Amazon-Web-Services

Exam Questions MLS-C01

AWS Certified Machine Learning - Specialty

NEW QUESTION 1

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: B

NEW QUESTION 2

A Machine Learning Specialist built an image classification deep learning model. However the Specialist ran into an overfitting problem in which the training and testing accuracies were 99% and 75%r respectively.
How should the Specialist address this issue and what is the reason behind it?

A. The learning rate should be increased because the optimization process was trapped at a local minimum.
B. The dropout rate at the flatten layer should be increased because the model is not generalized enough.
C. The dimensionality of dense layer next to the flatten layer should be increased because the model is not complex enough.
D. The epoch number should be increased because the optimization process was terminated before it reached the global minimum.

Answer: D

NEW QUESTION 3

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: A

NEW QUESTION 4

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: BD

NEW QUESTION 5

Example Corp has an annual sale event from October to December. The company has sequential sales data from the past 15 years and wants to use Amazon ML to predict the sales for this year's upcoming event. Which method should Example Corp use to split the data into a training dataset and evaluation dataset?

A. Pre-split the data before uploading to Amazon S3
B. Have Amazon ML split the data randomly.
C. Have Amazon ML split the data sequentially.
D. Perform custom cross-validation on the data

Answer: C

NEW QUESTION 6

An Machine Learning Specialist discover the following statistics while experimenting on a model.

Experiment 1
Baseline model
Train error = 5%
Test error = 16%

Experiment 2
The Specialist added more layers and neurons to the model and received the following results:
Train error = 5.2%
Test error = 15.7%

Experiment 3
The Specialist reverted back to the original number of neurons from Experiment 1 and implemented regularization in the neural network, which yielded the following results
Train error = 4.7%
Test error = 9.5%

What can the Specialist from the experiments?

A. The model In Experiment 1 had a high variance error lhat was reduced in Experiment 3 by regularization Experiment 2 shows that there is minimal bias error in Experiment 1
B. The model in Experiment 1 had a high bias error that was reduced in Experiment 3 by regularization Experiment 2 shows that there is minimal variance error in Experiment 1
C. The model in Experiment 1 had a high bias error and a high variance error that were reduced in Experiment 3 by regularization Experiment 2 shows thai high bias cannot be reduced by increasing layers and neurons in the model
D. The model in Experiment 1 had a high random noise error that was reduced in Expenment 3 by regularization Expenment 2 shows that random noise cannot be reduced by increasing layers and neurons in the model

Answer: C

NEW QUESTION 7
A Machine Learning Specialist is preparing data for training on Amazon SageMaker The Specialist 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

NEW QUESTION 8
A Machine Learning Specialist deployed a model that provides product recommendations on a company's website Initially, the model was performing very well and resulted in customers buying more products on average However within the past few months the Specialist has noticed that the effect of product recommendations has diminished and customers are starting to return to their original habits of spending less The Specialist is unsure of what happened, as the model has not changed from its initial deployment over a year ago
Which method should the Specialist try to improve model performance?

