MLS-C01 Dumps

AWS Certified Machine Learning - Specialty

https://www.certleader.com/MLS-C01-dumps.html

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
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: D

NEW QUESTION 4
A Machine Learning Specialist is required to build a supervised image-recognition model to identify a cat. The ML Specialist performs some tests and records the following results for a neural network-based image classifier:
Total number of images available = 1,000 Test set images = 100 (constant test set)
The ML Specialist notices that, in over 75% of the misclassified images, the cats were held upside down by their owners.
Which techniques can be used by the ML Specialist to improve this specific test error?

A. Increase the training data by adding variation in rotation for training images.
B. Increase the number of epochs for model training.
C. Increase the number of layers for the neural network.
D. Increase the dropout rate for the second-to-last layer.

Answer: B

NEW QUESTION 5
A large JSON dataset for a project has been uploaded to a private Amazon S3 bucket The Machine Learning Specialist wants to securely access and explore the data from an Amazon SageMaker notebook instance A new VPC was created and assigned to the Specialist
How can the privacy and integrity of the data stored in Amazon S3 be maintained while granting access to the Specialist for analysis?

A. Launch the SageMaker notebook instance within the VPC with SageMaker-provided internet access enabled Use an S3 ACL to open read privileges to the everyone group
B. Launch the SageMaker notebook instance within the VPC and create an S3 VPC endpoint for the notebook to access the data Copy the JSON dataset from Amazon S3 into the ML storage volume on the SageMaker notebook instance and work against the local dataset
C. Launch the SageMaker notebook instance within the VPC and create an S3 VPC endpoint for the notebook to access the data Define a custom S3 bucket policy to only allow requests from your VPC toaccess the S3 bucket
D. Launch the SageMaker notebook instance within the VPC with SageMaker-provided internet access enable
E. Generate an S3 pre-signed URL for access to data in the bucket

Answer: B

NEW QUESTION 6
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 7
A gaming company has launched an online game where people can start playing for free but they need to pay if they choose to use certain features The company needs to build an automated system to predict whether or not a new user will become a paid user within 1 year The company has gathered a labeled dataset from 1 million users
The training dataset consists of 1.000 positive samples (from users who ended up paying within 1 year) and 999.1 negative samples (from users who did not use any paid features) Each data sample consists of 200 features including user age, device, location, and play patterns
Using this dataset for training, the Data Science team trained a random forest model that converged with over 99% accuracy on the training set However, the prediction results on a test dataset were not satisfactory.
Which of the following approaches should the Data Science team take to mitigate this issue? (Select TWO.)

A. Add more deep trees to the random forest to enable the model to learn more features.
B. indicate a copy of the samples in the test database in the training dataset
C. Generate more positive samples by duplicating the positive samples and adding a small amount of noise to the duplicated data.
D. Change the cost function so that false negatives have a higher impact on the cost value than false positives
E. Change the cost function so that false positives have a higher impact on the cost value than false negatives

