

Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 1

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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:

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

Based on the model evaluation results, why is this a viable model for production?

- 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.

Show Suggested Answer

Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 2

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 3

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Mobile Network Operator is building an analytics platform to analyze and optimize a company's operations using Amazon Athena and Amazon S3.

The source systems send data in .CSV format in real time. The Data Engineering team wants to transform the data to the Apache Parquet format before storing it on Amazon S3.

Which solution takes the LEAST effort to implement?

- A. Ingest .CSV data using Apache Kafka Streams on Amazon EC2 instances and use Kafka Connect S3 to serialize data as Parquet
- B. Ingest .CSV data from Amazon Kinesis Data Streams and use Amazon Glue to convert data into Parquet.
- C. Ingest .CSV data using Apache Spark Structured Streaming in an Amazon EMR cluster and use Apache Spark to convert data into Parquet.
- D. Ingest .CSV data from Amazon Kinesis Data Streams and use Amazon Kinesis Data Firehose to convert data into Parquet.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 4

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A city wants to monitor its air quality to address the consequences of air pollution. A Machine Learning Specialist needs to forecast the air quality in parts per million of contaminates for the next 2 days in the city. As this is a prototype, only daily data from the last year is available.

Which model is MOST likely to provide the best results in Amazon SageMaker?

- A. Use the Amazon SageMaker k-Nearest-Neighbors (kNN) algorithm on the single time series consisting of the full year of data with a predictor_type of regressor.
- B. Use Amazon SageMaker Random Cut Forest (RCF) on the single time series consisting of the full year of data.
- C. Use the Amazon SageMaker Linear Learner algorithm on the single time series consisting of the full year of data with a predictor_type of regressor.
- D. Use the Amazon SageMaker Linear Learner algorithm on the single time series consisting of the full year of data with a predictor_type of classifier.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 5

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Data Engineer needs to build a model using a dataset containing customer credit card information

How can the Data Engineer ensure the data remains encrypted and the credit card information is secure?

- A. Use a custom encryption algorithm to encrypt the data and store the data on an Amazon SageMaker instance in a VPC. Use the SageMaker DeepAR algorithm to randomize the credit card numbers.
- B. Use an IAM policy to encrypt the data on the Amazon S3 bucket and Amazon Kinesis to automatically discard credit card numbers and insert fake credit card numbers.
- C. Use an Amazon SageMaker launch configuration to encrypt the data once it is copied to the SageMaker instance in a VPC. Use the SageMaker principal component analysis (PCA) algorithm to reduce the length of the credit card numbers.
- D. Use AWS KMS to encrypt the data on Amazon S3 and Amazon SageMaker, and redact the credit card numbers from the customer data with AWS Glue.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 6

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 7

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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 log 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.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 8

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 9

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Machine Learning Specialist is developing a custom video recommendation model for an application. The dataset used to train this model is very large with millions of data points and is hosted in an Amazon S3 bucket. The Specialist wants to avoid loading all of this data onto an Amazon SageMaker notebook instance because it would take hours to move and will exceed the attached 5 GB Amazon EBS volume on the notebook instance.

Which approach allows the Specialist to use all the data to train the model?

- A. Load a smaller subset of the data into the SageMaker notebook and train locally. Confirm that the training code is executing and the model parameters seem reasonable. Initiate a SageMaker training job using the full dataset from the S3 bucket using Pipe input mode.
- B. Launch an Amazon EC2 instance with an AWS Deep Learning AMI and attach the S3 bucket to the instance. Train on a small amount of the data to verify the training code and hyperparameters. Go back to Amazon SageMaker and train using the full dataset
- C. Use AWS Glue to train a model using a small subset of the data to confirm that the data will be compatible with Amazon SageMaker. Initiate a SageMaker training job using the full dataset from the S3 bucket using Pipe input mode.
- D. Load a smaller subset of the data into the SageMaker notebook and train locally. Confirm that the training code is executing and the model parameters seem reasonable. Launch an Amazon EC2 instance with an AWS Deep Learning AMI and attach the S3 bucket to train the full dataset.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 10

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 11

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Machine Learning Specialist receives customer data for an online shopping website. The data includes demographics, past visits, and locality information. The Specialist must develop a machine learning approach to identify the customer shopping patterns, preferences, and trends to enhance the website for better service and smart recommendations.

Which solution should the Specialist recommend?

- A. Latent Dirichlet Allocation (LDA) for the given collection of discrete data to identify patterns in the customer database.
- B. A neural network with a minimum of three layers and random initial weights to identify patterns in the customer database.
- C. Collaborative filtering based on user interactions and correlations to identify patterns in the customer database.
- D. Random Cut Forest (RCF) over random subsamples to identify patterns in the customer database.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 12

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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

Show Suggested Answer



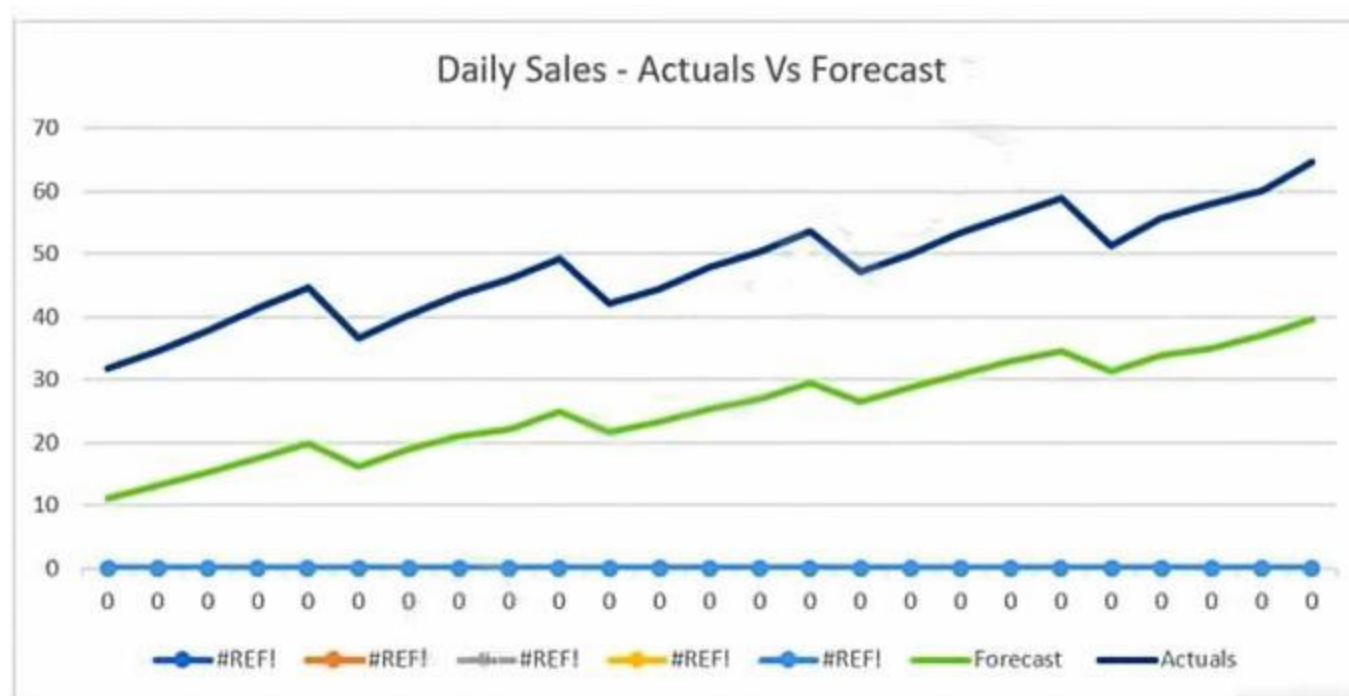
Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 13

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

The displayed graph is from a forecasting 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.

Show Suggested Answer

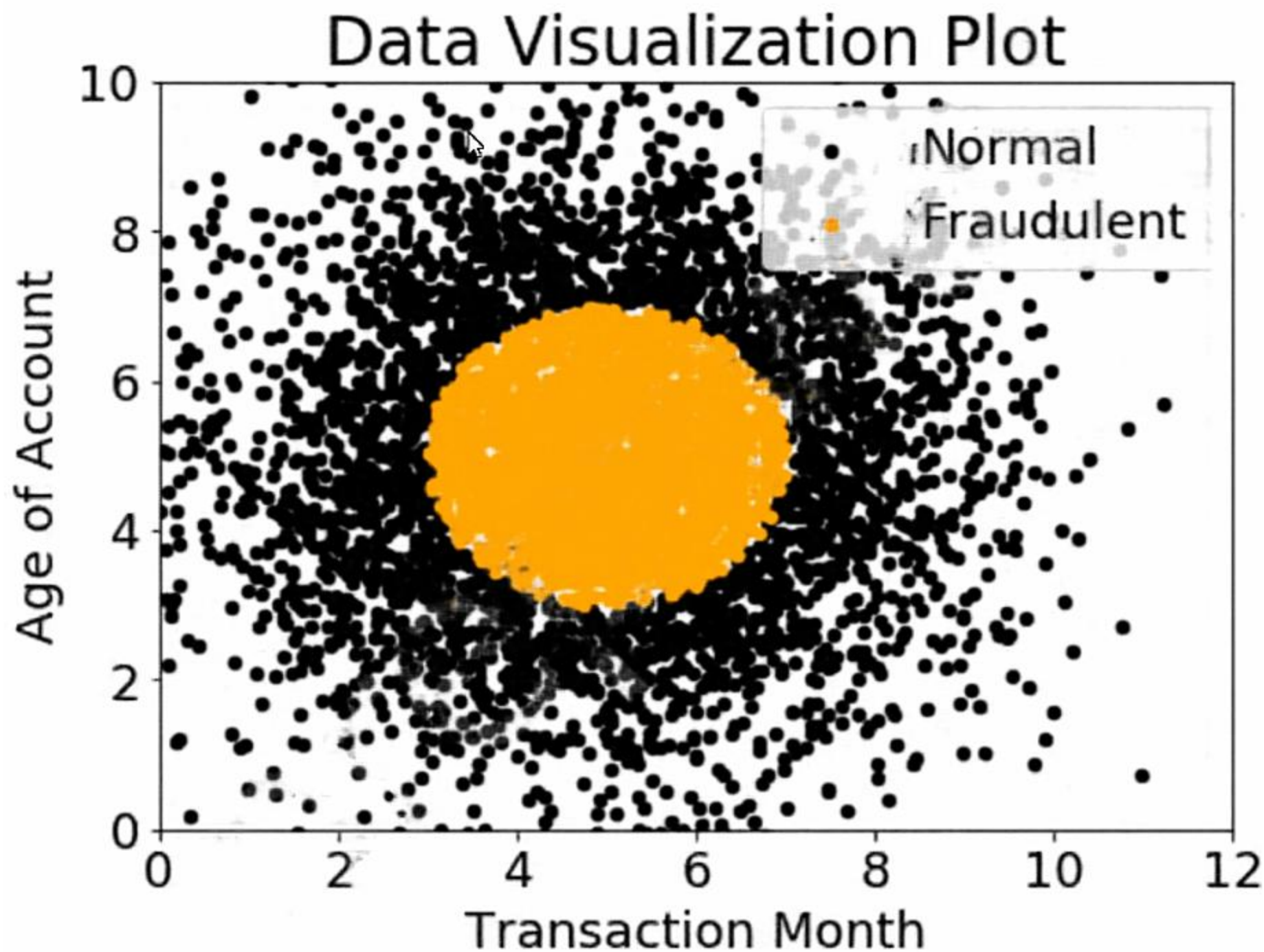
Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 14