A. The model needs to be completely re-engineered because it is unable to handle product inventory changes
B. The model's hyperparameters should be periodically updated to prevent drift
C. The model should be periodically retrained from scratch using the original data while adding a regularization term to handle product inventory changes
D. The model should be periodically retrained using the original training data plus new data as product inventory changes
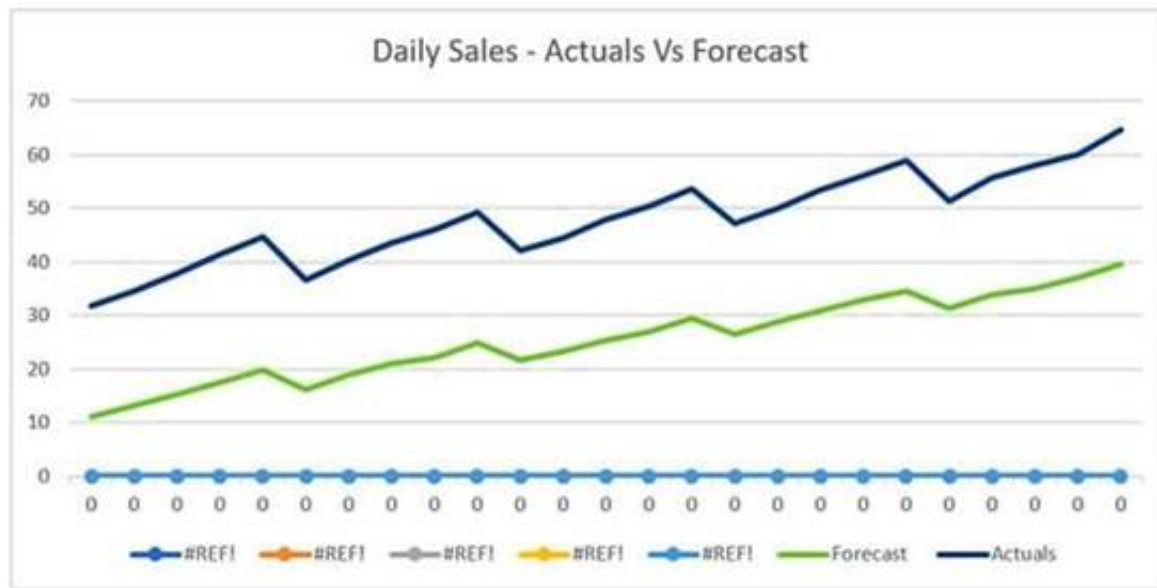
Answer: D

NEW QUESTION 9
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 10
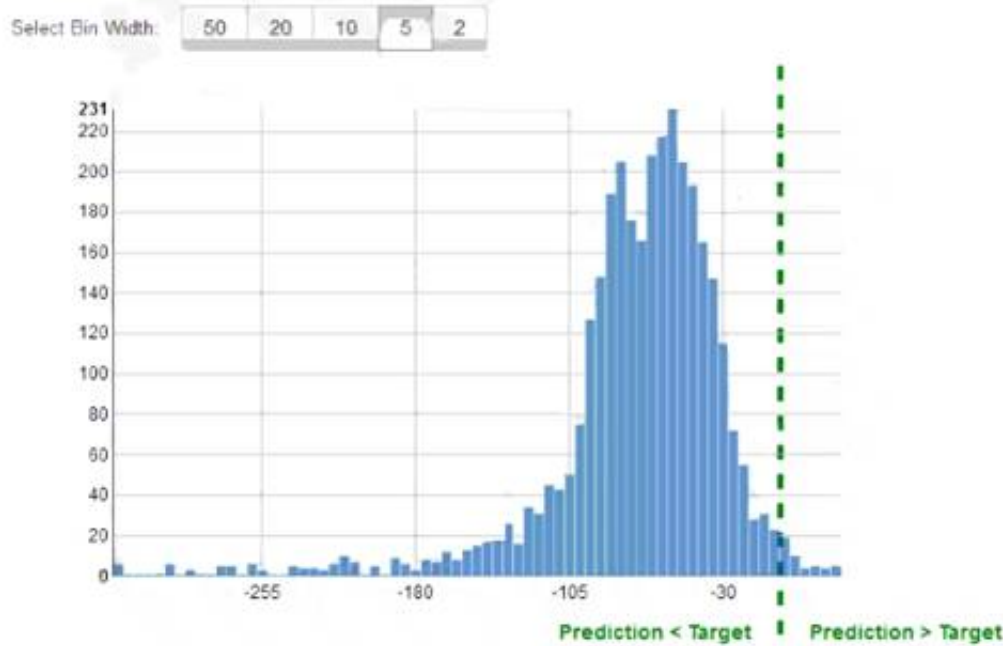The displayed graph is from a foresting model for testing a time series.

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 10

While reviewing the histogram for residuals on regression evaluation data a Machine Learning Specialist notices that the residuals do not form a zero-centered bell shape as shown What does this mean?

