

Amazon-Web-Services

Exam Questions MLS-C01

AWS Certified Machine Learning - Specialty



NEW QUESTION 1

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 to access the S3 bucket.
- D. Launch the SageMaker notebook instance within the VPC with SageMaker-provided internet access enabled.
- E. Generate an S3 pre-signed URL for access to data in the bucket.

Answer: B

NEW QUESTION 2

A Data Science team is designing a dataset repository where it will store a large amount of training data commonly used in its machine learning models. As Data Scientists may create an arbitrary number of new datasets every day, the solution has to scale automatically and be cost-effective. Also, it must be possible to explore the data using SQL.

Which storage scheme is MOST adapted to this scenario?

- A. Store datasets as files in Amazon S3.
- B. Store datasets as files in an Amazon EBS volume attached to an Amazon EC2 instance.
- C. Store datasets as tables in a multi-node Amazon Redshift cluster.
- D. Store datasets as global tables in Amazon DynamoDB.

Answer: A

NEW QUESTION 3

An Machine Learning Specialist discovers 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 learn from the experiments?

- A. The model in Experiment 1 had a high variance error that 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 that 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 Experiment 3 by regularization. Experiment 2 shows that random noise cannot be reduced by increasing layers and neurons in the model.

Answer: C

NEW QUESTION 4

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 5

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 6

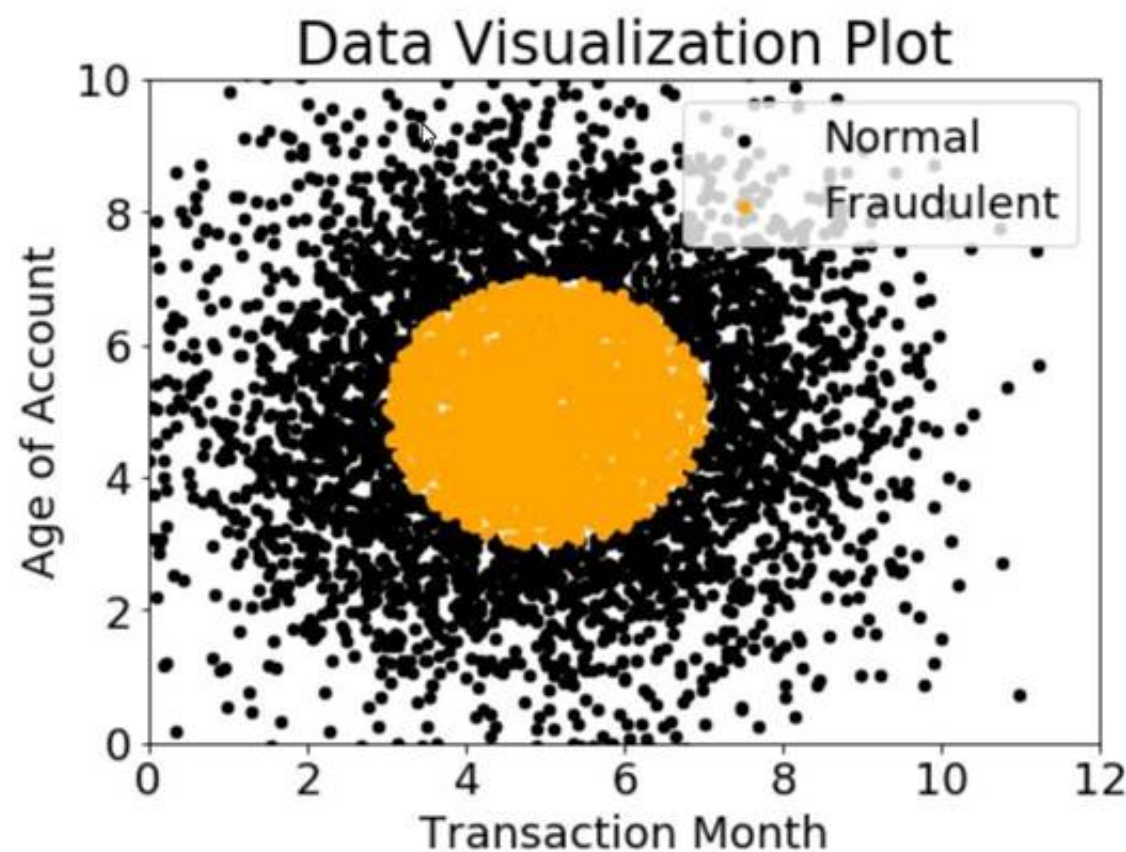
A Machine Learning Specialist deployed a model that provides product recommendations on a company's website Initially, the model was performing very well and resulted in customers buying more products on average However within the past few months the Specialist has noticed that the effect of product 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 7

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.