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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 (SELU)
- B. Logistic regression
- C. Support vector machine (SVM) with non-linear kernel
- D. Single perceptron with tanh activation function

Show Suggested Answer

Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 15

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Machine Learning Specialist at a company sensitive to security is preparing a dataset for model training. The dataset is stored in Amazon S3 and contains Personally Identifiable Information (PII).

The dataset:

- ⇒ Must be accessible from a VPC only.
- ⇒ Must not traverse the public internet.

How can these requirements be satisfied?

- A. Create a VPC endpoint and apply a bucket access policy that restricts access to the given VPC endpoint and the VPC.
- B. Create a VPC endpoint and apply a bucket access policy that allows access from the given VPC endpoint and an Amazon EC2 instance.
- C. Create a VPC endpoint and use Network Access Control Lists (NACLs) to allow traffic between only the given VPC endpoint and an Amazon EC2 instance.
- D. Create a VPC endpoint and use security groups to restrict access to the given VPC endpoint and an Amazon EC2 instance

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 16

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 17

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 18

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Machine Learning Specialist is packaging a custom ResNet model into a Docker container so the company can leverage Amazon SageMaker for training. The Specialist is using Amazon EC2 P3 instances to train the model and needs to properly configure the Docker container to leverage the NVIDIA GPUs.

What does the Specialist need to do?

- A. Bundle the NVIDIA drivers with the Docker image.
- B. Build the Docker container to be NVIDIA-Docker compatible.
- C. Organize the Docker container's file structure to execute on GPU instances.
- D. Set the GPU flag in the Amazon SageMaker CreateTrainingJob request body.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 19

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Machine Learning Specialist is building a logistic regression model that will predict whether or not a person will order a pizza. The Specialist is trying to build the optimal model with an ideal classification threshold.

What model evaluation technique should the Specialist use to understand how different classification thresholds will impact the model's performance?

- A. Receiver operating characteristic (ROC) curve
- B. Misclassification rate
- C. Root Mean Square Error (RMSE)
- D. L1 norm

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 20

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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 vectors that store edit distance with every other word.
- D. Download word embeddings pre-trained on a large corpus.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 21

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 22

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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 to 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.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 23

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Machine Learning Specialist is building a convolutional neural network (CNN) that will classify 10 types of animals. The Specialist has built a series of layers in a neural network that will take an input image of an animal, pass it through a series of convolutional and pooling layers, and then finally pass it through a dense and fully connected layer with 10 nodes. The Specialist would like to get an output from the neural network that is a probability distribution of how likely it is that the input image belongs to each of the 10 classes.

Which function will produce the desired output?

- A. Dropout
- B. Smooth L1 loss
- C. Softmax
- D. Rectified linear units (ReLU)

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 24

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Machine Learning Specialist trained a regression model, but the first iteration needs optimizing. The Specialist needs to understand whether the model is more frequently overestimating or underestimating the target.

What option can the Specialist use to determine whether it is overestimating or underestimating the target value?

- A. Root Mean Square Error (RMSE)
- B. Residual plots
- C. Area under the curve
- D. Confusion matrix

Show Suggested Answer



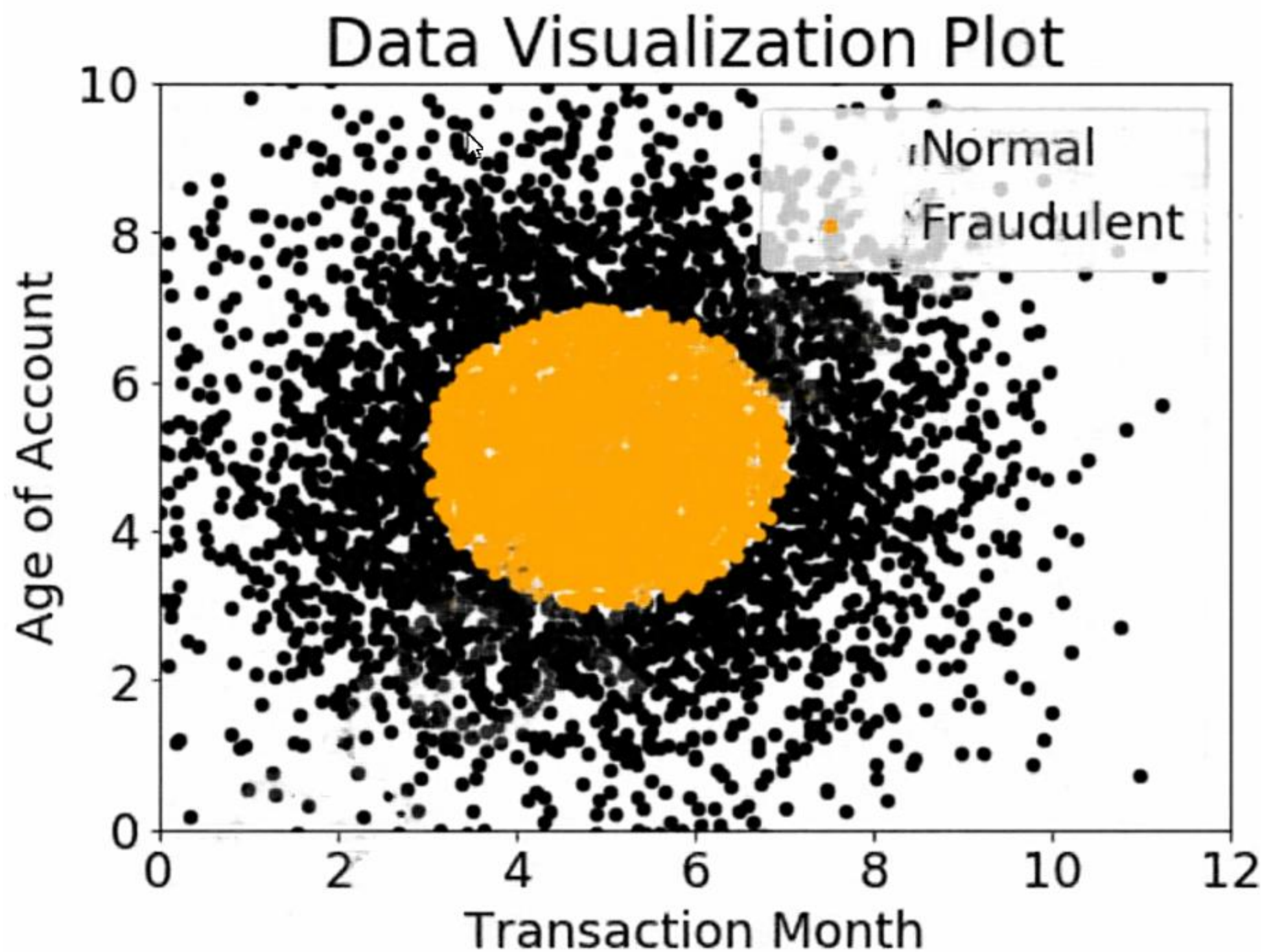
Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 25

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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

Show Suggested Answer

Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 26

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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 t-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.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 27

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Machine Learning Specialist is creating a new natural language processing application that processes a dataset comprised of 1 million sentences. The aim is to then run Word2Vec to generate embeddings of the sentences and enable different types of predictions.

Here is an example from the dataset:

"The quck BROWN FOX jumps over the lazy dog."

Which of the following are the operations the Specialist needs to perform to correctly sanitize and prepare the data in a repeatable manner? (Choose three.)

- A. Perform part-of-speech tagging and keep the action verb and the nouns only.
- B. Normalize all words by making the sentence lowercase.
- C. Remove stop words using an English stopword dictionary.
- D. Correct the typography on "quck" to "quick."
- E. One-hot encode all words in the sentence.
- F. Tokenize the sentence into words.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 28

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 29

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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? (Choose 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.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 30

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 31

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A monitoring service generates 1 TB of scale metrics record data every minute. A Research team performs queries on this data using Amazon Athena. The queries run slowly due to the large volume of data, and the team requires better performance.

How should the records be stored in Amazon S3 to improve query performance?

- A. CSV files
- B. Parquet files
- C. Compressed JSON
- D. RecordIO

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 32

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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

Show Suggested Answer

Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 33

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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 999,000 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? (Choose two.)

- A. Add more deep trees to the random forest to enable the model to learn more features.
- B. Include a copy of the samples in the test dataset 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.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 34

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 35

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 36

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 37

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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-real-time 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

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 38

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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? (Choose 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.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 39

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Machine Learning Specialist built an image classification deep learning model. However, the Specialist ran into an overfitting problem in which the training and testing accuracies were 99% and 75%, respectively.

How should the Specialist address this issue and what is the reason behind it?

