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.passapply.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

QUESTION 1

A machine learning (ML) specialist wants to secure calls to the Amazon SageMaker Service API. The specialist has configured Amazon VPC with a VPC interface endpoint for the Amazon SageMaker Service API and is attempting to secure traffic from specific sets of instances and IAM users. The VPC is configured with a single public subnet.

Which combination of steps should the ML specialist take to secure the traffic? (Choose two.)

A. Add a VPC endpoint policy to allow access to the IAM users.

B. Modify the users\\' IAM policy to allow access to Amazon SageMaker Service API calls only.

C. Modify the security group on the endpoint network interface to restrict access to the instances.

D. Modify the ACL on the endpoint network interface to restrict access to the instances.

E. Add a SageMaker Runtime VPC endpoint interface to the VPC.

Correct Answer: AC

Reference: https://aws.amazon.com/blogs/machine-learning/securing-all-amazon-sagemaker-api-calls-with-aws-privatelink/

QUESTION 2

A data scientist has 20 TB of data in CSV format in an Amazon S3 bucket. The data scientist needs to convert the data to Apache Parquet format.

How can the data scientist convert the file format with the LEAST amount of effort?

A. Use an AWS Glue crawler to convert the file format.

B. Write a script to convert the file format. Run the script as an AWS Glue job.

C. Write a script to convert the file format. Run the script on an Amazon EMR cluster.

D. Write a script to convert the file format. Run the script in an Amazon SageMaker notebook.

Correct Answer: B

QUESTION 3

A Machine Learning Specialist previously trained a logistic regression model using scikit-learn on a local machine, and the Specialist now wants to deploy it to production for inference only.

What steps should be taken to ensure Amazon SageMaker can host a model that was trained locally?

A. Build the Docker image with the inference code. Tag the Docker image with the registry hostname and upload it to Amazon ECR.

B. Serialize the trained model so the format is compressed for deployment. Tag the Docker image with the registry hostname and upload it to Amazon S3.

C. Serialize the trained model so the format is compressed for deployment. Build the image and upload it to Docker Hub.

D. Build the Docker image with the inference code. Configure Docker Hub and upload the image to Amazon ECR.

Correct Answer: A

https://sagemaker-workshop.com/custom/containers.html

QUESTION 4

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.

Correct Answer: C

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

QUESTION 5

A manufacturing company wants to monitor its devices for anomalous behavior. A data scientist has trained an Amazon SageMaker scikit-learn model that classifies a device as normal or anomalous based on its 4-day telemetry. The 4-day telemetry of each device is collected in a separate file and is placed in an Amazon S3 bucket once every hour. The total time to run the model across the telemetry for all devices is 5 minutes.

What is the MOST cost-effective solution for the company to use to run the model across the telemetry for all the devices?

A. SageMaker Batch Transform

B. SageMaker Asynchronous Inference

C. SageMaker Processing

D. A SageMaker multi-container endpoint

Correct Answer: A

https://docs.aws.amazon.com/sagemaker/latest/dg/inference-cost-optimization.html

"Use batch inference for workloads for which you need inference for a large set of data for processes that happen offline (that is, you don\\'t need a persistent endpoint). You pay for the instance for the duration of the batch inference job". As you pay for the batch job duration, cost should not be an issue with Batch transform.

"Use asynchronous inference for asynchronous workloads that process up to 1 GB of data (such as text corpus, image, video, and audio) that are latency insensitive and cost sensitive. With asynchronous inference, you can control costs by specifying a fixed number of instances for the optimal processing rate instead of provisioning for the peak. You can also scale down to zero to save additional costs."

QUESTION 6

A company is using a legacy telephony platform and has several years remaining on its contract. The company wants to move to AWS and wants to implement the following machine learning features:

1.

Call transcription in multiple languages

2.

Categorization of calls based on the transcript

3.

Detection of the main customer issues in the calls

4.

Customer sentiment analysis for each line of the transcript, with positive or negative indication and scoring of that sentiment

Which AWS solution will meet these requirements with the LEAST amount of custom model training?

A. Use Amazon Transcribe to process audio calls to produce transcripts, categorize calls, and detect issues. Use Amazon Comprehend to analyze sentiment.

B. Use Amazon Transcribe to process audio calls to produce transcripts. Use Amazon Comprehend to categorize calls, detect issues, and analyze sentiment

C. Use Contact Lens for Amazon Connect to process audio calls to produce transcripts, categorize calls, detect issues, and analyze sentiment.

D. Use Contact Lens for Amazon Connect to process audio calls to produce transcripts. Use Amazon Comprehend to categorize calls, detect issues, and analyze sentiment.