A. The model might have prediction errors over a range of target values.
B. The dataset cannot be accurately represented using the regression model
C. There are too many variables in the model
D. The model is predicting its target values perfectly.
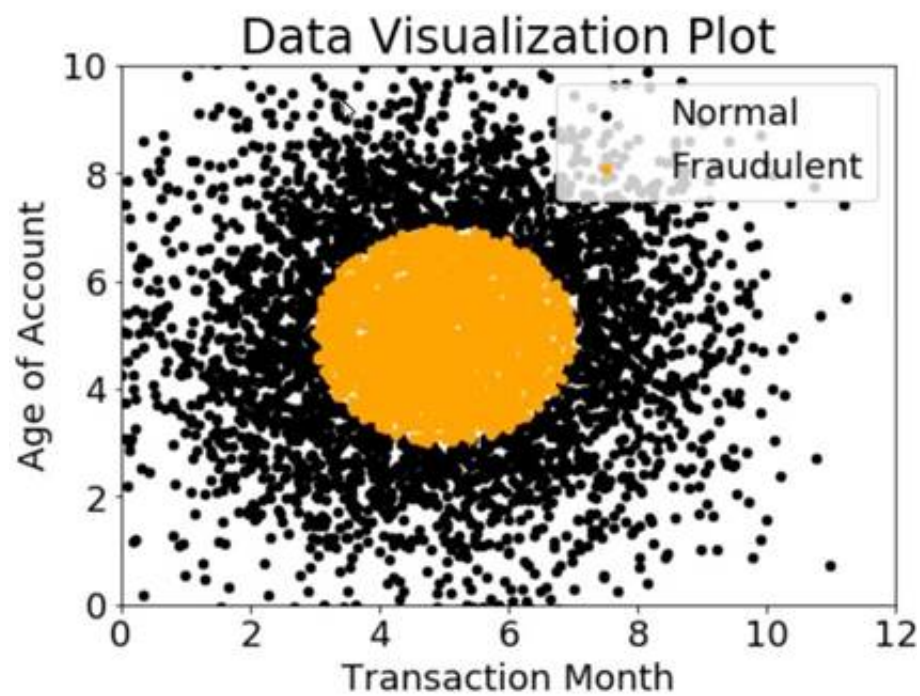
Answer: D

NEW QUESTION 15

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 17

A company wants to classify user behavior as either fraudulent or normal. Based on internal research, a Machine Learning Specialist would like to build a binary classifier based on two features: age of account and transaction month. The class distribution for these features is illustrated in the figure provided.

Based on this information, which model would have the HIGHEST recall with respect to the fraudulent class?

A. Decision tree
B. Linear support vector machine (SVM)
C. Naive Bayesian classifier
D. Single Perceptron with sigmoidal activation function

Answer: C

NEW QUESTION 18
A Machine Learning Specialist working for an online fashion company wants to build a data ingestion solution for the company's Amazon S3-based data lake.
The Specialist wants to create a set of ingestion mechanisms that will enable future capabilities comprised of:
• Real-time analytics
• Interactive analytics of historical data
• Clickstream analytics
• Product recommendations
Which services should the Specialist use?

A. AWS Glue as the data dialog; Amazon Kinesis Data Streams and Amazon Kinesis Data Analytics for real-time data insights; Amazon Kinesis Data Firehose for delivery to Amazon ES for clickstream analytics; Amazon EMR to generate personalized product recommendations
B. Amazon Athena as the data catalog; Amazon Kinesis Data Streams and Amazon Kinesis Data Analytics for near-realtime data insights; Amazon Kinesis Data Firehose for clickstream analytics; AWS Glue to generate personalized product recommendations
C. AWS Glue as the data catalog; Amazon Kinesis Data Streams and Amazon Kinesis Data Analytics for historical data insights; Amazon Kinesis Data Firehose for delivery to Amazon ES for clickstream analytics; Amazon EMR to generate personalized product recommendations
D. Amazon Athena as the data catalog; Amazon Kinesis Data Streams and Amazon Kinesis Data Analytics for historical data insights; Amazon DynamoDB streams for clickstream analytics; AWS Glue to generate personalized product recommendations

Answer: A

NEW QUESTION 23
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: BDE

NEW QUESTION 24
While working on a neural network project, a Machine Learning Specialist discovers thai some features in the data have very high magnitude resulting in this data being weighted more in the cost function What should the Specialist do to ensure better convergence during backpropagation?