- A. The learning rate should be increased because the optimization process was trapped at a local minimum.
- B. The dropout rate at the flatten layer should be increased because the model is not generalized enough.
- C. The dimensionality of dense layer next to the flatten layer should be increased because the model is not complex enough.
- D. The epoch number should be increased because the optimization process was terminated before it reached the global minimum.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 40

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 41

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 42

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Machine Learning Specialist is implementing a full Bayesian network on a dataset that describes public transit in New York City. One of the random variables is discrete, and represents the number of minutes New Yorkers wait for a bus given that the buses cycle every 10 minutes, with a mean of 3 minutes.

Which prior probability distribution should the ML Specialist use for this variable?

- A. Poisson distribution
- B. Uniform distribution
- C. Normal distribution
- D. Binomial distribution

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 43

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Data Science team within a large company uses Amazon SageMaker notebooks to access data stored in Amazon S3 buckets. The IT Security team is concerned that internet-enabled notebook instances create a security vulnerability where malicious code running on the instances could compromise data privacy.

The company mandates that all instances stay within a secured VPC with no internet access, and data communication traffic must stay within the AWS network.

How should the Data Science team configure the notebook instance placement to meet these requirements?

- A. Associate the Amazon SageMaker notebook with a private subnet in a VPC. Place the Amazon SageMaker endpoint and S3 buckets within the same VPC.
- B. Associate the Amazon SageMaker notebook with a private subnet in a VPC. Use IAM policies to grant access to Amazon S3 and Amazon SageMaker.
- C. Associate the Amazon SageMaker notebook with a private subnet in a VPC. Ensure the VPC has S3 VPC endpoints and Amazon SageMaker VPC endpoints attached to it.
- D. Associate the Amazon SageMaker notebook with a private subnet in a VPC. Ensure the VPC has a NAT gateway and an associated security group allowing only outbound connections to Amazon S3 and Amazon SageMaker.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 44

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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? (Choose three.)

- A. Decrease regularization.
- B. Increase regularization.
- C. Increase dropout.
- D. Decrease dropout.
- E. Increase feature combinations.
- F. Decrease feature combinations.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 45

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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. 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.
- B. 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.
- 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 inserts it into an Amazon RDS PostgreSQL database. Have the Analysts query and run dashboards from the RDS database.
- D. 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.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 46

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 47

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 48

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

What should the Specialist do to initialize the model to re-train it with the custom data?

- 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.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 49

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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. 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.
- B. 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. On each stream, use Amazon Rekognition Image to detect faces from a collection of known employees and alert when non-employees are detected.
- C. Install AWS DeepLens cameras and use the DeepLens_Kinesis_Video module to stream video to Amazon Kinesis Video Streams for each camera. 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.
- D. Install AWS DeepLens cameras and use the DeepLens_Kinesis_Video module to stream video to Amazon Kinesis Video Streams for each camera. 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.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 50

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Marketing Manager at a pet insurance company plans to launch a targeted marketing campaign on social media to acquire new customers. Currently, the company has the following data in Amazon Aurora:

- ⇒ Profiles for all past and existing customers
- ⇒ Profiles for all past and existing insured pets
- ⇒ Policy-level information
- ⇒ Premiums received
- ⇒ Claims paid

What steps should be taken to implement a machine learning model to identify potential new customers on social media?

- A. Use regression on customer profile data to understand key characteristics of consumer segments. Find similar profiles on social media
- B. Use clustering on customer profile data to understand key characteristics of consumer segments. Find similar profiles on social media
- C. Use a recommendation engine on customer profile data to understand key characteristics of consumer segments. Find similar profiles on social media.
- D. Use a decision tree classifier engine on customer profile data to understand key characteristics of consumer segments. Find similar profiles on social media.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 51

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 52

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 53

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company's Machine Learning Specialist needs to improve the training speed of a time-series forecasting model using TensorFlow. The training is currently implemented on a single-GPU machine and takes approximately 23 hours to complete. The training needs to be run daily.

The model accuracy is acceptable, but the company anticipates a continuous increase in the size of the training data and a need to update the model on an hourly, rather than a daily, basis. The company also wants to minimize coding effort and infrastructure changes.

What should the Machine Learning Specialist do to the training solution to allow it to scale for future demand?

- A. Do not change the TensorFlow code. Change the machine to one with a more powerful GPU to speed up the training.
- B. Change the TensorFlow code to implement a Horovod distributed framework supported by Amazon SageMaker. Parallelize the training to as many machines as needed to achieve the business goals.
- C. Switch to using a built-in AWS SageMaker DeepAR model. Parallelize the training to as many machines as needed to achieve the business goals.
- D. Move the training to Amazon EMR and distribute the workload to as many machines as needed to achieve the business goals.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 54

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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)

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 55

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 56

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Machine Learning Specialist is preparing data for training on Amazon SageMaker. The Specialist is using one of the SageMaker built-in algorithms for the training. The dataset is stored in .CSV format and 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.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 57

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Machine Learning Specialist is required to build a supervised image-recognition model to identify a cat. The ML Specialist performs some tests and records the following results for a neural network-based image classifier:

Total number of images available = 1,000

Test set images = 100 (constant test set)

The ML Specialist notices that, in over 75% of the misclassified images, the cats were held upside down by their owners.

Which techniques can be used by the ML Specialist to improve this specific test error?

- A. Increase the training data by adding variation in rotation for training images.
- B. Increase the number of epochs for model training
- C. Increase the number of layers for the neural network.
- D. Increase the dropout rate for the second-to-last layer.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 58

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Machine Learning Specialist needs to be able to ingest streaming data and store it in Apache Parquet files for exploration and analysis.

Which of the following services would both ingest and store this data in the correct format?

- A. AWS DMS
- B. Amazon Kinesis Data Streams
- C. Amazon Kinesis Data Firehose
- D. Amazon Kinesis Data Analytics

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 59

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A data scientist has explored and sanitized a dataset in preparation for the modeling phase of a supervised learning task. The statistical dispersion can vary widely between features, sometimes by several orders of magnitude. Before moving on to the modeling phase, the data scientist wants to ensure that the prediction performance on the production data is as accurate as possible.

Which sequence of steps should the data scientist take to meet these requirements?

- A. Apply random sampling to the dataset. Then split the dataset into training, validation, and test sets.
- B. Split the dataset into training, validation, and test sets. Then rescale the training set and apply the same scaling to the validation and test sets.
- C. Rescale the dataset. Then split the dataset into training, validation, and test sets.
- D. Split the dataset into training, validation, and test sets. Then rescale the training set, the validation set, and the test set independently.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 60

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 61

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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 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, 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.
- C. 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.
- D. 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.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 62

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 63

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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. AnXGBoost 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)

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 64

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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 analysis and entity detection
- B. Amazon SageMaker BlazingText cbow mode
- C. Natural Language Toolkit (NLTK) stemming and stop word removal
- D. Scikit-learn term frequency-inverse document frequency (TF-IDF) vectorizer

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 65

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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 bigram (OSB) transformation to apply a fixed-size sliding window to generate new features of a similar magnitude.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 66

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Machine Learning Specialist must build out a process to query a dataset on Amazon S3 using Amazon Athena. The dataset contains more than 800,000 records stored as plaintext CSV files. Each record contains 200 columns and is approximately 1.5 MB in size. Most queries will span 5 to 10 columns only.

How should the Machine Learning Specialist transform the dataset to minimize query runtime?

- A. Convert the records to Apache Parquet format.
- B. Convert the records to JSON format.
- C. Convert the records to GZIP CSV format.
- D. Convert the records to XML format.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 67

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Machine Learning Specialist is developing a daily ETL workflow containing multiple ETL jobs. The workflow consists of the following processes:

- * Start the workflow as soon as data is uploaded to Amazon S3.
- * When all the datasets are available in Amazon S3, start an ETL job to join the uploaded datasets with multiple terabyte-sized datasets already stored in Amazon S3.
- * Store the results of joining datasets in Amazon S3.
- * If one of the jobs fails, send a notification to the Administrator.

Which configuration will meet these requirements?

- A. Use AWS Lambda to trigger an AWS Step Functions workflow to wait for dataset uploads to complete in Amazon S3. Use AWS Glue to join the datasets. Use an Amazon CloudWatch alarm to send an SNS notification to the Administrator in the case of a failure.
- B. Develop the ETL workflow using AWS Lambda to start an Amazon SageMaker notebook instance. Use a lifecycle configuration script to join the datasets and persist the results in Amazon S3. Use an Amazon CloudWatch alarm to send an SNS notification to the Administrator in the case of a failure.
- C. Develop the ETL workflow using AWS Batch to trigger the start of ETL jobs when data is uploaded to Amazon S3. Use AWS Glue to join the datasets in Amazon S3. Use an Amazon CloudWatch alarm to send an SNS notification to the Administrator in the case of a failure.
- D. Use AWS Lambda to chain other Lambda functions to read and join the datasets in Amazon S3 as soon as the data is uploaded to Amazon S3. Use an Amazon CloudWatch alarm to send an SNS notification to the Administrator in the case of a failure.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 68

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

An agency collects census information within a country to determine healthcare and social program needs by province and city. The census form collects responses for approximately 500 questions from each citizen.

Which combination of algorithms would provide the appropriate insights? (Choose two.)

- A. The factorization machines (FM) algorithm
- B. The Latent Dirichlet Allocation (LDA) algorithm
- C. The principal component analysis (PCA) algorithm
- D. The k-means algorithm
- E. The Random Cut Forest (RCF) algorithm

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 69

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 70

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Machine Learning Specialist uploads a dataset to an Amazon S3 bucket protected with server-side encryption using AWS KMS. How should the ML Specialist define the Amazon SageMaker notebook instance so it can read the same dataset from Amazon S3?

- A. Define security group(s) to allow all HTTP inbound/outbound traffic and assign those security group(s) to the Amazon SageMaker notebook instance.
- B. Configure the Amazon SageMaker notebook instance to have access to the VPC. Grant permission in the KMS key policy to the notebook's KMS role.
- C. Assign an IAM role to the Amazon SageMaker notebook with S3 read access to the dataset. Grant permission in the KMS key policy to that role.
- D. Assign the same KMS key used to encrypt data in Amazon S3 to the Amazon SageMaker notebook instance.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 71

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Data Scientist needs to migrate an existing on-premises ETL process to the cloud. The current process runs at regular time intervals and uses PySpark to combine and format multiple large data sources into a single consolidated output for downstream processing.

