DUMPSBOSS.COM

AWS Certified Machine Learning - Specialty (MLS-C01)

Amazon AWS AWS-Certified-Machine-Learning-Specialty-MLS-C01

Version Demo

Total Demo Questions: 15

Total Premium Questions: 377

Buy Premium PDF

https://dumpsboss.com

support@dumpsboss.com

dumpsboss.com

QUESTION NO: 1

A data engineer at a bank is evaluating a new tabular dataset that includes customer data. The data engineer will use the customer data to create a new model to predict customer behavior. After creating a correlation matrix for the variables, the data engineer notices that many of the 100 features are highly correlated with each other.

Which steps should the data engineer take to address this issue? (Choose two.)

A. Use a linear-based algorithm to train the model.

B. Apply principal component analysis (PCA).

C. Remove a portion of highly correlated features from the dataset.

D. Apply min-max feature scaling to the dataset.

E. Apply one-hot encoding category-based variables.

ANSWER: B D

QUESTION NO: 2

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

Explanation:

Reference: https://docs.aws.amazon.com/sagemaker/latest/dg/gs-setup-working-env.html

QUESTION NO: 3

A Data Scientist is building a model to predict customer churn using a dataset of 100 continuous numerical features. The Marketing team has not provided any insight about which features are relevant for churn prediction. The Marketing team wants to interpret the model and see the direct impact of relevant features on the model outcome. While training a logistic regression model, the Data Scientist observes that there is a wide gap between the training and validation set accuracy.

Which methods can the Data Scientist use to improve the model performance and satisfy the Marketing team's needs? (Choose two.)

A. Add L1 regularization to the classifier

B. Add features to the dataset

C. Perform recursive feature elimination

D. Perform t-distributed stochastic neighbor embedding (t-SNE)

E. Perform linear discriminant analysis

ANSWER: B E

QUESTION NO: 4

A retail company is selling products through a global online marketplace. The company wants to use machine learning (ML) to analyze customer feedback and identify specific areas for improvement. A developer has built a tool that collects customer reviews from the online marketplace and stores them in an Amazon S3 bucket. This process yields a dataset of 40 reviews. A data scientist building the ML models must identify additional sources of data to increase the size of the dataset.

Which data sources should the data scientist use to augment the dataset of reviews? (Choose three.)

A. Emails exchanged by customers and the company's customer service agents

B. Social media posts containing the name of the company or its products

C. A publicly available collection of news articles

D. A publicly available collection of customer reviews

E. Product sales revenue figures for the company

F. Instruction manuals for the company's products

ANSWER: B D F

QUESTION NO: 5

A company is converting a large number of unstructured paper receipts into images. The company wants to create a model based on natural language processing (NLP) to find relevant entities such as date, location, and notes, as well as some custom entities such as receipt numbers.

The company is using optical character recognition (OCR) to extract text for data labeling. However, documents are in different structures and formats, and the company is facing challenges with setting up the manual workflows for each

document type. Additionally, the company trained a named entity recognition (NER) model for custom entity detection using a small sample size. This model has a very low confidence score and will require retraining with a large dataset.

Which solution for text extraction and entity detection will require the LEAST amount of effort?

A. Extract text from receipt images by using Amazon Textract. Use the Amazon SageMaker BlazingText algorithm to train on the text for entities and custom entities.

B. Extract text from receipt images by using a deep learning OCR model from the AWS Marketplace. Use the NER deep learning model to extract entities.

C. Extract text from receipt images by using Amazon Textract. Use Amazon Comprehend for entity detection, and use Amazon Comprehend custom entity recognition for custom entity detection.

D. Extract text from receipt images by using a deep learning OCR model from the AWS Marketplace. Use Amazon Comprehend for entity detection, and use Amazon Comprehend custom entity recognition for custom entity detection.

ANSWER: C

Explanation:

Reference: https://aws.amazon.com/blogs/machine-learning/building-an-nlp-powered-search-index-with-amazon-textract-and-amazon-comprehend/

QUESTION NO: 6

A Machine Learning Specialist is configuring Amazon SageMaker so multiple Data Scientists can access notebooks, train models, and deploy endpoints. To ensure the best operational performance, the Specialist needs to be able to track how often the Scientists are deploying models, GPU and CPU utilization on the deployed SageMaker endpoints, and all errors that are generated when an endpoint is invoked.