A. Dimensionality reduction
B. Data normalzation
C. Model regulanzation
D. Data augmentation for the minority class

Answer: D

NEW QUESTION 26
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 27
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 30
An insurance company is developing a new device for vehicles that uses a camera to observe drivers' behavior and alert them when they appear distracted The company created approximately 10,000 training images in a controlled environment that a Machine Learning Specialist will use to train and evaluate machine learning models
During the model evaluation the Specialist notices that the training error rate diminishes faster as the number of epochs increases and the model is not accurately inferring on the unseen test images
Which of the following should be used to resolve this issue? (Select TWO)

A. Add vanishing gradient to the model
B. Perform data augmentation on the training data
C. Make the neural network architecture complex.
D. Use gradient checking in the model
E. Add L2 regularization to the model

Answer: BD

NEW QUESTION 33
Which of the following metrics should a Machine Learning Specialist generally use to compare/evaluate machine learning classification models against each other?

A. Recall
B. Misclassification rate
C. Mean absolute percentage error (MAPE)
D. Area Under the ROC Curve (AUC)

Answer: A

NEW QUESTION 36
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 do1?

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 Create TrainingJob request body

Answer: A

NEW QUESTION 40
A company is setting up an Amazon SageMaker environment. The corporate data security policy does not allow communication over the internet.
How can the company enable the Amazon SageMaker service without enabling direct internet access to Amazon SageMaker notebook instances?

A. Create a NAT gateway within the corporate VPC.
B. Route Amazon SageMaker traffic through an on-premises network.

C. Create Amazon SageMaker VPC interface endpoints within the corporate VPC.
D. Create VPC peering with Amazon VPC hosting Amazon SageMaker.

Answer: A

NEW QUESTION 42
A retail company intends to use machine learning to categorize new products A labeled dataset of current products was provided to the Data Science team The dataset includes 1 200 products The labeled dataset has 15 features for each product such as title dimensions, weight, and price Each product is labeled as belonging to one of six categories such as books, games, electronics, and movies.
Which model should be used for categorizing new products using the provided dataset for training?

A. An XGBoost model where the objective parameter is set to multi: softmax
B. A deep convolutional neural network (CNN) with a softmax activation function for the last layer
C. A regression forest where the number of trees is set equal to the number of product categories
D. A DeepAR forecasting model based on a recurrent neural network (RNN)

Answer: B

NEW QUESTION 46
A monitoring service generates 1 TB of scale metrics record data every minute A Research team performs queries on this data using Amazon Athena The queries run slowly due to the large volume of data, and the team requires better performance
How should the records be stored in Amazon S3 to improve query performance?

A. CSV files
B. Parquet files
C. Compressed JSON
D. RecordIO

Answer: B

NEW QUESTION 51
A company is running an Amazon SageMaker training job that will access data stored in its Amazon S3 bucket A compliance policy requires that the data never be transmitted across the internet How should the company set up the job?

A. Launch the notebook instances in a public subnet and access the data through the public S3 endpoint
B. Launch the notebook instances in a private subnet and access the data through a NAT gateway
C. Launch the notebook instances in a public subnet and access the data through a NAT gateway
D. Launch the notebook instances in a private subnet and access the data through an S3 VPC endpoint.

Answer: D

NEW QUESTION 56
A Mobile Network Operator is building an analytics platform to analyze and optimize a company's operations using Amazon Athena and Amazon S3
The source systems send data in CSV format in real lime The Data Engineering team wants to transform the data to the Apache Parquet format before storing it on Amazon S3
Which solution takes the LEAST effort to implement?