The Data Scientist has been given the following requirements to the cloud solution:

- ⇒ Combine multiple data sources.
- ⇒ Reuse existing PySpark logic.
- ⇒ Run the solution on the existing schedule.
- ⇒ Minimize the number of servers that will need to be managed.

Which architecture should the Data Scientist use to build this solution?

- A. Write the raw data to Amazon S3. Schedule an AWS Lambda function to submit a Spark step to a persistent Amazon EMR cluster based on the existing schedule. Use the existing PySpark logic to run the ETL job on the EMR cluster. Output the results to a `processed` location in Amazon S3 that is accessible for downstream use.
- B. Write the raw data to Amazon S3. Create an AWS Glue ETL job to perform the ETL processing against the input data. Write the ETL job in PySpark to leverage the existing logic. Create a new AWS Glue trigger to trigger the ETL job based on the existing schedule. Configure the output target of the ETL job to write to a `processed` location in Amazon S3 that is accessible for downstream use.
- C. Write the raw data to Amazon S3. Schedule an AWS Lambda function to run on the existing schedule and process the input data from Amazon S3. Write the Lambda logic in Python and implement the existing PySpark logic to perform the ETL process. Have the Lambda function output the results to a `processed` location in Amazon S3 that is accessible for downstream use.
- D. Use Amazon Kinesis Data Analytics to stream the input data and perform real-time SQL queries against the stream to carry out the required transformations within the stream. Deliver the output results to a `processed` location in Amazon S3 that is accessible for downstream use.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 72

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 73

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

An aircraft engine manufacturing company is measuring 200 performance metrics in a time-series. Engineers want to detect critical manufacturing defects in near-real time during testing. All of the data needs to be stored for offline analysis.

What approach would be the MOST effective to perform near-real time defect detection?

- A. Use AWS IoT Analytics for ingestion, storage, and further analysis. Use Jupyter notebooks from within AWS IoT Analytics to carry out analysis for anomalies.
- B. Use Amazon S3 for ingestion, storage, and further analysis. Use an Amazon EMR cluster to carry out Apache Spark ML k-means clustering to determine anomalies.
- C. Use Amazon S3 for ingestion, storage, and further analysis. Use the Amazon SageMaker Random Cut Forest (RCF) algorithm to determine anomalies.
- D. Use Amazon Kinesis Data Firehose for ingestion and Amazon Kinesis Data Analytics Random Cut Forest (RCF) to perform anomaly detection. Use Kinesis Data Firehose to store data in Amazon S3 for further analysis.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 74

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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 algorithm-specific 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

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 75

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Machine Learning Specialist wants to determine the appropriate `SageMakerVariantInvocationsPerInstance` setting for an endpoint automatic scaling configuration. The Specialist has performed a load test on a single instance and determined that peak requests per second (RPS) without service degradation is about 20 RPS. As this is the first deployment, the Specialist intends to set the invocation safety factor to 0.5.

Based on the stated parameters and given that the invocations per instance setting is measured on a per-minute basis, what should the Specialist set as the `SageMakerVariantInvocationsPerInstance` setting?

- A. 10
- B. 30
- C. 600
- D. 2,400

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 76

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company uses a long short-term memory (LSTM) model to evaluate the risk factors of a particular energy sector. The model reviews multi-page text documents to analyze each sentence of the text and categorize it as either a potential risk or no risk. The model is not performing well, even though the Data Scientist has experimented with many different network structures and tuned the corresponding hyperparameters.

Which approach will provide the MAXIMUM performance boost?

- A. Initialize the words by term frequency-inverse document frequency (TF-IDF) vectors pretrained on a large collection of news articles related to the energy sector.
- B. Use gated recurrent units (GRUs) instead of LSTM and run the training process until the validation loss stops decreasing.
- C. Reduce the learning rate and run the training process until the training loss stops decreasing.
- D. Initialize the words by word2vec embeddings pretrained on a large collection of news articles related to the energy sector.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 77

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Machine Learning Specialist needs to move and transform data in preparation for training. Some of the data needs to be processed in near-real time, and other data can be moved hourly. There are existing Amazon EMR MapReduce jobs to clean and feature engineering to perform on the data.

Which of the following services can feed data to the MapReduce jobs? (Choose two.)

- A. AWS DMS
- B. Amazon Kinesis
- C. AWS Data Pipeline
- D. Amazon Athena
- E. Amazon ES

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 78

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 79

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A trucking company is collecting live image data from its fleet of trucks across the globe. The data is growing rapidly and approximately 100 GB of new data is generated every day. The company wants to explore machine learning uses cases while ensuring the data is only accessible to specific IAM users.

Which storage option provides the most processing flexibility and will allow access control with IAM?

- A. Use a database, such as Amazon DynamoDB, to store the images, and set the IAM policies to restrict access to only the desired IAM users.
- B. Use an Amazon S3-backed data lake to store the raw images, and set up the permissions using bucket policies.
- C. Setup up Amazon EMR with Hadoop Distributed File System (HDFS) to store the files, and restrict access to the EMR instances using IAM policies.
- D. Configure Amazon EFS with IAM policies to make the data available to Amazon EC2 instances owned by the IAM users.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 80

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A credit card company wants to build a credit scoring model to help predict whether a new credit card applicant will default on a credit card payment. The company has collected data from a large number of sources with thousands of raw attributes. Early experiments to train a classification model revealed that many attributes are highly correlated, the large number of features slows down the training speed significantly, and that there are some overfitting issues.

The Data Scientist on this project would like to speed up the model training time without losing a lot of information from the original dataset.

Which feature engineering technique should the Data Scientist use to meet the objectives?

- A. Run self-correlation on all features and remove highly correlated features
- B. Normalize all numerical values to be between 0 and 1
- C. Use an autoencoder or principal component analysis (PCA) to replace original features with new features
- D. Cluster raw data using k-means and use sample data from each cluster to build a new dataset

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 81

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Data Scientist is training a multilayer perception (MLP) on a dataset with multiple classes. The target class of interest is unique compared to the other classes within the dataset, but it does not achieve an acceptable recall metric. The Data Scientist has already tried varying the number and size of the MLP's hidden layers, which has not significantly improved the results. A solution to improve recall must be implemented as quickly as possible.

Which techniques should be used to meet these requirements?

- A. Gather more data using Amazon Mechanical Turk and then retrain
- B. Train an anomaly detection model instead of an MLP
- C. Train an XGBoost model instead of an MLP
- D. Add class weights to the MLP's loss function and then retrain

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 82

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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 fraudulent.

How should the Specialist frame this business problem?

- A. Streaming classification
- B. Binary classification
- C. Multi-category classification
- D. Regression classification

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 83

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A real estate company wants to create a machine learning model for predicting housing prices based on a historical dataset. The dataset contains 32 features. Which model will meet the business requirement?

- A. Logistic regression
- B. Linear regression
- C. K-means
- D. Principal component analysis (PCA)

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 84

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Machine Learning Specialist is applying a linear least squares regression model to a dataset with 1,000 records and 50 features. Prior to training, the ML Specialist notices that two features are perfectly linearly dependent.

Why could this be an issue for the linear least squares regression model?

- A. It could cause the backpropagation algorithm to fail during training
- B. It could create a singular matrix during optimization, which fails to define a unique solution
- C. It could modify the loss function during optimization, causing it to fail during training
- D. It could introduce non-linear dependencies within the data, which could invalidate the linear assumptions of the model

Show Suggested Answer



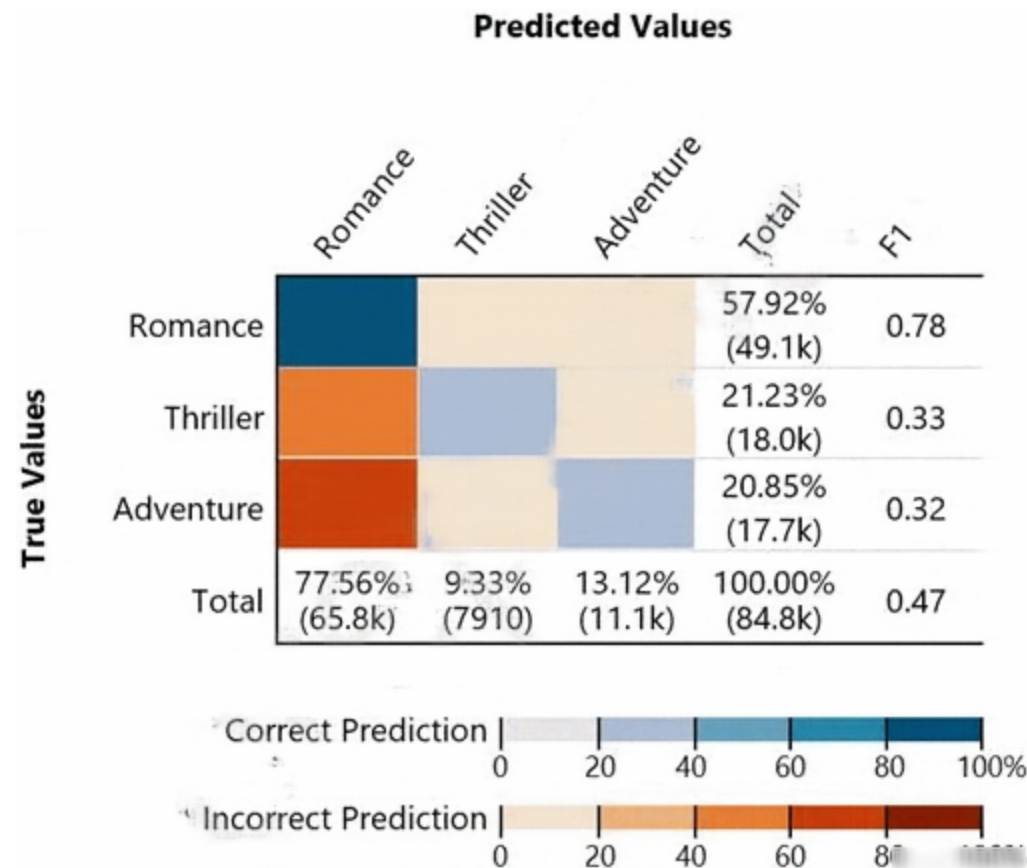
Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 85

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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 $77.56\% \div 0.78$ and the predicted class frequency for Adventure is $20.85\% \div 0.32$

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 86

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Machine Learning Specialist wants to bring a custom algorithm to Amazon SageMaker. The Specialist implements the algorithm in a Docker container supported by Amazon SageMaker.