Answer: BD

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
The displayed graph is from a foresting model for testing a time series.
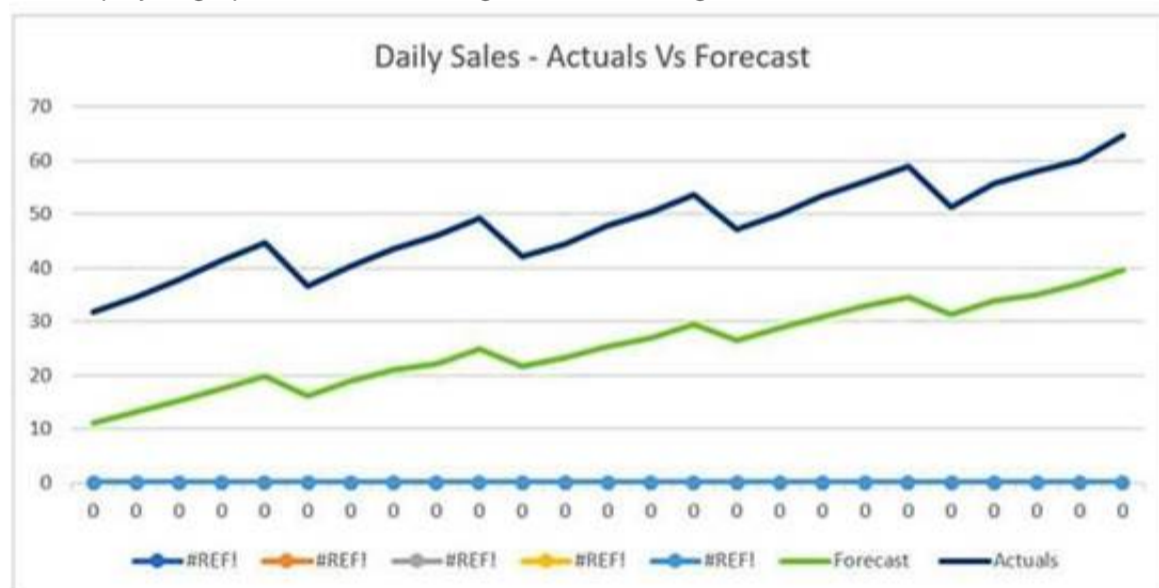
Considering the graph only, which conclusion should a Machine Learning Specialist make about the behavior of the model?

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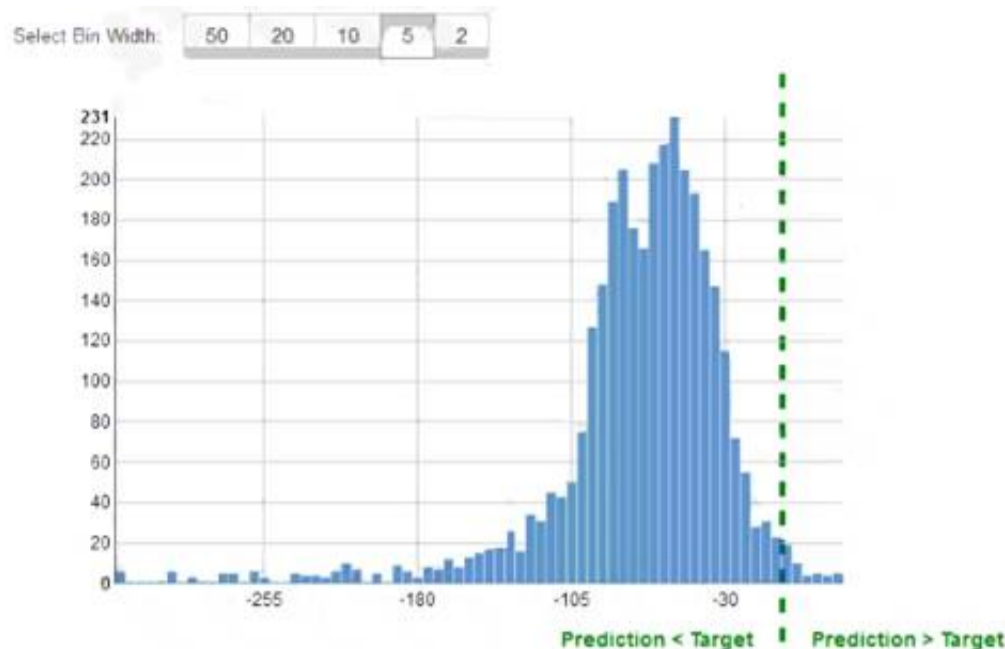
A. The model might have prediction errors over a range of target values.
B. The dataset cannot be accurately represented using the regression model
C. There are too many variables in the model
D. The model is predicting its target values perfectly.

Answer: D

NEW QUESTION 10
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 14
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 18
A large consumer goods manufacturer has the following products on sale
• 34 different toothpaste variants
• 48 different toothbrush variants
• 43 different mouthwash variants
The entire sales history of all these products is available in Amazon S3 Currently, the company is using custom-built autoregressive integrated moving average (ARIMA) models to forecast demand for these products The company wants to predict the demand for a new product that will soon be launched
Which solution should a Machine Learning Specialist apply?