Correct Answer: C

https://aws.amazon.com/connect/contact-lens/

QUESTION 7

A machine learning (ML) specialist at a retail company is forecasting sales for one of the company\\'s stores. The ML specialist is using data from the past 10 years. The company has provided a dataset that includes the total amount of money in sales each day for the store. Approximately 5% of the days are missing sales data.

The ML specialist builds a simple forecasting model with the dataset and discovers that the model performs poorly. The performance is poor around the time of seasonal events, when the model consistently predicts sales figures that are too low or too high.

Which actions should the ML specialist take to try to improve the model\\'s performance? (Choose two.)

A. Add information about the store\\'s sales periods to the dataset.

B. Aggregate sales figures from stores in the same proximity.

C. Apply smoothing to correct for seasonal variation.

D. Change the forecast frequency from daily to weekly.

E. Replace missing values in the dataset by using linear interpolation.

Correct Answer: AE

A to improve model in seasonal periods E to fill missing data

QUESTION 8

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

Correct Answer: B

QUESTION 9

A Machine Learning Specialist is given a structured dataset on the shopping habits of a company\\'s customer base. The dataset contains thousands of columns of data and hundreds of numerical columns for each customer. The Specialist wants to identify whether there are natural groupings for these columns across all customers and visualize the results as quickly as possible.

What approach should the Specialist take to accomplish these tasks?

A. Embed the numerical features using the t-distributed stochastic neighbor embedding (t-SNE) algorithm and create a scatter plot.

B. Run k-means using the Euclidean distance measure for different values of k and create an elbow plot.

C. Embed the numerical features using the t-distributed stochastic neighbor embedding (t-SNE) algorithm and create a line graph.

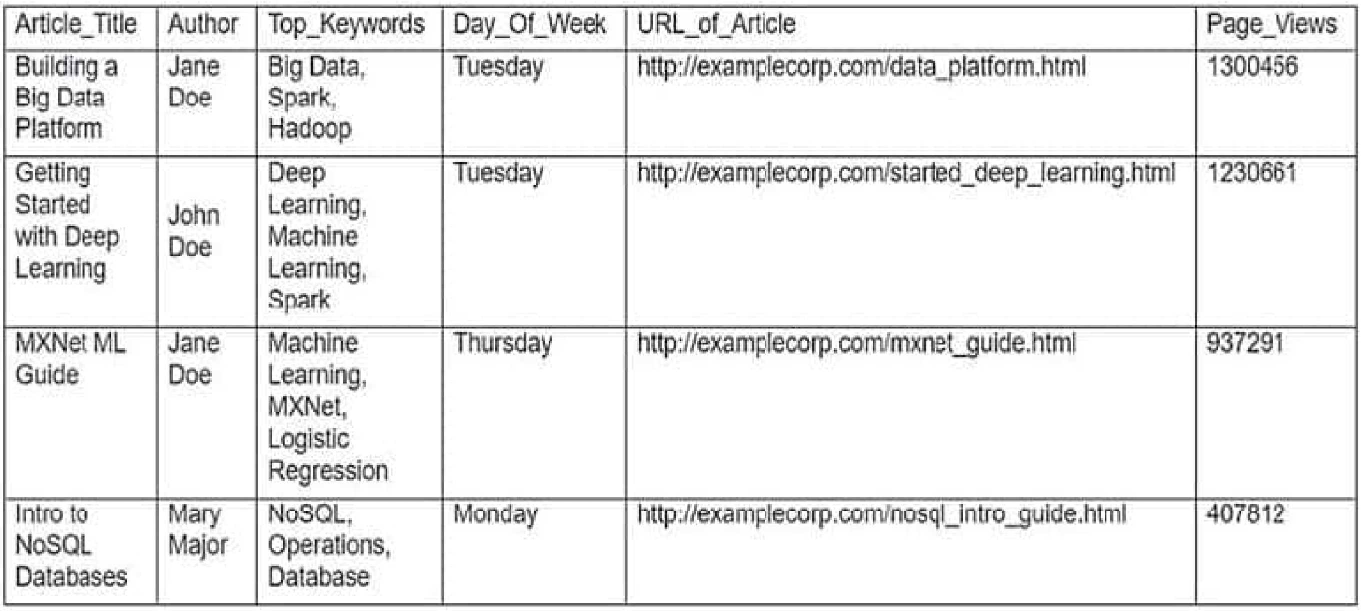
D. Run k-means using the Euclidean distance measure for different values of k and create box plots for each numerical column within each cluster.