How should the Specialist package the Docker container so that Amazon SageMaker can launch the training correctly?

- A. Modify the `bash_profile` file in the container and add a bash command to start the training program
- B. Use `CMD config` in the Dockerfile to add the training program as a `CMD` of the image
- C. Configure the training program as an `ENTRYPOINT` named `train`
- D. Copy the training program to directory `/opt/ml/train`

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 87

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Data Scientist needs to analyze employment data. The dataset contains approximately 10 million observations on people across 10 different features. During the preliminary analysis, the Data Scientist notices that income and age distributions are not normal. While income levels shows a right skew as expected, with fewer individuals having a higher income, the age distribution also shows a right skew, with fewer older individuals participating in the workforce.

Which feature transformations can the Data Scientist apply to fix the incorrectly skewed data? (Choose two.)

- A. Cross-validation
- B. Numerical value binning
- C. High-degree polynomial transformation
- D. Logarithmic transformation
- E. One hot encoding

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 88

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 89

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 90

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Machine Learning Specialist is planning to create a long-running Amazon EMR cluster. The EMR cluster will have 1 master node, 10 core nodes, and 20 task nodes. To save on costs, the Specialist will use Spot Instances in the EMR cluster.

Which nodes should the Specialist launch on Spot Instances?

- A. Master node
- B. Any of the core nodes
- C. Any of the task nodes
- D. Both core and task nodes

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 91

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 92

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company wants to predict the sale prices of houses based on available historical sales data. The target variable in the company's dataset is the sale price. The features include parameters such as the lot size, living area measurements, non-living area measurements, number of bedrooms, number of bathrooms, year built, and postal code. The company wants to use multi-variable linear regression to predict house sale prices.

Which step should a machine learning specialist take to remove features that are irrelevant for the analysis and reduce the model's complexity?

- A. Plot a histogram of the features and compute their standard deviation. Remove features with high variance.
- B. Plot a histogram of the features and compute their standard deviation. Remove features with low variance.
- C. Build a heatmap showing the correlation of the dataset against itself. Remove features with low mutual correlation scores.
- D. Run a correlation check of all features against the target variable. Remove features with low target variable correlation scores.

Show Suggested Answer



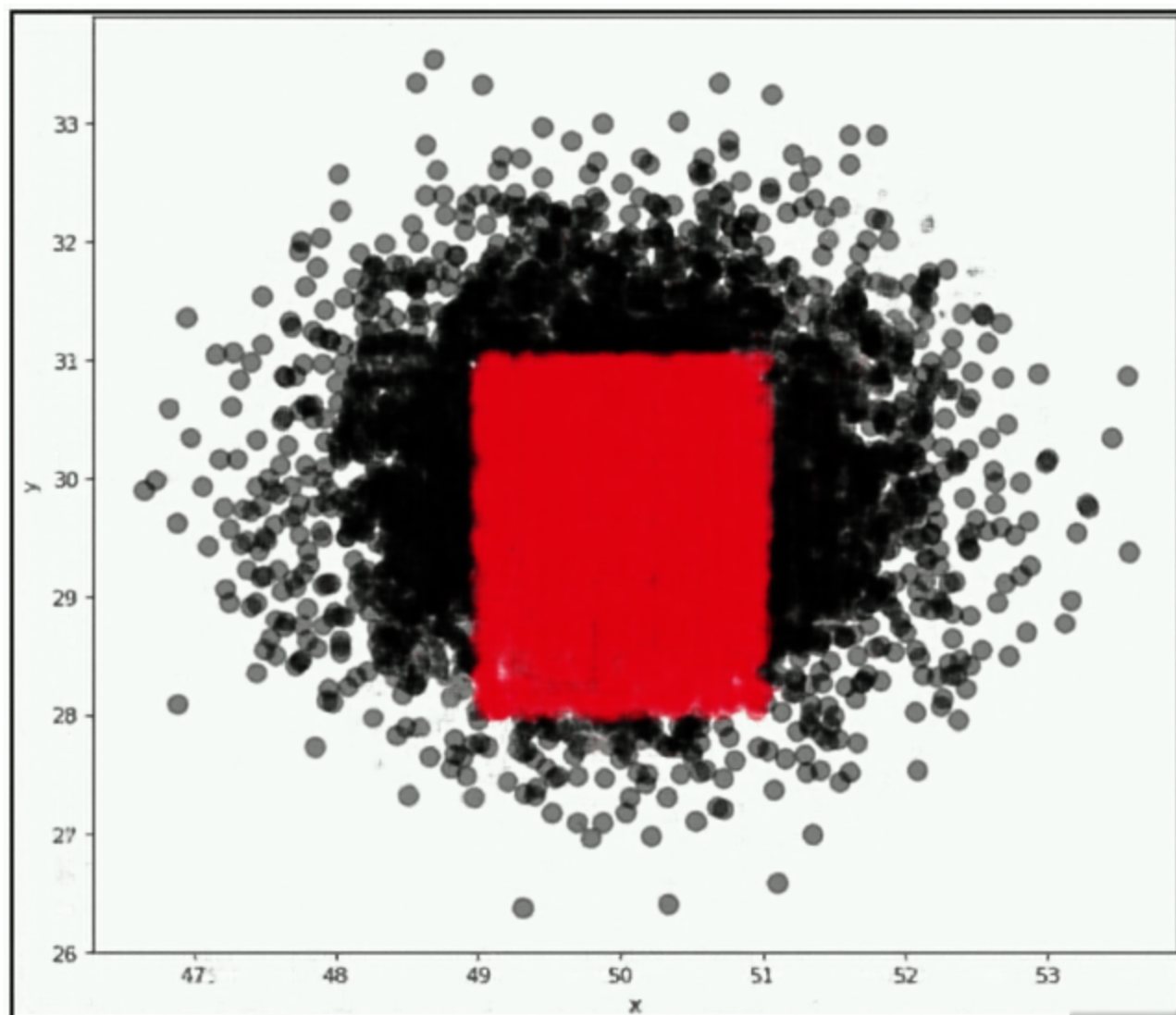
Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 93

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company wants to classify user behavior as either fraudulent or normal. Based on internal research, a machine learning specialist will build a binary classifier based on two features: age of account, denoted by x , and transaction month, denoted by y . The class distributions are illustrated in the provided figure. The positive class is portrayed in red, while the negative class is portrayed in black.



Which model would have the HIGHEST accuracy?

- A. Linear support vector machine (SVM)
- B. Decision tree
- C. Support vector machine (SVM) with a radial basis function kernel
- D. Single perceptron with a Tanh activation function

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 94

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A health care company is planning to use neural networks to classify their X-ray images into normal and abnormal classes. The labeled data is divided into a training set of 1,000 images and a test set of 200 images. The initial training of a neural network model with 50 hidden layers yielded 99% accuracy on the training set, but only 55% accuracy on the test set.

What changes should the Specialist consider to solve this issue? (Choose three.)

- A. Choose a higher number of layers
- B. Choose a lower number of layers
- C. Choose a smaller learning rate
- D. Enable dropout
- E. Include all the images from the test set in the training set
- F. Enable early stopping

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 95

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

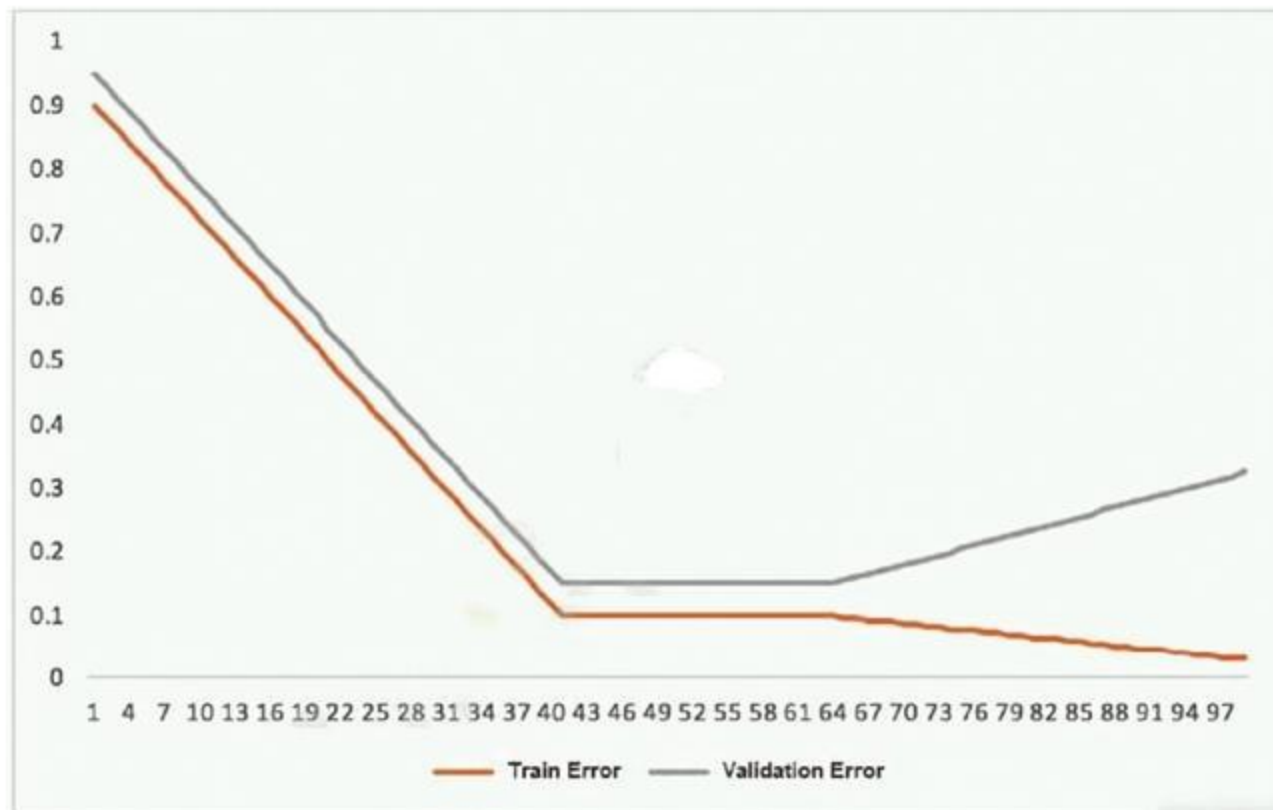
This graph shows the training and validation loss against the epochs for a neural network.