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

Answer: B

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

NEW QUESTION 23
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 27
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 31
A Machine Learning Specialist is using Apache Spark for pre-processing training data As part of the Spark pipeline, the Specialist wants to use Amazon SageMaker for training a model and hosting it Which of the following would the Specialist do to integrate the Spark application with SageMaker? (Select THREE)

A. Download the AWS SDK for the Spark environment
B. Install the SageMaker Spark library in the Spark environment.
C. Use the appropriate estimator from the SageMaker Spark Library to train a model.
D. Compress the training data into a ZIP file and upload it to a pre-defined Amazon S3 bucket.
E. Use the sageMakerMode
F. transform method to get inferences from the model hosted in SageMaker
G. Convert the DataFrame object to a CSV file, and use the CSV file as input for obtaining inferences from SageMaker.

Answer: DEF

NEW QUESTION 35
A Machine Learning Specialist is configuring automatic model tuning in Amazon SageMaker
When using the hyperparameter optimization feature, which of the following guidelines should be followed to improve optimization?
Choose the maximum number of hyperparameters supported by

A. Amazon SageMaker to search the largest number of combinations possible
B. Specify a very large hyperparameter range to allow Amazon SageMaker to cover every possible value.
C. Use log-scaled hyperparameters to allow the hyperparameter space to be searched as quickly as possible
D. Execute only one hyperparameter tuning job at a time and improve tuning through successive rounds of experiments

Answer: C

NEW QUESTION 39
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 41
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 43
A Machine Learning Specialist is developing a custom video recommendation model for an application The dataset used to train this model is very large with millions of data points and is hosted in an Amazon S3 bucket The Specialist wants to avoid loading all of this data onto an Amazon SageMaker notebook instance because it would take hours to move and will exceed the attached 5 GB Amazon EBS volume on the notebook instance.
Which approach allows the Specialist to use all the data to train the model?

A. Load a smaller subset of the data into the SageMaker notebook and train locall
B. Confirm that the training code is executing and the model parameters seem reasonabl
C. Initiate a SageMaker training job using the full dataset from the S3 bucket using Pipe input mode.
D. Launch an Amazon EC2 instance with an AWS Deep Learning AMI and attach the S3 bucket to theinstanc
E. Train on a small amount of the data to verify the training code and hyperparameter
F. Go back toAmazon SageMaker and train using the full dataset
G. Use AWS Glue to train a model using a small subset of the data to confirm that the data will be compatiblewith Amazon SageMake
H. Initiate a SageMaker training job using the full dataset from the S3 bucket usingPipe input mode.
I. Load a smaller subset of the data into the SageMaker notebook and train locall
J. Confirm that the training code is executing and the model parameters seem reasonabl
K. Launch an Amazon EC2 instance with an AWS Deep Learning AMI and attach the S3 bucket to train the full dataset.

Answer: A

NEW QUESTION 46
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 48
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 50
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 53
A Data Scientist is developing a machine learning model to classify whether a financial transaction is fraudulent. The labeled data available for training consists of 100,000 non-fraudulent observations and 1,000 fraudulent observations.
The Data Scientist applies the XGBoost algorithm to the data, resulting in the following confusion matrix when the trained model is applied to a previously unseen validation dataset. The accuracy of the model is 99.1%, but the Data Scientist has been asked to reduce the number of false negatives.

Predicted   0   1
Actual   0 99,966 | 34
          1   877 | 123

Which combination of steps should the Data Scientist take to reduce the number of false positive predictions by the model? (Select TWO.)