Which services are integrated with Amazon SageMaker to track this information? (Choose two.)

A. AWS CloudTrail

B. AWS Health

C. AWS Trusted Advisor

D. Amazon CloudWatch

E. AWS Config

ANSWER: A D

Explanation:

Reference: https://aws.amazon.com/sagemaker/faqs/

QUESTION NO: 7

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.

ANSWER: B

QUESTION NO: 8

A Machine Learning team runs its own training algorithm on Amazon SageMaker. The training algorithm requires external assets. The team needs to submit both its own algorithm code and algorithmspecific parameters to Amazon SageMaker.

What combination of services should the team use to build a custom algorithm in Amazon SageMaker? (Choose two.)

A. AWS Secrets Manager

B. AWS CodeStar

C. Amazon ECR

D. Amazon ECS

E. Amazon S3

ANSWER: C E

QUESTION NO: 9

A logistics company needs a forecast model to predict next month's inventory requirements for a single item in 10 warehouses. A machine learning specialist uses Amazon Forecast to develop a forecast model from 3 years of monthly data. There is no missing data. The specialist selects the DeepAR+ algorithm to train a predictor. The predictor means absolute percentage error (MAPE) is much larger than the MAPE produced by the current human forecasters.

Which changes to the CreatePredictor API call could improve the MAPE? (Choose two.)

A. Set PerformAutoML to true.

B. Set ForecastHorizon to 4.

C. Set ForecastFrequency to W for weekly.

D. Set PerformHPO to true.

E. Set FeaturizationMethodName to filling.

ANSWER: C D

Explanation:

Reference: https://docs.aws.amazon.com/forecast/latest/dg/forecast.dg.pdf

QUESTION NO: 10

A company is building a demand forecasting model based on machine learning (ML). In the development stage, an ML specialist uses an Amazon SageMaker notebook to perform feature engineering during work hours that consumes low amounts of CPU and memory resources. A data engineer uses the same notebook to perform data preprocessing once a day on average that requires very high memory and completes in only 2 hours. The data preprocessing is not configured to use GPU. All the processes are running well on an ml.m5.4xlarge notebook instance.

The company receives an AWS Budgets alert that the billing for this month exceeds the allocated budget.

Which solution will result in the MOST cost savings?

A. Change the notebook instance type to a memory optimized instance with the same vCPU number as the ml.m5.4xlarge instance has. Stop the notebook when it is not in use. Run both data preprocessing and feature engineering development on that instance.

B. Keep the notebook instance type and size the same. Stop the notebook when it is not in use. Run data preprocessing on a P3 instance type with the same memory as the ml.m5.4xlarge instance by using Amazon SageMaker Processing.

C. Change the notebook instance type to a smaller general purpose instance. Stop the notebook when it is not in use. Run data preprocessing on an ml.r5 instance with the same memory size as the ml.m5.4xlarge instance by using Amazon SageMaker Processing.

D. Change the notebook instance type to a smaller general purpose instance. Stop the notebook when it is not in use. Run data preprocessing on an R5 instance with the same memory size as the ml.m5.4xlarge instance by using the Reserved Instance option.

ANSWER: B

QUESTION NO: 11

A media company with a very large archive of unlabeled images, text, audio, and video footage wishes to index its assets to allow rapid identification of relevant content by the Research team. The company wants to use machine learning to accelerate the efforts of its in-house researchers who have limited machine learning expertise.

Which is the FASTEST route to index the assets?

A. Use Amazon Rekognition, Amazon Comprehend, and Amazon Transcribe to tag data into distinct categories/classes.

B. Create a set of Amazon Mechanical Turk Human Intelligence Tasks to label all footage.

C. Use Amazon Transcribe to convert speech to text. Use the Amazon SageMaker Neural Topic Model (NTM) and Object Detection algorithms to tag data into distinct categories/classes.