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

Show Suggested Answer

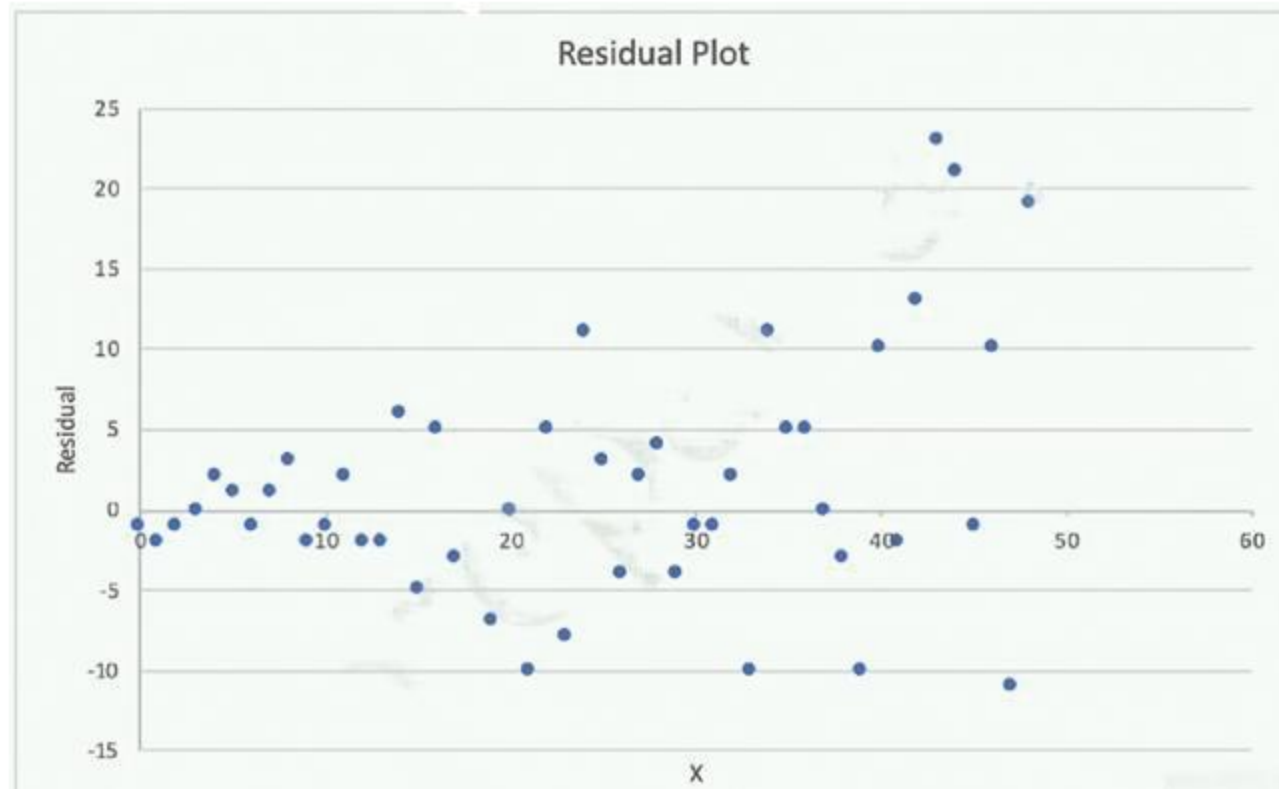
Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 96

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Machine Learning Specialist is attempting to build a linear regression model.



Given the displayed residual plot only, what is the MOST likely problem with the model?

- A. Linear regression is inappropriate. The residuals do not have constant variance.
- B. Linear regression is inappropriate. The underlying data has outliers.
- C. Linear regression is appropriate. The residuals have a zero mean.
- D. Linear regression is appropriate. The residuals have constant variance.

Show Suggested Answer

Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 97

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A large company has developed a BI application that generates reports and dashboards using data collected from various operational metrics. The company wants to provide executives with an enhanced experience so they can use natural language to get data from the reports. The company wants the executives to be able ask questions using written and spoken interfaces.

Which combination of services can be used to build this conversational interface? (Choose three.)

- A. Alexa for Business
- B. Amazon Connect
- C. Amazon Lex
- D. Amazon Polly
- E. Amazon Comprehend
- F. Amazon Transcribe

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 98

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A machine learning specialist works for a fruit processing company and needs to build a system that categorizes apples into three types. The specialist has collected a dataset that contains 150 images for each type of apple and applied transfer learning on a neural network that was pretrained on ImageNet with this dataset.

The company requires at least 85% accuracy to make use of the model.

After an exhaustive grid search, the optimal hyperparameters produced the following:

- ⇒ 68% accuracy on the training set
- ⇒ 67% accuracy on the validation set

What can the machine learning specialist do to improve the system's accuracy?

- A. Upload the model to an Amazon SageMaker notebook instance and use the Amazon SageMaker HPO feature to optimize the model's hyperparameters.
- B. Add more data to the training set and retrain the model using transfer learning to reduce the bias.
- C. Use a neural network model with more layers that are pretrained on ImageNet and apply transfer learning to increase the variance.
- D. Train a new model using the current neural network architecture.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 99

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company uses camera images of the tops of items displayed on store shelves to determine which items were removed and which ones still remain. After several hours of data labeling, the company has a total of 1,000 hand-labeled images covering 10 distinct items. The training results were poor.

Which machine learning approach fulfills the company's long-term needs?

- A. Convert the images to grayscale and retrain the model
- B. Reduce the number of distinct items from 10 to 2, build the model, and iterate
- C. Attach different colored labels to each item, take the images again, and build the model
- D. Augment training data for each item using image variants like inversions and translations, build the model, and iterate.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 100

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Data Scientist is developing a binary classifier to predict whether a patient has a particular disease on a series of test results. The Data Scientist has data on 400 patients randomly selected from the population. The disease is seen in 3% of the population.

Which cross-validation strategy should the Data Scientist adopt?

- A. A k-fold cross-validation strategy with $k=5$
- B. A stratified k-fold cross-validation strategy with $k=5$
- C. A k-fold cross-validation strategy with $k=5$ and 3 repeats
- D. An 80/20 stratified split between training and validation

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 101

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A technology startup is using complex deep neural networks and GPU compute to recommend the company's products to its existing customers based upon each customer's habits and interactions. The solution currently pulls each dataset from an Amazon S3 bucket before loading the data into a TensorFlow model pulled from the company's Git repository that runs locally. This job then runs for several hours while continually outputting its progress to the same S3 bucket. The job can be paused, restarted, and continued at any time in the event of a failure, and is run from a central queue.

Senior managers are concerned about the complexity of the solution's resource management and the costs involved in repeating the process regularly. They ask for the workload to be automated so it runs once a week, starting Monday and completing by the close of business Friday.

Which architecture should be used to scale the solution at the lowest cost?

- A. Implement the solution using AWS Deep Learning Containers and run the container as a job using AWS Batch on a GPU-compatible Spot Instance
- B. Implement the solution using a low-cost GPU-compatible Amazon EC2 instance and use the AWS Instance Scheduler to schedule the task
- C. Implement the solution using AWS Deep Learning Containers, run the workload using AWS Fargate running on Spot Instances, and then schedule the task using the built-in task scheduler
- D. Implement the solution using Amazon ECS running on Spot Instances and schedule the task using the ECS service scheduler

Show Suggested Answer



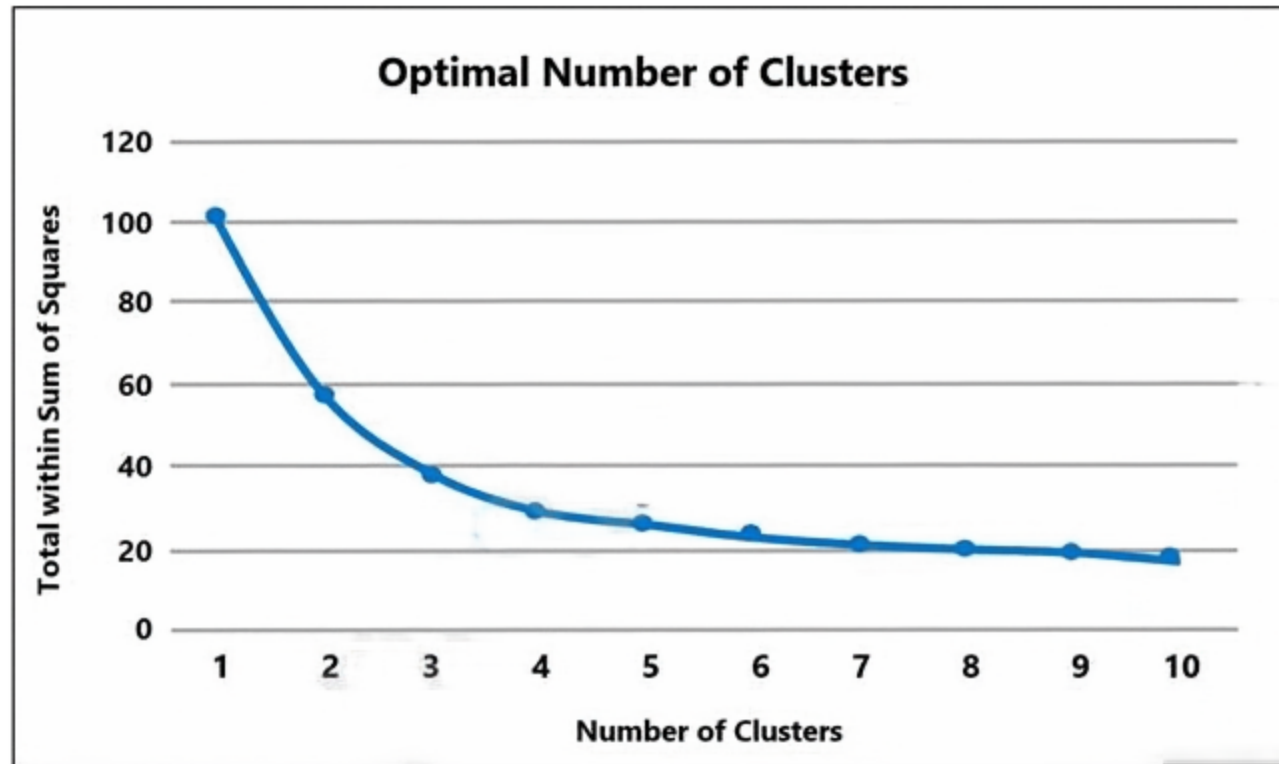
Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 102

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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



Considering the graph, what is a reasonable selection for the optimal choice of k ?