A. Change the XGBoost eval_metric parameter to optimize based on rmse instead of error.
B. Increase the XGBoost scale_pos_weight parameter to adjust the balance of positive and negative weights.
C. Increase the XGBoost max_depth parameter because the model is currently underfitting the data.
D. Change the XGBoost evaljnetric parameter to optimize based on AUC instead of error.
E. Decrease the XGBoost max_depth parameter because the model is currently overfitting the data.

Answer: DE

NEW QUESTION 56
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: C

NEW QUESTION 58
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 60
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 65
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 68
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 71
Given the following confusion matrix for a movie classification model, what is the true class frequency for Romance and the predicted class frequency for Adventure?
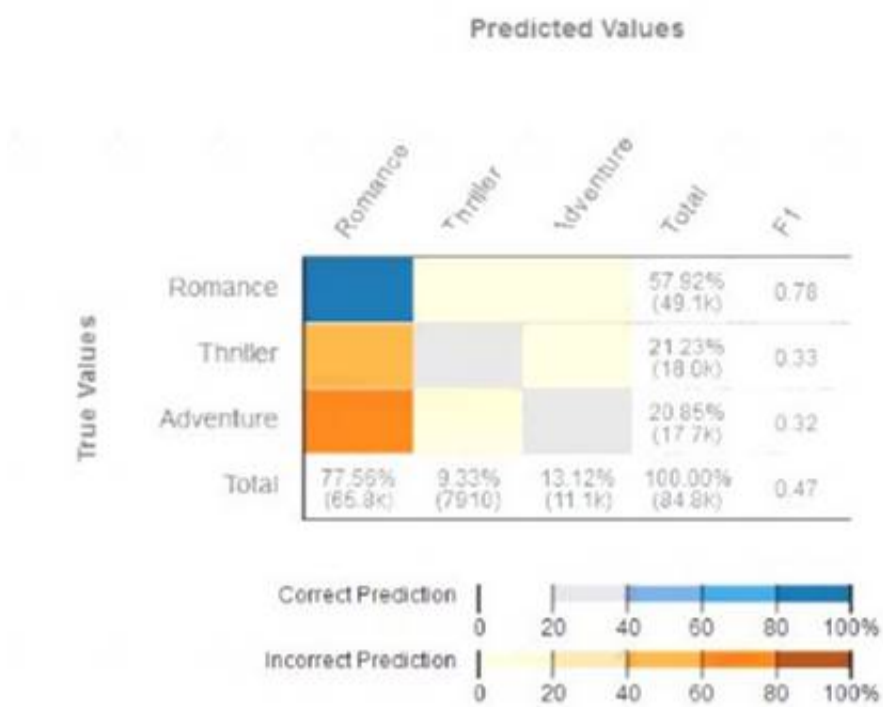
A. The true class frequency for Romance is 77.56% and the predicted class frequency for Adventure is 20 85%
B. The true class frequency for Romance is 57.92% and the predicted class frequency for Adventure is 1312%
C. The true class frequency for Romance is 0 78 and the predicted class frequency for Adventure is (0 47 - 0.32).
D. The true class frequency for Romance is 77.56% * 0.78 and the predicted class frequency for Adventure is 20 85% ' 0.32

Answer: A

NEW QUESTION 72
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 75
A Machine Learning Specialist receives customer data for an online shopping website. The data includes demographics, past visits, and locality information. The Specialist must develop a machine learning approach to identify the customer shopping patterns, preferences and trends to enhance the website for better service and smart recommendations.
Which solution should the Specialist recommend?

A. Latent Dirichlet Allocation (LDA) for the given collection of discrete data to identify patterns in the customer database.
B. A neural network with a minimum of three layers and random initial weights to identify patterns in the customer database
C. Collaborative filtering based on user interactions and correlations to identify patterns in the customer database
D. Random Cut Forest (RCF) over random subsamples to identify patterns in the customer database