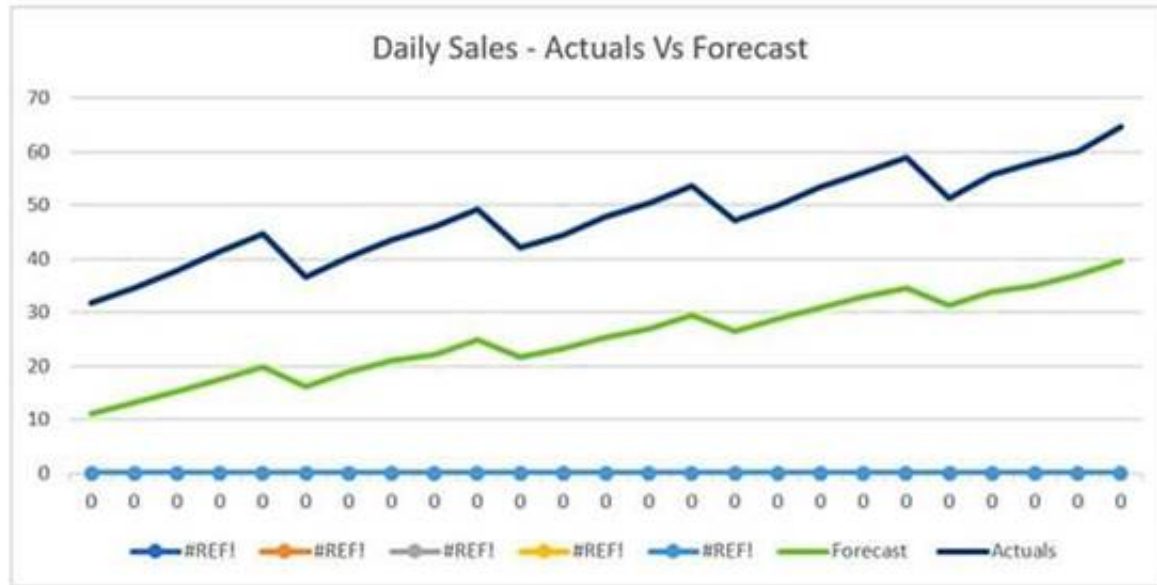
Based on this information which model would have the HIGHEST accuracy?

- A. Long short-term memory (LSTM) model with scaled exponential linear unit (SELL))
- B. Logistic regression
- C. Support vector machine (SVM) with non-linear kernel
- D. Single perceptron with tanh activation function

Answer: B

NEW QUESTION 8

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 9

A Machine Learning Specialist is working for a credit card processing company and receives an unbalanced dataset containing credit card transactions. It contains 99,000 valid transactions and 1,000 fraudulent transactions. The Specialist is asked to score a model that was run against the dataset. The Specialist has been advised that identifying valid transactions is equally as important as identifying fraudulent transactions. What metric is BEST suited to score the model?

- A. Precision
- B. Recall
- C. Area Under the ROC Curve (AUC)
- D. Root Mean Square Error (RMSE)

Answer: A

NEW QUESTION 10

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 catalog; 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 10

A Machine Learning Specialist is working with a large company to leverage machine learning within its products. The company wants to group its customers into categories based on which customers will and will not churn within the next 6 months. The company has labeled the data available to the Specialist. Which machine learning model type should the Specialist use to accomplish this task?

- A. Linear regression
- B. Classification
- C. Clustering
- D. Reinforcement learning

Answer: B

Explanation:

The goal of classification is to determine to which class or category a data point (customer in our case) belongs to. For classification problems, data scientists would use historical data with predefined target variables AKA labels (churner/non-churner) – answers that need to be predicted – to train an algorithm. With classification, businesses can answer the following questions:

- > Will this customer churn or not?
- > Will a customer renew their subscription?
- > Will a user downgrade a pricing plan?
- > Are there any signs of unusual customer behavior?