Correct Answer: A

https://towardsdatascience.com/an-introduction-to-t-sne-with-python-example-5a3a293108d1

QUESTION 10

Machine Learning Specialist is working with a media company to perform classification on popular articles from the company\\'s website. The company is using random forests to classify how popular an article will be before it is published. A sample of the data being used is below.



Article_Title	Author	Top_Keywords	Day_Of_Week	URL_of_Article	Page_Views
Building a Big Data Platform	Jane Doe	Big Data, Spark, Hadoop	Tuesday	http://examplecorp.com/data_platform.html	1300456
Getting Started with Deep Learning	John Doe	Deep Learning, Machine Learning, Spark	Tuesday	http://examplecorp.com/started_deep_learning.html	1230661
MXNet ML Guide	Jane Doe	Machine Learning, MXNet, Logistic Regression	Thursday	http://examplecorp.com/mxnet_guide.html	937291
Intro to NoSQL Databases	Mary Major	NoSQL, Operations, Database	Monday	http://examplecorp.com/nosql_intro_guide.html	407812

Given the dataset, the Specialist wants to convert the Day_Of_Week column to binary values. What technique should be used to convert this column to binary values?

A. Binarization

B. One-hot encoding

C. Tokenization

D. Normalization transformation

Correct Answer: B

QUESTION 11

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

Correct Answer: B

SSML is specific to that particular doument, like W3C an be pronounced as "World Wide Web Consortium" using $_{\text{W3C}}$ in that specific document and when you create a new document, you need to format again. But with LEXICONS, you can upload a lexicon file once and ALL the FUTURE documents can just have W3C and that will be pronounced as "World Wide Web Consortium".. so answer is B, because the question asks for "future" documents.

QUESTION 12

A machine learning (ML) engineer is integrating a production model with a customer metadata repository for real-time inference. The repository is hosted in Amazon SageMaker Feature Store. The engineer wants to retrieve only the latest version of the customer metadata record for a single customer at a time.

Which solution will meet these requirements?

A. Use the SageMaker Feature Store BatchGetRecord API with the record identifier. Filter to find the latest record.

B. Create an Amazon Athena query to retrieve the data from the feature table.

C. Create an Amazon Athena query to retrieve the data from the feature table. Use the write_time value to find the latest record.

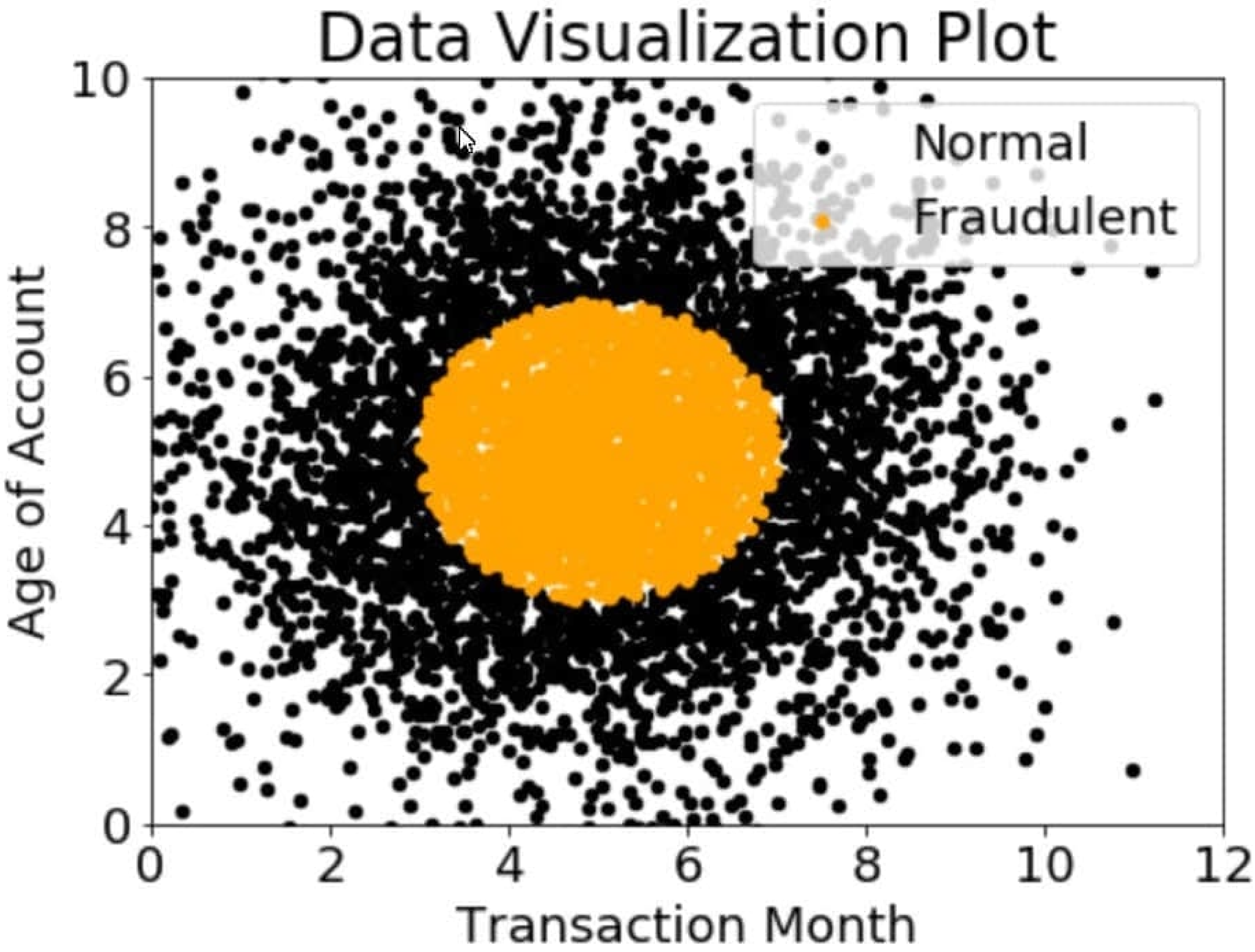
D. Use the SageMaker Feature Store GetRecord API with the record identifier.

