

MLS-C01 Dumps

AWS Certified Machine Learning - Specialty

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



NEW QUESTION 1

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 2

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 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 3

A Machine Learning team uses Amazon SageMaker to train an Apache MXNet handwritten digit classifier model using a research dataset. The team wants to receive a notification when the model is overfitting. Auditors want to view the Amazon SageMaker log activity report to ensure there are no unauthorized API calls. What should the Machine Learning team do to address the requirements with the least amount of code and fewest steps?

- A. Implement an AWS Lambda function to log Amazon SageMaker API calls to Amazon S3. Add code to push a custom metric to Amazon CloudWatch
- B. Create an alarm in CloudWatch with Amazon SNS to receive a notification when the model is overfitting.
- C. Use AWS CloudTrail to log Amazon SageMaker API calls to Amazon S3. Add code to push a custom metric to Amazon CloudWatch
- D. Create an alarm in CloudWatch with Amazon SNS to receive a notification when the model is overfitting.
- E. Implement an AWS Lambda function to log Amazon SageMaker API calls to AWS CloudTrail
- F. Add code to push a custom metric to Amazon CloudWatch
- G. Create an alarm in CloudWatch with Amazon SNS to receive a notification when the model is overfitting.
- H. Use AWS CloudTrail to log Amazon SageMaker API calls to Amazon S3. Set up Amazon SNS to receive a notification when the model is overfitting.

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 Machine Learning Specialist is developing recommendation engine for a photography blog Given a picture, the recommendation engine should show a picture that captures similar objects The Specialist would like to create a numerical representation feature to perform nearest-neighbor searches

What actions would allow the Specialist to get relevant numerical representations?

- A. Reduce image resolution and use reduced resolution pixel values as features

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

Answer: A

NEW QUESTION 6

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 7

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 8

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 9

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 10

A large mobile network operating company is building a machine learning model to predict customers who are likely to unsubscribe from the service. The company plans to offer an incentive for these customers as the cost of churn is far greater than the cost of the incentive.

The model produces the following confusion matrix after evaluating on a test dataset of 100 customers: Based on the model evaluation results, why is this a viable model for production?

n = 100	PREDICTED CHURN	
	Yes	No
ACTUAL Churn Yes	10	4
Actual No	10	76

- A. The model is 86% accurate and the cost incurred by the company as a result of false negatives is less than the false positives.
- B. The precision of the model is 86%, which is less than the accuracy of the model.
- C. The model is 86% accurate and the cost incurred by the company as a result of false positives is less than the false negatives.
- D. The precision of the model is 86%, which is greater than the accuracy of the model.

Answer: B

NEW QUESTION 10

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 11

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 16

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 18

A Machine Learning Specialist is building a prediction model for a large number of features using linear models, such as linear regression and logistic regression During exploratory data analysis the Specialist observes that many features are highly correlated with each other This may make the model unstable What should be done to reduce the impact of having such a large number of features?

- A. Perform one-hot encoding on highly correlated features
- B. Use matrix multiplication on highly correlated features.
- C. Create a new feature space using principal component analysis (PCA)
- D. Apply the Pearson correlation coefficient

Answer: C

NEW QUESTION 21

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 25

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 30

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 35

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

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 39

When submitting Amazon SageMaker training jobs using one of the built-in algorithms, which common parameters **MUST** be specified? (Select THREE.)

- A. The training channel identifying the location of training data on an Amazon S3 bucket.
- B. The validation channel identifying the location of validation data on an Amazon S3 bucket.
- C. The IAM role that Amazon SageMaker can assume to perform tasks on behalf of the users.
- D. Hyperparameters in a JSON array as documented for the algorithm used.
- E. The Amazon EC2 instance class specifying whether training will be run using CPU or GPU.
- F. The output path specifying where on an Amazon S3 bucket the trained model will persist.

Answer: AEF

NEW QUESTION 44

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 45

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 47

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 48

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 52

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 these 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 55

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 56

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 version.
- B. Redirect production traffic to the new endpoint by updating the client configuration.
- 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 version.
- 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 variant.
- 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 60

A manufacturing company asks its Machine Learning Specialist to develop a model that classifies defective parts into one of eight defect types. The company has provided roughly 100000 images per defect type for training. During the initial training of the image classification model, the Specialist notices that the validation accuracy is 80%, while the training accuracy is 90%. It is known that human-level performance for this type of image classification is around 90%. What should the Specialist consider to fix this issue?

- A. A longer training time
- B. Making the network larger
- C. Using a different optimizer
- D. Using some form of regularization

Answer: D

NEW QUESTION 64

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 65

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?