NEW QUESTION 12

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 15

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 16

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 19

A company is running a machine learning prediction service that generates 100 TB of predictions every day. A Machine Learning Specialist must generate a visualization of the daily precision-recall curve from the predictions, and forward a read-only version to the Business team.

Which solution requires the LEAST coding effort?

- A. Run a daily Amazon EMR workflow to generate precision-recall data, and save the results in Amazon S3. Give the Business team read-only access to S3.
- B. Generate daily precision-recall data in Amazon QuickSight, and publish the results in a dashboard shared with the Business team.
- C. Run a daily Amazon EMR workflow to generate precision-recall data, and save the results in Amazon S3. Visualize the arrays in Amazon QuickSight, and publish them in a dashboard shared with the Business team.
- D. Generate daily precision-recall data in Amazon ES, and publish the results in a dashboard shared with the Business team.

Answer: C

NEW QUESTION 22

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 25

An office security agency conducted a successful pilot using 100 cameras installed at key locations within the main office. Images from the cameras were uploaded to Amazon S3 and tagged using Amazon Rekognition, and the results were stored in Amazon ES. The agency is now looking to expand the pilot into a full production system using thousands of video cameras in its office locations globally. The goal is to identify activities performed by non-employees in real time. Which solution should the agency consider?

- A. Use a proxy server at each local office and for each camera, and stream the RTSP feed to a unique Amazon Kinesis Video Streams video stream
- B. On each stream, use Amazon Rekognition Video and create a stream processor to detect faces from a collection of known employees, and alert when non-employees are detected.
- C. Use a proxy server at each local office and for each camera, and stream the RTSP feed to a unique Amazon Kinesis Video Streams video stream
- D. On each stream, use Amazon Rekognition Image to detect faces from a collection of known employees and alert when non-employees are detected.
- E. Install AWS DeepLens cameras and use the DeepLens_Kinesis_Video module to stream video to Amazon Kinesis Video Streams for each camera
- F. On each stream, use Amazon Rekognition Video and create a stream processor to detect faces from a collection on each stream, and alert when non-employees are detected.
- G. Install AWS DeepLens cameras and use the DeepLens_Kinesis_Video module to stream video to Amazon Kinesis Video Streams for each camera
- H. On each stream, run an AWS Lambda function to capture image fragments and then call Amazon Rekognition Image to detect faces from a collection of known employees, and alert when non-employees are detected.

Answer: D

NEW QUESTION 28

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 30

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 33

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 to 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 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 Firehose stream that transforms raw record attributes into simple transformed values using SQL.

Answer: D

NEW QUESTION 37

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 38

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 39

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 use?

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

Answer: D

NEW QUESTION 40

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 44

A company has raw user and transaction data stored in Amazon S3, 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 46

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	1 34
	1 877	1 123

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 eval_metric 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 48

A Machine Learning Specialist is working with multiple data sources containing billions of records that need to be joined. What feature engineering and model development approach should the Specialist take with a dataset this large?

- A. Use an Amazon SageMaker notebook for both feature engineering and model development
- B. Use an Amazon SageMaker notebook for feature engineering and Amazon ML for model development
- C. Use Amazon EMR for feature engineering and Amazon SageMaker SDK for model development
- D. Use Amazon ML for both feature engineering and model development.

Answer: B

NEW QUESTION 52

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: D

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 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 QuickSight 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 to Amazon ES and use Kibana to query and visualize the log data.

Answer: B

NEW QUESTION 60

A Machine Learning Specialist is using Amazon SageMaker to host a model for a highly available customer-facing application .
The Specialist has trained a new version of the model, validated it with historical data, and now wants to deploy it to production To limit any risk of a negative customer experience, the Specialist wants to be able to monitor the model and roll it back, if needed
What is the SIMPLEST approach with the LEAST risk to deploy the model and roll it back, if needed?

