MLS-C01^{Q&As}

AWS Certified Machine Learning - Specialty (MLS-C01)

Pass Amazon MLS-C01 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.leads4pass.com/aws-certified-machine-learning-specialty.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Amazon Official Exam Center

⚙ **Instant Download** After Purchase

⚙ **100% Money Back** Guarantee

⚙ **365 Days** Free Update

⚙ **800,000+** Satisfied Customers

QUESTION 1

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)

Correct Answer: A

QUESTION 2

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

Correct Answer: D

https://docs.aws.amazon.com/zh_tw/machine-learning/latest/dg/regression-model-insights.html

QUESTION 3

A company is training machine learning (ML) models on Amazon SageMaker by using 200 TB of data that is stored in Amazon S3 buckets. The training data consists of individual files that are each larger than 200 MB in size. The company needs a data access solution that offers the shortest processing time and the least amount of setup.

Which solution will meet these requirements?

A. Use File mode in SageMaker to copy the dataset from the S3 buckets to the ML instance storage.

B. Create an Amazon FSx for Lustre file system. Link the file system to the S3 buckets.

C. Create an Amazon Elastic File System (Amazon EFS) file system. Mount the file system to the training instances.

D. Use FastFile mode in SageMaker to stream the files on demand from the S3 buckets.

Correct Answer: D

For larger datasets with larger files (more than 50 MB per file), the first option is to try fast file mode, which is more straightforward to use than FSx for Lustre because it doesn\'t require creating a file system, or connecting to a VPC. Fast file mode is ideal for large file containers (more than 150 MB), and might also do well with files more than 50 MB.

https://docs.aws.amazon.com/sagemaker/latest/dg/model-access-training-data.html#model-access-training-data-best-practices

---

QUESTION 4

IT leadership wants Jo transition a company\\\'s existing machine learning data storage environment to AWS as a temporary ad hoc solution The company currently uses a custom software process that heavily leverages SOL as a query language and exclusively stores generated csv documents for machine learning

The ideal state for the company would be a solution that allows it to continue to use the current workforce of SQL experts The solution must also support the storage of csv and JSON files, and be able to query over semi-structured data The following are high priorities for the company:

1.

 Solution simplicity

2.

 Fast development time

3.

 Low cost

4.

 High flexibility

What technologies meet the company\\\'s requirements?

A. Amazon S3 and Amazon Athena

B. Amazon Redshift and AWS Glue

C. Amazon DynamoDB and DynamoDB Accelerator (DAX)

D. Amazon RDS and Amazon ES

Correct Answer: B

QUESTION 5

A finance company needs to forecast the price of a commodity. The company has compiled a dataset of historical daily prices. A data scientist must train various forecasting models on 80% of the dataset and must validate the efficacy of those models on the remaining 20% of the dataset.

How should the data scientist split the dataset into a training dataset and a validation dataset to compare model performance?

A. Pick a date so that 80% of the data points precede the date. Assign that group of data points as the training dataset. Assign all the remaining data points to the validation dataset.

B. Pick a date so that 80% of the data points occur after the date. Assign that group of data points as the training dataset. Assign all the remaining data points to the validation dataset.

C. Starting from the earliest date in the dataset, pick eight data points for the training dataset and two data points for the validation dataset. Repeat this stratified sampling until no data points remain.

D. Sample data points randomly without replacement so that 80% of the data points are in the training dataset. Assign all the remaining data points to the validation dataset.

Correct Answer: A

QUESTION 6

A manufacturer is operating a large number of factories with a complex supply chain relationship where unexpected downtime of a machine can cause production to stop at several factories. A data scientist wants to analyze sensor data from the factories to identify equipment in need of preemptive maintenance and then dispatch a service team to prevent unplanned downtime. The sensor readings from a single machine can include up to 200 data points including temperatures, voltages, vibrations, RPMs, and pressure readings.

To collect this sensor data, the manufacturer deployed Wi-Fi and LANs across the factories. Even though many factory locations do not have reliable or high-speed internet connectivity, the manufacturer would like to maintain near-real-time inference capabilities.

Which deployment architecture for the model will address these business requirements?

A. Deploy the model in Amazon SageMaker. Run sensor data through this model to predict which machines need maintenance.

B. Deploy the model on AWS IoT Greengrass in each factory. Run sensor data through this model to infer which machines need maintenance.

C. Deploy the model to an Amazon SageMaker batch transformation job. Generate inferences in a daily batch report to identify machines that need maintenance.

D. Deploy the model in Amazon SageMaker and use an IoT rule to write data to an Amazon DynamoDB table. Consume a DynamoDB stream from the table with an AWS Lambda function to invoke the endpoint.