A. Ingest .CSV data using Apache Kafka Streams on Amazon EC2 instances and use Kafka Connect S3 to serialize data as Parquet
B. Ingest .CSV data from Amazon Kinesis Data Streams and use Amazon Glue to convert data into Parquet.
C. Ingest .CSV data using Apache Spark Structured Streaming in an Amazon EMR cluster and use Apache Spark to convert data into Parquet.
D. Ingest .CSV data from Amazon Kinesis Data Streams and use Amazon Kinesis Data Firehose to convert data into Parquet.

Answer: C

NEW QUESTION 57
A retail chain has been ingesting purchasing records from its network of 20,000 stores to Amazon S3 using Amazon Kinesis Data Firehose To support training an improved machine learning model, training records will require new but simple transformations, and some attributes will be combined The model needs lo be retrained daily
Given the large number of stores and the legacy data ingestion, which change will require the LEAST amount of development effort?

A. Require that the stores to switch to capturing their data locally on AWS Storage Gateway for loading into Amazon S3 then use AWS Glue to do the transformation
B. Deploy an Amazon EMR cluster running Apache Spark with the transformation logic, and have the cluster run each day on the accumulating records in Amazon S3, outputting new/transformed records to Amazon S3
C. Spin up a fleet of Amazon EC2 instances with the transformation logic, have them transform the data records accumulating on Amazon S3, and output the transformed records to Amazon S3.
D. Insert an Amazon Kinesis Data Analytics stream downstream of the Kinesis Data Firehouse stream that transforms raw record attributes into simple transformed values using SQL.

Answer: D

NEW QUESTION 58
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 60
A Machine Learning Specialist has completed a proof of concept for a company using a small data sample and now the Specialist is ready to implement an end-to-end solution in AWS using Amazon SageMaker The historical training data is stored in Amazon RDS
Which approach should the Specialist use for training a model using that data?

A. Write a direct connection to the SQL database within the notebook and pull data in
B. Push the data from Microsoft SQL Server to Amazon S3 using an AWS Data Pipeline and provide the S3 location within the notebook.
C. Move the data to Amazon DynamoDB and set up a connection to DynamoDB within the notebook to pull data in
D. Move the data to Amazon ElastiCache using AWS DMS and set up a connection within the notebook to pull data in for fast access.

Answer: B

NEW QUESTION 64
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 68
A company has raw user and transaction data stored in AmazonS3 a MySQL database, and Amazon RedShift A Data Scientist needs to perform an analysis by joining the three datasets from Amazon S3, MySQL, and Amazon RedShift, and then calculating the average-of a few selected columns from the joined data
Which AWS service should the Data Scientist use?

A. Amazon Athena
B. Amazon Redshift Spectrum
C. AWS Glue
D. Amazon QuickSight

Answer: A

NEW QUESTION 70
A Machine Learning Specialist is building a model that will perform time series forecasting using Amazon SageMaker The Specialist has finished training the model and is now planning to perform load testing on the endpoint so they can configure Auto Scaling for the model variant
Which approach will allow the Specialist to review the latency, memory utilization, and CPU utilization during the load test"?

A. Review SageMaker logs that have been written to Amazon S3 by leveraging Amazon Athena and Amazon OuickSight to visualize logs as they are being produced
B. Generate an Amazon CloudWatch dashboard to create a single view for the latency, memory utilization,and CPU utilization metrics that are outputted by Amazon SageMaker
C. Build custom Amazon CloudWatch Logs and then leverage Amazon ES and Kibana to query and visualize the data as it is generated by Amazon SageMaker
D. Send Amazon CloudWatch Logs that were generated by Amazon SageMaker lo Amazon ES and use Kibana to query and visualize the log data.

Answer: B

NEW QUESTION 75
A financial services company is building a robust serverless data lake on Amazon S3. The data lake should be flexible and meet the following requirements:
* Support querying old and new data on Amazon S3 through Amazon Athena and Amazon Redshift Spectrum.
* Support event-driven ETL pipelines.
* Provide a quick and easy way to understand metadata. Which approach meets trfese requirements?