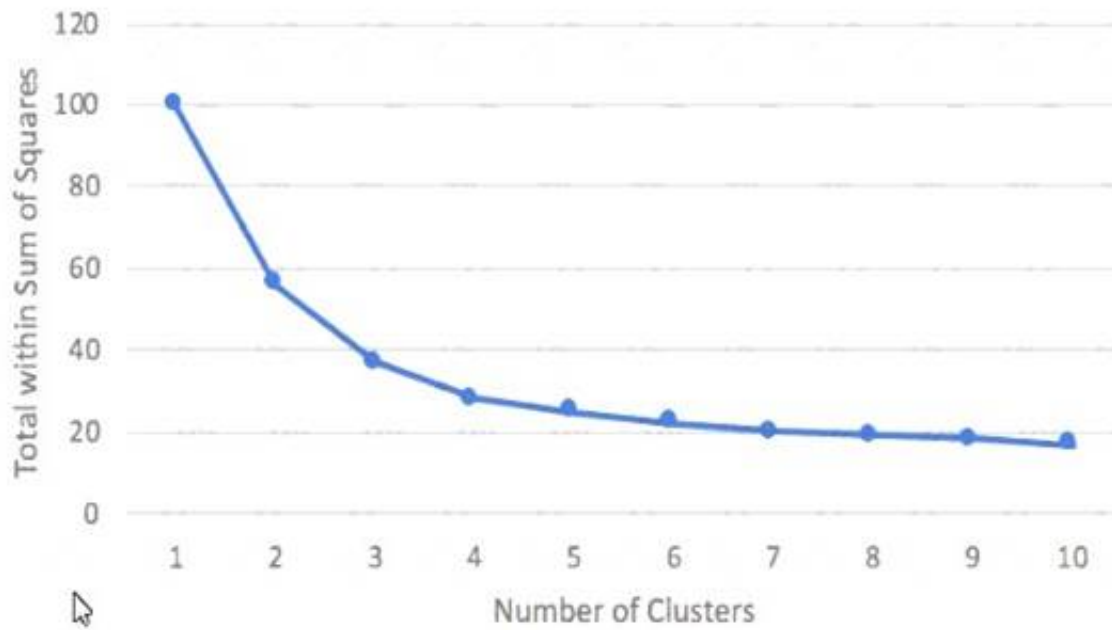
- A. Create a SageMaker endpoint and configuration for the new model versio
- B. Redirect production traffic to the new endpoint by updating the client configuratio
- C. Revert traffic to the last version if the model does not perform as expected.
- D. Create a SageMaker endpoint and configuration for the new model versio
- E. Redirect production traffic to the new endpoint by using a load balancer Revert traffic to the last version if the model does not perform as expected.
- F. Update the existing SageMaker endpoint to use a new configuration that is weighted to send 5% of the traffic to the new varian
- G. Revert traffic to the last version by resetting the weights if the model does not perform as expected.
- H. Update the existing SageMaker endpoint to use a new configuration that is weighted to send 100% of the traffic to the new variant Revert traffic to the last version by resetting the weights if the model does not perform as expected.

Answer: A

NEW QUESTION 65

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

Optimal Number of Clusters



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 69

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 73

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 75

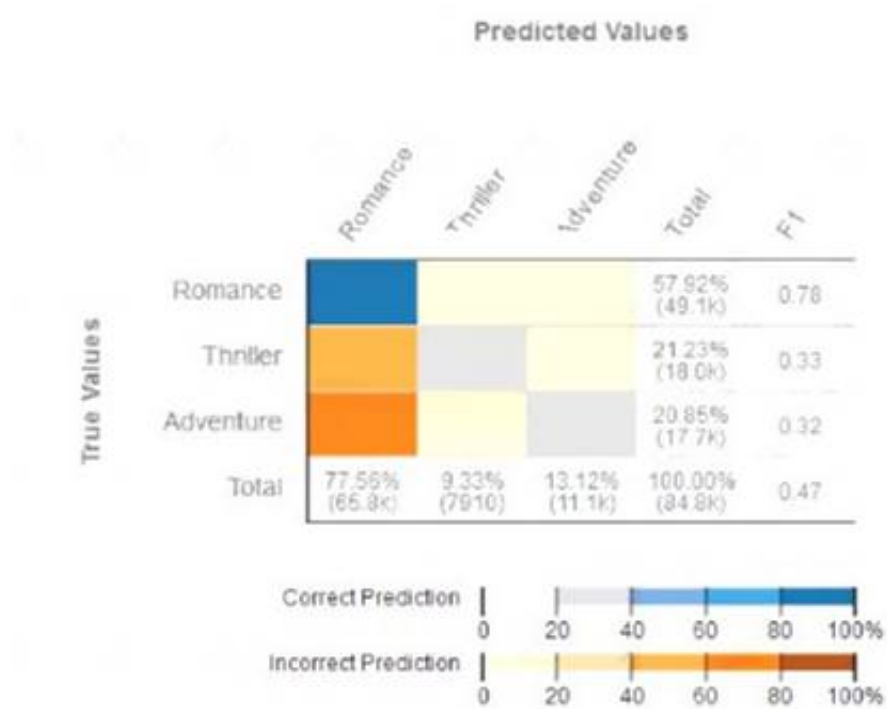
An e-commerce company needs a customized training model to classify images of its shirts and pants products. The company needs a proof of concept in 2 to 3 days with good accuracy. Which compute choice should the Machine Learning Specialist select to train and achieve good accuracy on the model quickly?