D. Use the AWS Deep Learning AMI and Amazon EC2 GPU instances to create custom models for audio transcription and topic modeling, and use object detection to tag data into distinct categories/ classes.

ANSWER: A

QUESTION NO: 12

A company has video feeds and images of a subway train station. The company wants to create a deep learning model that will alert the station manager if any passenger crosses the yellow safety line when there is no train in the station. The alert will be based on the video feeds. The company wants the model to detect the yellow line, the passengers who cross the yellow line, and the trains in the video feeds. This task requires labeling. The video data must remain confidential.

A data scientist creates a bounding box to label the sample data and uses an object detection model. However, the object detection model cannot clearly demarcate the yellow line, the passengers who cross the yellow line, and the trains.

Which labeling approach will help the company improve this model?

A. Use Amazon Rekognition Custom Labels to label the dataset and create a custom Amazon Rekognition object detection model. Create a private workforce. Use Amazon Augmented AI (Amazon A2I) to review the low-confidence predictions and retrain the custom Amazon Rekognition model.

B. Use an Amazon SageMaker Ground Truth object detection labeling task. Use Amazon Mechanical Turk as the labeling workforce.

C. Use Amazon Rekognition Custom Labels to label the dataset and create a custom Amazon Rekognition object detection model. Create a workforce with a third-party AWS Marketplace vendor. Use Amazon Augmented AI (Amazon A2I) to review the low-confidence predictions and retrain the custom Amazon Rekognition model.

D. Use an Amazon SageMaker Ground Truth semantic segmentation labeling task. Use a private workforce as the labeling workforce.

ANSWER: B

Explanation:

Reference: https://docs.aws.amazon.com/sagemaker/latest/dg/sms-workforce-management-public.html

QUESTION NO: 13

A machine learning (ML) specialist wants to create a data preparation job that uses a PySpark script with complex window aggregation operations to create data for training and testing. The ML specialist needs to evaluate the impact of the number of features and the sample count on model performance.

Which approach should the ML specialist use to determine the ideal data transformations for the model?

A. Add an Amazon SageMaker Debugger hook to the script to capture key metrics. Run the script as an AWS Glue job.

B. Add an Amazon SageMaker Experiments tracker to the script to capture key metrics. Run the script as an AWS Glue job.

C. Add an Amazon SageMaker Debugger hook to the script to capture key parameters. Run the script as a SageMaker processing job.

D. Add an Amazon SageMaker Experiments tracker to the script to capture key parameters. Run the script as a SageMaker processing job.

ANSWER: B

QUESTION NO: 14

A company ingests machine learning (ML) data from web advertising clicks into an Amazon S3 data lake. Click data is added to an Amazon Kinesis data stream by using the Kinesis Producer Library (KPL). The data is loaded into the S3 data lake from the data stream by using an Amazon Kinesis Data Firehose delivery stream. As the data volume increases, an ML specialist notices that the rate of data ingested into Amazon S3 is relatively constant. There also is an increasing backlog of data for Kinesis Data Streams and Kinesis Data Firehose to ingest.

Which next step is MOST likely to improve the data ingestion rate into Amazon S3?

A. Increase the number of S3 prefixes for the delivery stream to write to.

B. Decrease the retention period for the data stream.

C. Increase the number of shards for the data stream.

D. Add more consumers using the Kinesis Client Library (KCL).

ANSWER: C

QUESTION NO: 15

A company has an ecommerce website with a product recommendation engine built in TensorFlow. The recommendation engine endpoint is hosted by Amazon SageMaker. Three compute-optimized instances support the expected peak load of the website.

Response times on the product recommendation page are increasing at the beginning of each month. Some users are encountering errors. The website receives the majority of its traffic between 8 AM and 6 PM on weekdays in a single time zone.

Which of the following options are the MOST effective in solving the issue while keeping costs to a minimum? (Choose two.)

A. Configure the endpoint to use Amazon Elastic Inference (EI) accelerators.

B. Create a new endpoint configuration with two production variants.

C. Configure the endpoint to automatically scale with the InvocationsPerInstance metric.

D. Deploy a second instance pool to support a blue/green deployment of models.

E. Reconfigure the endpoint to use burstable instances.

ANSWER: B D