- A. 1
- B. 4
- C. 7
- D. 10

Show Suggested Answer

Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 103

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 104

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 105

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Machine Learning Specialist is deciding between building a naive Bayesian model or a full Bayesian network for a classification problem. The Specialist computes the Pearson correlation coefficients between each feature and finds that their absolute values range between 0.1 to 0.95.

Which model describes the underlying data in this situation?

- A. A naive Bayesian model, since the features are all conditionally independent.
- B. A full Bayesian network, since the features are all conditionally independent.
- C. A naive Bayesian model, since some of the features are statistically dependent.
- D. A full Bayesian network, since some of the features are statistically dependent.

Show Suggested Answer



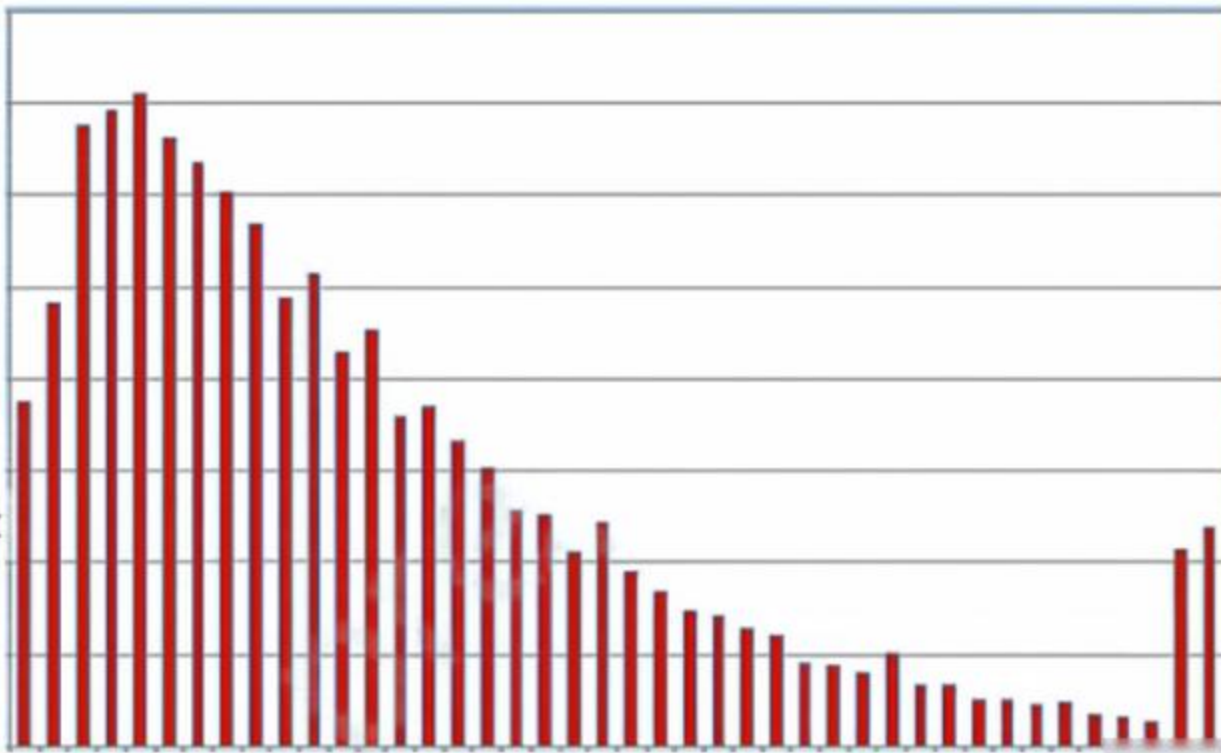
Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 106

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Data Scientist is building a linear regression model and will use resulting p-values to evaluate the statistical significance of each coefficient. Upon inspection of the dataset, the Data Scientist discovers that most of the features are normally distributed. The plot of one feature in the dataset is shown in the graphic.



What transformation should the Data Scientist apply to satisfy the statistical assumptions of the linear regression model?

- A. Exponential transformation
- B. Logarithmic transformation
- C. Polynomial transformation
- D. Sinusoidal transformation

Show Suggested Answer

Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 107

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Machine Learning Specialist is assigned to a Fraud Detection team and must tune an XGBoost model, which is working appropriately for test data. However, with unknown data, it is not working as expected. The existing parameters are provided as follows.

```
param = {  
    'eta': 0.05, # the training step for each iteration  
    'silent': 1, # logging mode - quiet  
    'n_estimators': 2000,  
    'max_depth': 30,  
    'min_child_weight': 3,  
    'gamma': 0,  
    'subsample': 0.8,  
    'objective': 'multi:softprob', # error evaluation for multiclass training  
    'num_class': 201} # the number of classes that exist in this dataset  
num_round = 60 # the number of training iterations
```

Which parameter tuning guidelines should the Specialist follow to avoid overfitting?

- A. Increase the max_depth parameter value.
- B. Lower the max_depth parameter value.
- C. Update the objective to binary:logistic.
- D. Lower the min_child_weight parameter value.

Show Suggested Answer

Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 108

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A data scientist is developing a pipeline to ingest streaming web traffic data. The data scientist needs to implement a process to identify unusual web traffic patterns as part of the pipeline. The patterns will be used downstream for alerting and incident response. The data scientist has access to unlabeled historic data to use, if needed.

The solution needs to do the following:

⇒ Calculate an anomaly score for each web traffic entry.

Adapt unusual event identification to changing web patterns over time.

▪

Which approach should the data scientist implement to meet these requirements?

- A. Use historic web traffic data to train an anomaly detection model using the Amazon SageMaker Random Cut Forest (RCF) built-in model. Use an Amazon Kinesis Data Stream to process the incoming web traffic data. Attach a preprocessing AWS Lambda function to perform data enrichment by calling the RCF model to calculate the anomaly score for each record.
- B. Use historic web traffic data to train an anomaly detection model using the Amazon SageMaker built-in XGBoost model. Use an Amazon Kinesis Data Stream to process the incoming web traffic data. Attach a preprocessing AWS Lambda function to perform data enrichment by calling the XGBoost model to calculate the anomaly score for each record.
- C. Collect the streaming data using Amazon Kinesis Data Firehose. Map the delivery stream as an input source for Amazon Kinesis Data Analytics. Write a SQL query to run in real time against the streaming data with the k-Nearest Neighbors (kNN) SQL extension to calculate anomaly scores for each record using a tumbling window.
- D. Collect the streaming data using Amazon Kinesis Data Firehose. Map the delivery stream as an input source for Amazon Kinesis Data Analytics. Write a SQL query to run in real time against the streaming data with the Amazon Random Cut Forest (RCF) SQL extension to calculate anomaly scores for each record using a sliding window.

Show Suggested Answer

Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 109

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Data Scientist received a set of insurance records, each consisting of a record ID, the final outcome among 200 categories, and the date of the final outcome. Some partial information on claim contents is also provided, but only for a few of the 200 categories. For each outcome category, there are hundreds of records distributed over the past 3 years. The Data Scientist wants to predict how many claims to expect in each category from month to month, a few months in advance. What type of machine learning model should be used?

- A. Classification month-to-month using supervised learning of the 200 categories based on claim contents.
- B. Reinforcement learning using claim IDs and timestamps where the agent will identify how many claims in each category to expect from month to month.
- C. Forecasting using claim IDs and timestamps to identify how many claims in each category to expect from month to month.
- D. Classification with supervised learning of the categories for which partial information on claim contents is provided, and forecasting using claim IDs and timestamps for all other categories.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 110

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company that promotes healthy sleep patterns by providing cloud-connected devices currently hosts a sleep tracking application on AWS. The application collects device usage information from device users. The company's Data Science team is building a machine learning model to predict if and when a user will stop utilizing the company's devices. Predictions from this model are used by a downstream application that determines the best approach for contacting users.

The Data Science team is building multiple versions of the machine learning model to evaluate each version against the company's business goals. To measure long-term effectiveness, the team wants to run multiple versions of the model in parallel for long periods of time, with the ability to control the portion of inferences served by the models.

Which solution satisfies these requirements with MINIMAL effort?

- A. Build and host multiple models in Amazon SageMaker. Create multiple Amazon SageMaker endpoints, one for each model. Programmatically control invoking different models for inference at the application layer.
- B. Build and host multiple models in Amazon SageMaker. Create an Amazon SageMaker endpoint configuration with multiple production variants. Programmatically control the portion of the inferences served by the multiple models by updating the endpoint configuration.
- C. Build and host multiple models in Amazon SageMaker Neo to take into account different types of medical devices. Programmatically control which model is invoked for inference based on the medical device type.
- D. Build and host multiple models in Amazon SageMaker. Create a single endpoint that accesses multiple models. Use Amazon SageMaker batch transform to control invoking the different models through the single endpoint.

Show Suggested Answer

Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 111

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

An agricultural company is interested in using machine learning to detect specific types of weeds in a 100-acre grassland field. Currently, the company uses tractor-mounted cameras to capture multiple images of the field as 10 × 10 grids. The company also has a large training dataset that consists of annotated images of popular weed classes like broadleaf and non-broadleaf docks.

The company wants to build a weed detection model that will detect specific types of weeds and the location of each type within the field. Once the model is ready, it will be hosted on Amazon SageMaker endpoints. The model will perform real-time inferencing using the images captured by the cameras.

Which approach should a Machine Learning Specialist take to obtain accurate predictions?