- 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 80

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 13.12%
 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 81

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 (t-SNE) to visualize the large number of input variables in an easier-to-read dimension.
 C. A scatter plot showing the 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 85

A web-based company wants to improve its conversion rate on its landing page. Using a large historical dataset of customer visits, the company has repeatedly trained a multi-class deep learning network algorithm on Amazon SageMaker. However, there is an overfitting problem: training data shows 90% accuracy in predictions, while test data shows 70% accuracy only.

The company needs to boost the generalization of its model before deploying it into production to maximize conversions of visits to purchases.

Which action is recommended to provide the HIGHEST accuracy model for the company's test and validation data?

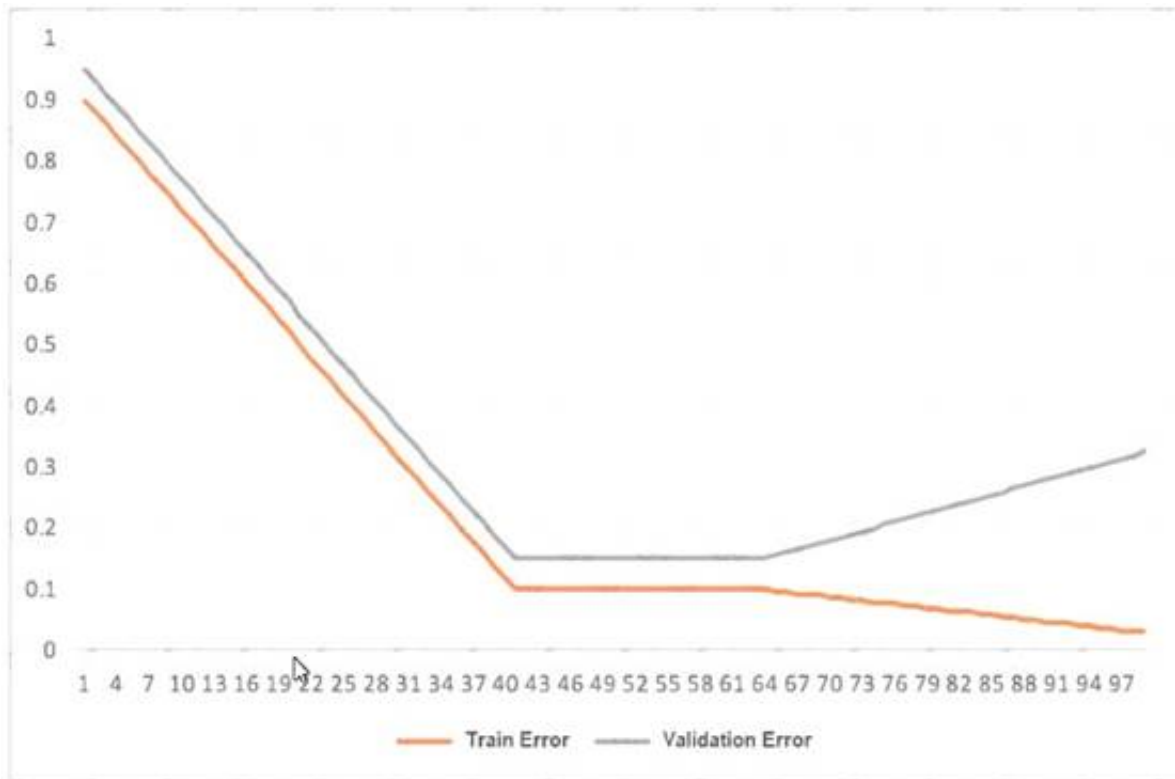
- A. Increase the randomization of training data in the mini-batches used in training.
 B. Allocate a higher proportion of the overall data to the training dataset.
 C. Apply L1 or L2 regularization and dropouts to the training.
 D. Reduce the number of layers and units (or neurons) from the deep learning network.