A. Use an AWS Glue crawler to crawl S3 data, an AWS Lambda function to trigger an AWS Glue ETL job, and an AWS Glue Data catalog to search and discover metadata.
B. Use an AWS Glue crawler to crawl S3 data, an AWS Lambda function to trigger an AWS Batch job, and an external Apache Hive metastore to search and discover metadata.
C. Use an AWS Glue crawler to crawl S3 data, an Amazon CloudWatch alarm to trigger an AWS Batch job, and an AWS Glue Data Catalog to search and discover metadata.
D. Use an AWS Glue crawler to crawl S3 data, an Amazon CloudWatch alarm to trigger an AWS Glue ETL job, and an external Apache Hive metastore to search and discover metadata.

Answer: B

NEW QUESTION 78
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
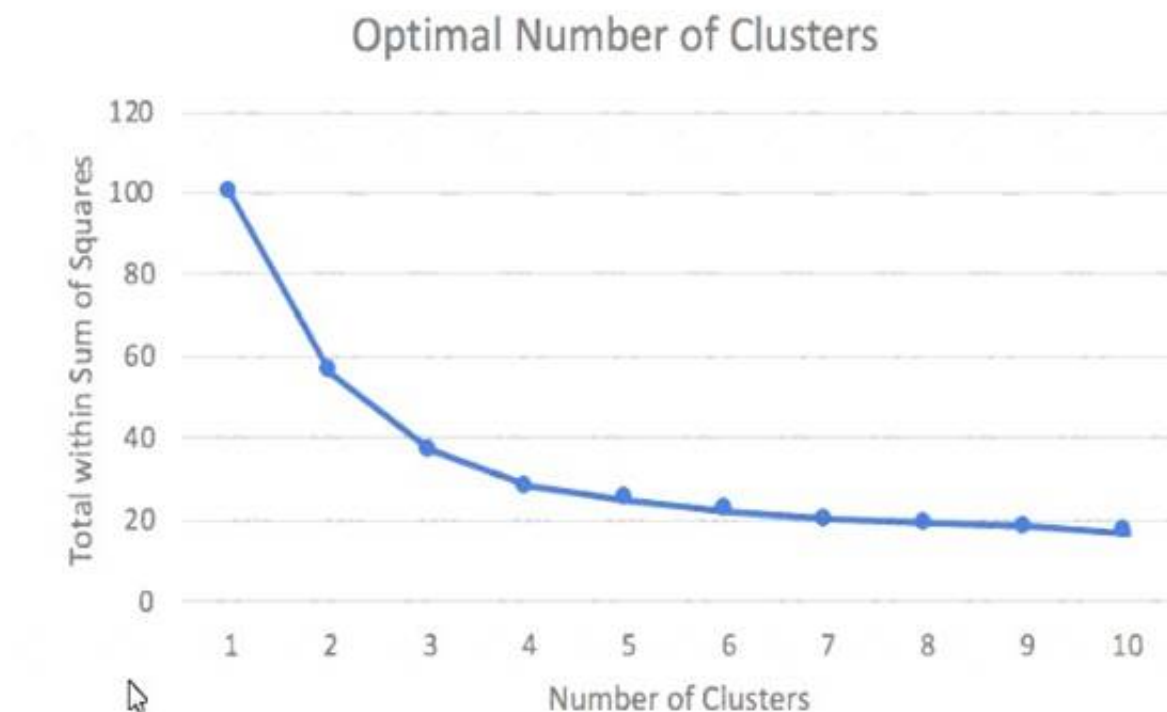
dataset could be used to reconstruct the missing data
Which reconstruction approach should the Specialist use to preserve the integrity of the dataset?

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

Answer: C

NEW QUESTION 82
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 83
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 87
A company is observing low accuracy while training on the default built-in image classification algorithm in Amazon SageMaker. The Data Science team wants to use an Inception neural network architecture instead of a ResNet architecture.
Which of the following will accomplish this? (Select TWO.)