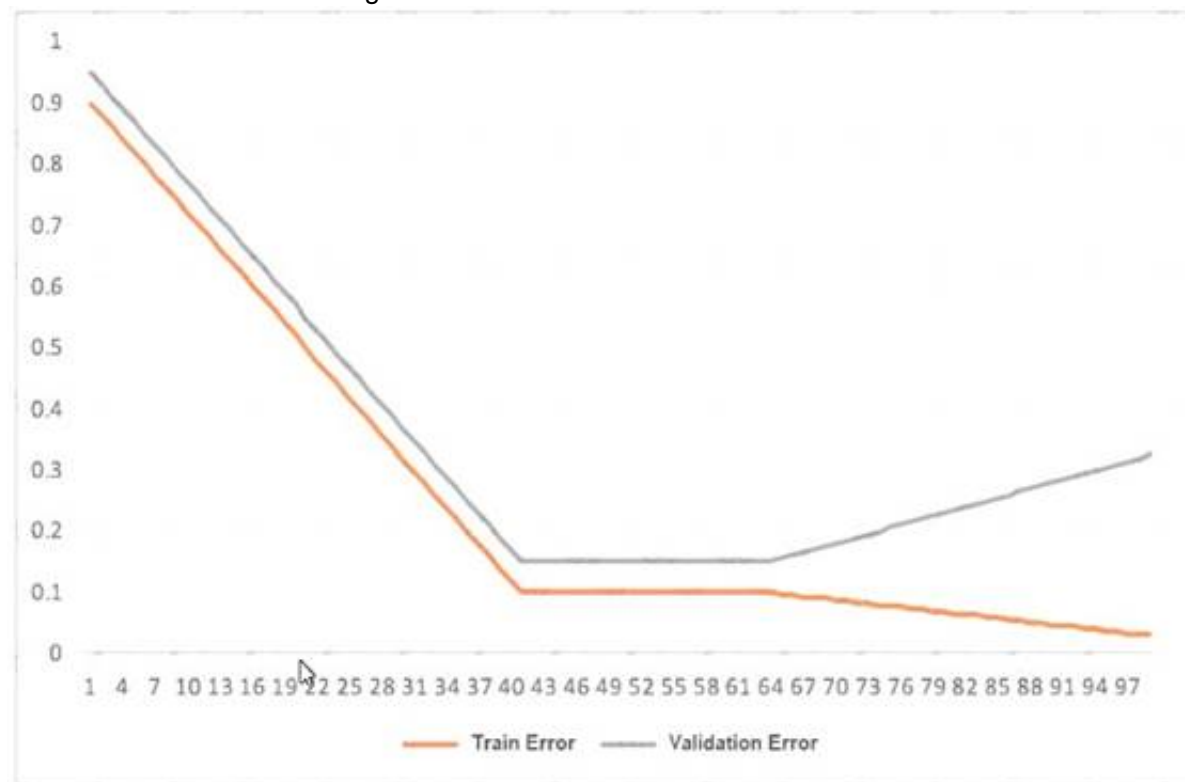
- A. . m5 4xlarge (general purpose)
- B. r5.2xlarge (memory optimized)
- C. p3.2xlarge (GPU accelerated computing)
- D. p3 8xlarge (GPU accelerated computing)

Answer: C

NEW QUESTION 70

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 73

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 78

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 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 85

A Machine Learning Specialist is working with a large cybersecurity company that manages security events in real time for companies around the world. The cybersecurity company wants to design a solution that will allow it to use machine learning to score malicious events as anomalies on the data as it is being ingested. The company also wants to be able to save the results in its data lake for later processing and analysis.

What is the MOST efficient way to accomplish these tasks?

- A. Ingest the data using Amazon Kinesis Data Firehose, and use Amazon Kinesis Data Analytics Random Cut Forest (RCF) for anomaly detection. Then use Kinesis Data Firehose to stream the results to Amazon S3.
- B. Ingest the data into Apache Spark Streaming using Amazon EMR.
- C. and use Spark MLlib with k-means to perform anomaly detection. Then store the results in an Apache Hadoop Distributed File System (HDFS) using Amazon EMR with a replication factor of three as the data lake.
- D. Ingest the data and store it in Amazon S3. Use AWS Batch along with the AWS Deep Learning AMIs to train a k-means model using TensorFlow on the data in Amazon S3.
- E. Ingest the data and store it in Amazon S3. Have an AWS Glue job that is triggered on demand transform the new data. Then use the built-in Random Cut Forest (RCF) model within Amazon SageMaker to detect anomalies in the data.

Answer: B

NEW QUESTION 87

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 91

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 94

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 96

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 101

.....

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>