Answer: C

NEW QUESTION 78
This graph shows the training and validation loss against the epochs for a neural network The network being trained is as follows
• Two dense layers one output neuron
• 100 neurons in each layer
• 100 epochs
• Random initialization of weights
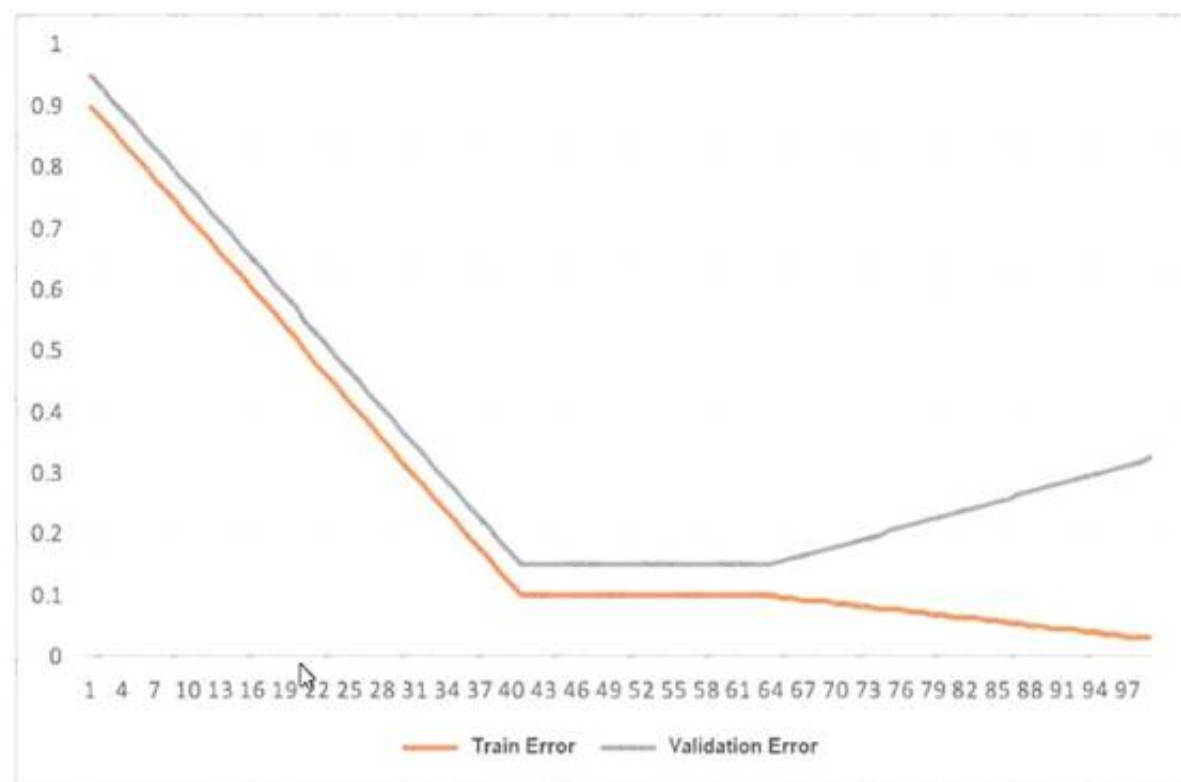
Which technique can be used to improve model performance in terms of accuracy in the validation set?

A. Early stopping
B. Random initialization of weights with appropriate seed
C. Increasing the number of epochs
D. Adding another layer with the 100 neurons

Answer: D

NEW QUESTION 80
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 83
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 86
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 89
A company is using Amazon Polly to translate plaintext documents to speech for automated company announcements However company acronyms are being mispronounced in the current documents How should a Machine Learning Specialist address this issue for future documents'?

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

Answer: A

NEW QUESTION 92
......

Thank You for Trying Our Product

* 100% Pass or Money Back

    All our products come with a 90-day Money Back Guarantee.

* One year free update

    You can enjoy free update one year. 24x7 online support.

* Trusted by Millions

    We currently serve more than 30,000,000 customers.

* Shop Securely

    All transactions are protected by VeriSign!

100% Pass Your MLS-C01 Exam with Our Prep Materials Via below:

https://www.certleader.com/MLS-C01-dumps.html