Correct Answer: D

QUESTION 13

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 recall with respect to the fraudulent class?

A. Decision tree

B. Linear support vector machine (SVM)

C. Naive Bayesian classifier

D. Single Perceptron with sigmoidal activation function

Correct Answer: C

https://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html#sphx-glr-auto-examples-classification-plot-classifier-comparison-py

QUESTION 14

An automotive company uses computer vision in its autonomous cars. The company trained its object detection models successfully by using transfer learning from a convolutional neural network (CNN). The company trained the models by

using PyTorch through the Amazon SageMaker SDK.

The vehicles have limited hardware and compute power. The company wants to optimize the model to reduce memory, battery, and hardware consumption without a significant sacrifice in accuracy.

Which solution will improve the computational efficiency of the models?

A. Use Amazon CloudWatch metrics to gain visibility into the SageMaker training weights, gradients, biases, and activation outputs. Compute the filter ranks based on the training information. Apply pruning to remove the low-ranking filters. Set new weights based on the pruned set of filters. Run a new training job with the pruned model.

B. Use Amazon SageMaker Ground Truth to build and run data labeling workflows. Collect a larger labeled dataset with the labelling workflows. Run a new training job that uses the new labeled data with previous training data.

C. Use Amazon SageMaker Debugger to gain visibility into the training weights, gradients, biases, and activation outputs. Compute the filter ranks based on the training information. Apply pruning to remove the low-ranking filters. Set the new weights based on the pruned set of filters. Run a new training job with the pruned model.

D. Use Amazon SageMaker Model Monitor to gain visibility into the ModelLatency metric and OverheadLatency metric of the model after the company deploys the model. Increase the model learning rate. Run a new training job.

Correct Answer: C

QUESTION 15

A company wants to segment a large group of customers into subgroups based on shared characteristics. The company\\'s data scientist is planning to use the Amazon SageMaker built-in k-means clustering algorithm for this task. The data scientist needs to determine the optimal number of subgroups (k) to use.

Which data visualization approach will MOST accurately determine the optimal value of k?

A. Calculate the principal component analysis (PCA) components. Run the k-means clustering algorithm for a range of k by using only the first two PCA components. For each value of k, create a scatter plot with a different color for each cluster. The optimal value of k is the value where the clusters start to look reasonably separated.

B. Calculate the principal component analysis (PCA) components. Create a line plot of the number of components against the explained variance. The optimal value of k is the number of PCA components after which the curve starts decreasing in a linear fashion.

C. Create a t-distributed stochastic neighbor embedding (t-SNE) plot for a range of perplexity values. The optimal value of k is the value of perplexity, where the clusters start to look reasonably separated.

D. Run the k-means clustering algorithm for a range of k. For each value of k, calculate the sum of squared errors (SSE). Plot a line chart of the SSE for each value of k. The optimal value of k is the point after which the curve starts decreasing in a linear fashion.

Correct Answer: D