'eta': 0.05,   # the training step for each iteration
    'silent': 1,   # logging mode - quiet
    'n_estimators':2000,
    'max_depth':30,
    'min_child_weight': 3,
    'gamma': 0,
    'subsample': 0.8,
    'objective': 'multi:softprob',   # error evaluation for multiclass training
    'num_class': 201}   # the number of classes that exist in this dataset
num_round = 60   # the number of training iterations
```

Which parameter tuning guidelines should the Specialist follow to avoid overfitting?

A. Increase the max_depth parameter value.
B. Lower the max_depth parameter value.
C. Update the objective to binary:logistic.
D. Lower the min_child_weight parameter value.

Answer: B

NEW QUESTION 148
A Machine Learning Specialist is packaging a custom ResNet model into a Docker container so the company can leverage Amazon SageMaker for training. The Specialist is using Amazon EC2 P3 instances to train the model and needs to properly configure the Docker container to leverage the NVIDIA GPUs.
What does the Specialist need to do?

A. Bundle the NVIDIA drivers with the Docker image.
B. Build the Docker container to be NVIDIA-Docker compatible.
C. Organize the Docker container's file structure to execute on GPU instances.
D. Set the GPU flag in the Amazon SageMaker CreateTrainingJob request body

Answer: B

NEW QUESTION 150
A manufacturing company has structured and unstructured data stored in an Amazon S3 bucket. A Machine Learning Specialist wants to use SQL to run queries on this data.
Which solution requires the LEAST effort to be able to query this data?

A. Use AWS Data Pipeline to transform the data and Amazon RDS to run queries.
B. Use AWS Glue to catalogue the data and Amazon Athena to run queries.
C. Use AWS Batch to run ETL on the data and Amazon Aurora to run the queries.
D. Use AWS Lambda to transform the data and Amazon Kinesis Data Analytics to run queries.

Answer: B

NEW QUESTION 153
A data scientist is using an Amazon SageMaker notebook instance and needs to securely access data stored in a specific Amazon S3 bucket.
How should the data scientist accomplish this?

A. Add an S3 bucket policy allowing GetObject, PutObject, and ListBucket permissions to the AmazonSageMaker notebook ARN as principal.
B. Encrypt the objects in the S3 bucket with a custom AWS Key Management Service (AWS KMS) key that only the notebook owner has access to.
C. Attach the policy to the IAM role associated with the notebook that allows GetObject, PutObject, and ListBucket operations to the specific S3 bucket.
D. Use a script in a lifecycle configuration to configure the AWS CLI on the instance with an access key ID and secret.

Answer: C

NEW QUESTION 157
A Machine Learning Specialist is developing recommendation engine for a photography blog Given a picture, the recommendation engine should show a picture that captures similar objects The Specialist would like to create a numerical representation feature to perform nearest-neighbor searches
What actions would allow the Specialist to get relevant numerical representations?

A. Reduce image resolution and use reduced resolution pixel values as features
B. Use Amazon Mechanical Turk to label image content and create a one-hot representation indicating the presence of specific labels
C. Run images through a neural network pie-trained on ImageNet, and collect the feature vectors from the penultimate layer
D. Average colors by channel to obtain three-dimensional representations of images.

Answer: A

NEW QUESTION 162
For the given confusion matrix, what is the recall and precision of the model?

		Actual	
		Yes	No
Predicted	Yes	12	3
	No	1	9

A. Recall = 0.92 Precision = 0.84
B. Recall = 0.84 Precision = 0.8
C. Recall = 0.92 Precision = 0.8
D. Recall = 0.8 Precision = 0.92

Answer: C

NEW QUESTION 165
......

Thank You for Trying Our Product

We offer two products:

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

2nd - Questons and Answers in PDF Format

AWS-Certified-Machine-Learning-Specialty Practice Exam Features:

* AWS-Certified-Machine-Learning-Specialty Questions and Answers Updated Frequently

* AWS-Certified-Machine-Learning-Specialty Practice Questions Verified by Expert Senior Certified Staff

* AWS-Certified-Machine-Learning-Specialty Most Realistic Questions that Guarantee you a Pass on Your FirstTry

* AWS-Certified-Machine-Learning-Specialty Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

100% Actual & Verified — Instant Download, Please Click
Order The AWS-Certified-Machine-Learning-Specialty Practice Test Here