A. Customize the built-in image classification algorithm to use Inception and use this for model training.
B. Create a support case with the SageMaker team to change the default image classification algorithm to Inception.
C. Bundle a Docker container with TensorFlow Estimator loaded with an Inception network and use this for model training.
D. Use custom code in Amazon SageMaker with TensorFlow Estimator to load the model with an Inception network and use this for model training.
E. Download and apt-get install the inception network code into an Amazon EC2 instance and use this instance as a Jupyter notebook in Amazon SageMaker.

Answer: AD

NEW QUESTION 88
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 89
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 94
A Machine Learning Specialist kicks off a hyperparameter tuning job for a tree-based ensemble model using Amazon SageMaker with Area Under the ROC Curve (AUC) as the objective metric This workflow will eventually be deployed in a pipeline that retrains and tunes hyperparameters each night to model click-through on data that goes stale every 24 hours
With the goal of decreasing the amount of time it takes to train these models, and ultimately to decrease costs, the Specialist wants to reconfigure the input hyperparameter range(s)
Which visualization will accomplish this?

A. A histogram showing whether the most important input feature is Gaussian.
B. A scatter plot with points colored by target variable that uses (-Distributed Stochastic Neighbor Embedding (I-SNE) to visualize the large number of input variables in an easier-to-read dimension.
C. A scatter plot showing (he performance of the objective metric over each training iteration
D. A scatter plot showing the correlation between maximum tree depth and the objective metric.

Answer: B

NEW QUESTION 97
A Marketing Manager at a pet insurance company plans to launch a targeted marketing campaign on social media to acquire new customers Currently, the company has the following data in Amazon Aurora
• Profiles for all past and existing customers
• Profiles for all past and existing insured pets
• Policy-level information
• Premiums received
• Claims paid
What steps should be taken to implement a machine learning model to identify potential new customers on social media?

A. Use regression on customer profile data to understand key characteristics of consumer segments Find similar profiles on social media.
B. Use clustering on customer profile data to understand key characteristics of consumer segments Find similar profiles on social media.
C. Use a recommendation engine on customer profile data to understand key characteristics of consumer segment
D. Find similar profiles on social media
E. Use a decision tree classifier engine on customer profile data to understand key characteristics of consumer segment
F. Find similar profiles on social media

Answer: C

NEW QUESTION 102
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: D

NEW QUESTION 106
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 service accounts.
D. Amazon SageMaker notebook instances are based on AWS ECS instances running within AWS service accounts.

Answer: C

NEW QUESTION 110
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: B

NEW QUESTION 114
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 quenes
D. Use AWS Lambda to transform the data and Amazon Kinesis Data Analytics to run queries

Answer: D

NEW QUESTION 116
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: A

NEW QUESTION 118
A Data Engineer needs to build a model using a dataset containing customer credit card information.
How can the Data Engineer ensure the data remains encrypted and the credit card information is secure? Use a custom encryption algorithm to encrypt the data and store the data on an Amazon SageMaker instance in a VPC. Use the SageMaker DeepAR algorithm to randomize the credit card numbers.

A. Use an IAM policy to encrypt the data on the Amazon S3 bucket and Amazon Kinesis to automatically discard credit card numbers and insert fake credit card numbers.
B. Use an Amazon SageMaker launch configuration to encrypt the data once it is copied to the SageMaker instance in a VP
C. Use the SageMaker principal component analysis (PCA) algorithm to reduce the length of the credit card numbers.
D. Use AWS KMS to encrypt the data on Amazon S3

Answer: C

NEW QUESTION 120
......

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

MLS-C01 Practice Exam Features:

* MLS-C01 Questions and Answers Updated Frequently

* MLS-C01 Practice Questions Verified by Expert Senior Certified Staff

* MLS-C01 Most Realistic Questions that Guarantee you a Pass on Your FirstTry

* MLS-C01 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

100% Actual & Verified — Instant Download, Please Click
Order The MLS-C01 Practice Test Here