Answer: A

NEW QUESTION 88

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 90

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 format
- 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 database
- 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 92

A Machine Learning Specialist needs to create a data repository to hold a large amount of time-based training data for a new model. In the source system, new files are added every hour. Throughout a single 24-hour period, the volume of hourly updates will change significantly. The Specialist always wants to train on the last 24 hours of the data.

Which type of data repository is the MOST cost-effective solution?

- A. An Amazon EBS-backed Amazon EC2 instance with hourly directories
- B. An Amazon RDS database with hourly table partitions
- C. An Amazon S3 data lake with hourly object prefixes
- D. An Amazon EMR cluster with hourly hive partitions on Amazon EBS volumes

Answer: C

NEW QUESTION 94

A bank's Machine Learning team is developing an approach for credit card fraud detection. The company has a large dataset of historical data labeled as fraudulent. The goal is to build a model to take the information from new transactions and predict whether each transaction is fraudulent or not.

Which built-in Amazon SageMaker machine learning algorithm should be used for modeling this problem?

- A. Seq2seq
- B. XGBoost
- C. K-means
- D. Random Cut Forest (RCF)

Answer: C

NEW QUESTION 96

A manufacturing company has a large set of labeled historical sales data. The manufacturer would like to predict how many units of a particular part should be produced each quarter. Which machine learning approach should be used to solve this problem?

- A. Logistic regression
- B. Random Cut Forest (RCF)
- C. Principal component analysis (PCA)
- D. Linear regression

Answer: B

NEW QUESTION 101

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 104

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: D

NEW QUESTION 105

A Machine Learning Specialist has built a model using Amazon SageMaker built-in algorithms and is not getting expected accurate results. The Specialist wants to use hyperparameter optimization to increase the model's accuracy. Which method is the MOST repeatable and requires the LEAST amount of effort to achieve this?

- A. Launch multiple training jobs in parallel with different hyperparameters.
- B. Create an AWS Step Functions workflow that monitors the accuracy in Amazon CloudWatch Logs and relaunches the training job with a defined list of hyperparameters.
- C. Create a hyperparameter tuning job and set the accuracy as an objective metric.
- D. Create a random walk in the parameter space to iterate through a range of values that should be used for each individual hyperparameter.

Answer: B

NEW QUESTION 107

A Data Scientist is developing a machine learning model to predict future patient outcomes based on information collected about each patient and their treatment plans. The model should output a continuous value as its prediction. The data available includes labeled outcomes for a set of 4,000 patients. The study was conducted on a group of individuals over the age of 65 who have a particular disease that is known to worsen with age.

Initial models have performed poorly. While reviewing the underlying data, the Data Scientist notices that, out of 4,000 patient observations, there are 450 where the patient age has been input as 0. The other features for these observations appear normal compared to the rest of the sample population.

How should the Data Scientist correct this issue?

- A. Drop all records from the dataset where age has been set to 0.
- B. Replace the age field value for records with a value of 0 with the mean or median value from the dataset.
- C. Drop the age feature from the dataset and train the model using the rest of the features.
- D. Use k-means clustering to handle missing features.

Answer: A

NEW QUESTION 112

A Machine Learning Specialist is assigned a TensorFlow project using Amazon SageMaker for training, and needs to continue working for an extended period with no Wi-Fi access.

Which approach should the Specialist use to continue working?

- A. Install Python 3 and boto3 on their laptop and continue the code development using that environment.
- B. Download the TensorFlow Docker container used in Amazon SageMaker from GitHub to their local environment, and use the Amazon SageMaker Python SDK to test the code.
- C. Download TensorFlow from tensorflow.org to emulate the TensorFlow kernel in the SageMaker environment.
- D. Download the SageMaker notebook to their local environment then install Jupyter Notebooks on their laptop and continue the development in a local notebook.

Answer: A

NEW QUESTION 114

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 117

.....

Thank You for Trying Our Product

We offer two products:

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

2nd - Questions 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](#)