Correct Answer: B

https://aws.amazon.com/blogs/iot/using-aws-iot-for-predictive-maintenance/ https://aws.amazon.com/blogs/iot/industrial-iot-from-condition-based-monitoring-to-predictive-quality-to-digitize-your-factory-with-aws-iot-services/

QUESTION 7

A machine learning (ML) specialist needs to solve a binary classification problem for a marketing dataset. The ML specialist must maximize the Area Under the ROC Curve (AUC) of the algorithm by training an XGBoost algorithm. The ML specialist must find values for the eta, alpha, min_child_weight, and max_depth hyperparameters that will generate

the most accurate model.

Which approach will meet these requirements with the LEAST operational overhead?

A. Use a bootstrap script to install scikit-learn on an Amazon EMR cluster. Deploy the EMR cluster. Apply k-fold cross-validation methods to the algorithm.

B. Deploy Amazon SageMaker prebuilt Docker images that have scikit-learn installed. Apply k-fold cross-validation methods to the algorithm.

C. Use Amazon SageMaker automatic model tuning (AMT). Specify a range of values for each hyperparameter.

D. Subscribe to an AUC algorithm that is on AWS Marketplace. Specify a range of values for each hyperparameter.

Correct Answer: C

QUESTION 8

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

Which solution should a machine learning specialist implement to meet these requirements?

A. Install cameras compatible with Amazon Kinesis Video Streams to stream the data to AWS over the restaurant\\'s existing internet connection. Write an AWS Lambda function to take an image and send it to Amazon Rekognition to count the number of faces in the image. Send an Amazon Simple Notification Service (Amazon SNS) notification if the line is too long.

B. Deploy AWS DeepLens cameras in the restaurant to capture video. Enable Amazon Rekognition on the AWS DeepLens device, and use it to trigger a local AWS Lambda function when a person is recognized. Use the Lambda function to send an Amazon Simple Notification Service (Amazon SNS) notification if the line is too long.

C. Build a custom model in Amazon SageMaker to recognize the number of people in an image. Install cameras compatible with Amazon Kinesis Video Streams in the restaurant. Write an AWS Lambda function to take an image. Use the SageMaker endpoint to call the model to count people. Send an Amazon Simple Notification Service (Amazon SNS) notification if the line is too long.

D. Build a custom model in Amazon SageMaker to recognize the number of people in an image. Deploy AWS DeepLens cameras in the restaurant. Deploy the model to the cameras. Deploy an AWS Lambda function to the cameras to use the model to count people and send an Amazon Simple Notification Service (Amazon SNS) notification if the line is too long.

Correct Answer: D

QUESTION 9

A company wants to use machine learning (ML) to improve its customer churn prediction model. The company stores data in an Amazon Redshift data warehouse.

A data science team wants to use Amazon Redshift machine learning (Amazon Redshift ML) to build a model and run predictions for new data directly within the data warehouse.

Which combination of steps should the company take to use Amazon Redshift ML to meet these requirements? (Choose three.)

A. Define the feature variables and target variable for the churn prediction model.

B. Use the SOL EXPLAIN_MODEL function to run predictions.

C. Write a CREATE MODEL SQL statement to create a model.

D. Use Amazon Redshift Spectrum to train the model.

E. Manually export the training data to Amazon S3.

F. Use the SQL prediction function to run predictions.

Correct Answer: ACF

QUESTION 10

A chemical company has developed several machine learning (ML) solutions to identify chemical process abnormalities. The time series values of independent variables and the labels are available for the past 2 years and are sufficient to accurately model the problem.

The regular operation label is marked as 0. The abnormal operation label is marked as 1 . Process abnormalities have a significant negative effect on the companys profits. The company must avoid these abnormalities.

Which metrics will indicate an ML solution that will provide the GREATEST probability of detecting an abnormality?

A. Precision = 0.91 Recall = 0.6

B. Precision = 0.61 Recall = 0.98

C. Precision = 0.7 Recall = 0.9

D. Precision = 0.98 Recall = 0.8

Correct Answer: B

The metrics that will indicate an ML solution that will provide the greatest probability of detecting an abnormality are precision and recall. Precision is the ratio of true positives (TP) to the total number of predicted positives (TP + FP), where FP is false positives. Recall is the ratio of true positives (TP) to the total number of actual positives (TP + FN), where FN is false negatives. A high precision means that the ML solution has a low rate of false alarms, while a high recall means that the ML solution has a high rate of true detections. For the chemical company, the goal is to avoid process abnormalities, which are marked as 1 in the labels. Therefore, the company needs an ML solution that has a high recall for the positive class, meaning that it can detect most of the abnormalities and minimize the false negatives. Among the four options, option B has the highest recall for the positive class, which is 0.98. This means that the ML solution can detect 98% of the abnormalities and miss only 2%. Option B also has a reasonable precision for the positive class, which is 0.61. This means that the ML solution has a false alarm rate of 39%, which may be acceptable for the company, depending on the cost and benefit analysis. The other options have lower recall for the positive class, which means that they have higher false negative rates, which can be more detrimental for the company than false positive rates. References:

1: AWS Certified Machine Learning - Specialty uide

2: AWS Training - Machine Learning on AWS

3: AWS Whitepaper - An Overview of Machine Learning on AWS

4: Precision and recall

QUESTION 11

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

Correct Answer: B

In transfer learning, a pre-trained model is used as a starting point to train a new model on a different task, typically using a smaller dataset. The pre-trained model contains weights that have been learned from a large amount of data on a related task, and these weights can be leveraged to train the new model more efficiently.

To re-train the model with the custom data, the Specialist should initialize the model with pre-trained weights in all layers, as these weights can provide a good starting point for the new task. The Specialist should then replace the last fully connected layer, which is responsible for making the final predictions, as this layer will likely need to be modified to reflect the new task. By keeping the pre-trained weights in the other layers, the Specialist can take advantage of the knowledge learned from the previous task, and potentially speed up the training process.

QUESTION 12

A machine learning (ML) specialist at a manufacturing company uses Amazon SageMaker DeepAR to forecast input materials and energy requirements for the company. Most of the data in the training dataset is missing values for the target

variable. The company stores the training dataset as JSON files.

The ML specialist develop a solution by using Amazon SageMaker DeepAR to account for the missing values in the training dataset.

Which approach will meet these requirements with the LEAST development effort?

A. Impute the missing values by using the linear regression method. Use the entire dataset and the imputed values to train the DeepAR model.

B. Replace the missing values with not a number (NaN). Use the entire dataset and the encoded missing values to train the DeepAR model.

C. Impute the missing values by using a forward fill. Use the entire dataset and the imputed values to train the DeepAR model.

D. Impute the missing values by using the mean value. Use the entire dataset and the imputed values to train the

DeepAR model.

Correct Answer: D

QUESTION 13

The chief editor for a product catalog wants the research and development team to build a machine learning system that can be used to detect whether or not individuals in a collection of images are wearing the company\\'s retail brand. The team has a set of training data.

Which machine learning algorithm should the researchers use that BEST meets their requirements?

A. Latent Dirichlet Allocation (LDA)

B. Recurrent neural network (RNN)

C. K-means

D. Convolutional neural network (CNN)

Correct Answer: D

QUESTION 14

A Machine Learning Specialist is working for an online retailer that wants to run analytics on every customer visit, processed through a machine learning pipeline. The data needs to be ingested by Amazon Kinesis Data Streams at up to 100 transactions per second, and the JSON data blob is 100 KB in size.

What is the MINIMUM number of shards in Kinesis Data Streams the Specialist should use to successfully ingest this data?

A. 1 shards

B. 10 shards

C. 100 shards

D. 1,000 shards

Correct Answer: B

1 shard can Ingest 1 MB/second or 1,000 records/second. So 100 KB * 100 = 10 MB (10 shards required)

QUESTION 15

A music streaming company is building a pipeline to extract features. The company wants to store the features for offline model training and online inference. The company wants to track feature history and to give the company\\'s data science teams access to the features.

Which solution will meet these requirements with the MOST operational efficiency?

A. Use Amazon SageMaker Feature Store to store features for model training and inference. Create an online store for online inference. Create an offline store for model training. Create an IAM role for data scientists to access and search through feature groups.

B. Use Amazon SageMaker Feature Store to store features for model training and inference. Create an online store for both online inference and model training. Create an IAM role for data scientists to access and search through feature groups.

C. Create one Amazon S3 bucket to store online inference features. Create a second S3 bucket to store offline model training features. Turn on versioning for the S3 buckets and use tags to specify which tags are for online inference features and which are for offline model training features. Use Amazon Athena to query the S3 bucket for online inference. Connect the S3 bucket for offline model training to a SageMaker training job. Create an IAM policy that allows data scientists to access both buckets.

D. Create two separate Amazon DynamoDB tables to store online inference features and offline model training features. Use time-based versioning on both tables. Query the DynamoDB table for online inference. Move the data from DynamoDB to Amazon S3 when a new SageMaker training job is launched. Create an IAM policy that allows data scientists to access both tables.

Correct Answer: B

[MLS-C01 PDF Dumps](#)   [MLS-C01 Practice Test](#)   [MLS-C01 Exam Questions](#)