- A. Prepare the images in RecordIO format and upload them to Amazon S3. Use Amazon SageMaker to train, test, and validate the model using an image classification algorithm to categorize images into various weed classes.
- B. Prepare the images in Apache Parquet format and upload them to Amazon S3. Use Amazon SageMaker to train, test, and validate the model using an object-detection single-shot multibox detector (SSD) algorithm.
- C. Prepare the images in RecordIO format and upload them to Amazon S3. Use Amazon SageMaker to train, test, and validate the model using an object-detection single-shot multibox detector (SSD) algorithm.
- D. Prepare the images in Apache Parquet format and upload them to Amazon S3. Use Amazon SageMaker to train, test, and validate the model using an image classification algorithm to categorize images into various weed classes.

Show Suggested Answer

Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 112

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 113

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A Machine Learning Specialist is designing a scalable data storage solution for Amazon SageMaker. There is an existing TensorFlow-based model implemented as a train.py script that relies on static training data that is currently stored as TFRecords.

Which method of providing training data to Amazon SageMaker would meet the business requirements with the LEAST development overhead?

- A. Use Amazon SageMaker script mode and use train.py unchanged. Point the Amazon SageMaker training invocation to the local path of the data without reformatting the training data.
- B. Use Amazon SageMaker script mode and use train.py unchanged. Put the TFRecord data into an Amazon S3 bucket. Point the Amazon SageMaker training invocation to the S3 bucket without reformatting the training data.
- C. Rewrite the train.py script to add a section that converts TFRecords to protobuf and ingests the protobuf data instead of TFRecords.
- D. Prepare the data in the format accepted by Amazon SageMaker. Use AWS Glue or AWS Lambda to reformat and store the data in an Amazon S3 bucket.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 114

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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)

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 115

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A retail company is using Amazon Personalize to provide personalized product recommendations for its customers during a marketing campaign. The company sees a significant increase in sales of recommended items to existing customers immediately after deploying a new solution version, but these sales decrease a short time after deployment. Only historical data from before the marketing campaign is available for training.

How should a data scientist adjust the solution?

- A. Use the event tracker in Amazon Personalize to include real-time user interactions.
- B. Add user metadata and use the HRNN-Metadata recipe in Amazon Personalize.
- C. Implement a new solution using the built-in factorization machines (FM) algorithm in Amazon SageMaker.
- D. Add event type and event value fields to the interactions dataset in Amazon Personalize.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 116

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 117

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

An e-commerce company wants to launch a new cloud-based product recommendation feature for its web application. Due to data localization regulations, any sensitive data must not leave its on-premises data center, and the product recommendation model must be trained and tested using nonsensitive data only. Data transfer to the cloud must use IPsec. The web application is hosted on premises with a PostgreSQL database that contains all the data. The company wants the data to be uploaded securely to Amazon S3 each day for model retraining.

How should a machine learning specialist meet these requirements?

- A. Create an AWS Glue job to connect to the PostgreSQL DB instance. Ingest tables without sensitive data through an AWS Site-to-Site VPN connection directly into Amazon S3.
- B. Create an AWS Glue job to connect to the PostgreSQL DB instance. Ingest all data through an AWS Site-to-Site VPN connection into Amazon S3 while removing sensitive data using a PySpark job.
- C. Use AWS Database Migration Service (AWS DMS) with table mapping to select PostgreSQL tables with no sensitive data through an SSL connection. Replicate data directly into Amazon S3.
- D. Use PostgreSQL logical replication to replicate all data to PostgreSQL in Amazon EC2 through AWS Direct Connect with a VPN connection. Use AWS Glue to move data from Amazon EC2 to Amazon S3.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 118

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 119

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A data scientist wants to use Amazon Forecast to build a forecasting model for inventory demand for a retail company. The company has provided a dataset of historic inventory demand for its products as a .csv file stored in an Amazon S3 bucket. The table below shows a sample of the dataset.

timestamp	item_id	demand	category	lead_time
2019-12-14	uni_000736	120	hardware	90
2020-01-31	uni_003429	98	hardware	30
2020-03-04	uni_000211	234	accessories	10

How should the data scientist transform the data?

- A. Use ETL jobs in AWS Glue to separate the dataset into a target time series dataset and an item metadata dataset. Upload both datasets as .csv files to Amazon S3.
- B. Use a Jupyter notebook in Amazon SageMaker to separate the dataset into a related time series dataset and an item metadata dataset. Upload both datasets as tables in Amazon Aurora.
- C. Use AWS Batch jobs to separate the dataset into a target time series dataset, a related time series dataset, and an item metadata dataset. Upload them directly to Forecast from a local machine.
- D. Use a Jupyter notebook in Amazon SageMaker to transform the data into the optimized protobuf recordIO format. Upload the dataset in this format to Amazon S3.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 120

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A machine learning specialist is running an Amazon SageMaker endpoint using the built-in object detection algorithm on a P3 instance for real-time predictions in a company's production application. When evaluating the model's resource utilization, the specialist notices that the model is using only a fraction of the GPU. Which architecture changes would ensure that provisioned resources are being utilized effectively?

- A. Redeploy the model as a batch transform job on an M5 instance.
- B. Redeploy the model on an M5 instance. Attach Amazon Elastic Inference to the instance.
- C. Redeploy the model on a P3dn instance.
- D. Deploy the model onto an Amazon Elastic Container Service (Amazon ECS) cluster using a P3 instance.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 121

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A data scientist uses an Amazon SageMaker notebook instance to conduct data exploration and analysis. This requires certain Python packages that are not natively available on Amazon SageMaker to be installed on the notebook instance.

How can a machine learning specialist ensure that required packages are automatically available on the notebook instance for the data scientist to use?

- A. Install AWS Systems Manager Agent on the underlying Amazon EC2 instance and use Systems Manager Automation to execute the package installation commands.
- B. Create a Jupyter notebook file (.ipynb) with cells containing the package installation commands to execute and place the file under the /etc/init directory of each Amazon SageMaker notebook instance.
- C. Use the conda package manager from within the Jupyter notebook console to apply the necessary conda packages to the default kernel of the notebook.
- D. Create an Amazon SageMaker lifecycle configuration with package installation commands and assign the lifecycle configuration to the notebook instance.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 122

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A data scientist needs to identify fraudulent user accounts for a company's ecommerce platform. The company wants the ability to determine if a newly created account is associated with a previously known fraudulent user. The data scientist is using AWS Glue to cleanse the company's application logs during ingestion.

Which strategy will allow the data scientist to identify fraudulent accounts?

- A. Execute the built-in FindDuplicates Amazon Athena query.
- B. Create a FindMatches machine learning transform in AWS Glue.
- C. Create an AWS Glue crawler to infer duplicate accounts in the source data.
- D. Search for duplicate accounts in the AWS Glue Data Catalog.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 123

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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 needs to reduce the number of false negatives.

	Predicted 0	Predicted 1
Actual 0	99,966	34
Actual 1	877	123

Which combination of steps should the Data Scientist take to reduce the number of false negative predictions by the model? (Choose two.)

- A. Change the XGBoost `eval_metric` parameter to optimize based on Root Mean Square Error (RMSE).
- 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 Area Under the ROC Curve (AUC).
- E. Decrease the XGBoost `max_depth` parameter because the model is currently overfitting the data.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 124

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A data scientist has developed a machine learning translation model for English to Japanese by using Amazon SageMaker's built-in seq2seq algorithm with 500,000 aligned sentence pairs. While testing with sample sentences, the data scientist finds that the translation quality is reasonable for an example as short as five words. However, the quality becomes unacceptable if the sentence is 100 words long.

Which action will resolve the problem?

- A. Change preprocessing to use n-grams.
- B. Add more nodes to the recurrent neural network (RNN) than the largest sentence's word count.
- C. Adjust hyperparameters related to the attention mechanism.
- D. Choose a different weight initialization type.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 125

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A financial company is trying to detect credit card fraud. The company observed that, on average, 2% of credit card transactions were fraudulent. A data scientist trained a classifier on a year's worth of credit card transactions data. The model needs to identify the fraudulent transactions (positives) from the regular ones (negatives). The company's goal is to accurately capture as many positives as possible.

Which metrics should the data scientist use to optimize the model? (Choose two.)

- A. Specificity
- B. False positive rate
- C. Accuracy
- D. Area under the precision-recall curve
- E. True positive rate

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 126

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A machine learning specialist is developing a proof of concept for government users whose primary concern is security. The specialist is using Amazon SageMaker to train a convolutional neural network (CNN) model for a photo classifier application. The specialist wants to protect the data so that it cannot be accessed and transferred to a remote host by malicious code accidentally installed on the training container.

Which action will provide the MOST secure protection?

- A. Remove Amazon S3 access permissions from the SageMaker execution role.
- B. Encrypt the weights of the CNN model.
- C. Encrypt the training and validation dataset.
- D. Enable network isolation for training jobs.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 127

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A medical imaging company wants to train a computer vision model to detect areas of concern on patients' CT scans. The company has a large collection of unlabeled CT scans that are linked to each patient and stored in an Amazon S3 bucket. The scans must be accessible to authorized users only. A machine learning engineer needs to build a labeling pipeline.

Which set of steps should the engineer take to build the labeling pipeline with the LEAST effort?

- A. Create a workforce with AWS Identity and Access Management (IAM). Build a labeling tool on Amazon EC2 Queue images for labeling by using Amazon Simple Queue Service (Amazon SQS). Write the labeling instructions.
- B. Create an Amazon Mechanical Turk workforce and manifest file. Create a labeling job by using the built-in image classification task type in Amazon SageMaker Ground Truth. Write the labeling instructions.
- C. Create a private workforce and manifest file. Create a labeling job by using the built-in bounding box task type in Amazon SageMaker Ground Truth. Write the labeling instructions.
- D. Create a workforce with Amazon Cognito. Build a labeling web application with AWS Amplify. Build a labeling workflow backend using AWS Lambda. Write the labeling instructions.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 128

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company is using Amazon Textract to extract textual data from thousands of scanned text-heavy legal documents daily. The company uses this information to process loan applications automatically. Some of the documents fail business validation and are returned to human reviewers, who investigate the errors. This activity increases the time to process the loan applications.

What should the company do to reduce the processing time of loan applications?

- A. Configure Amazon Textract to route low-confidence predictions to Amazon SageMaker Ground Truth. Perform a manual review on those words before performing a business validation.
- B. Use an Amazon Textract synchronous operation instead of an asynchronous operation.
- C. Configure Amazon Textract to route low-confidence predictions to Amazon Augmented AI (Amazon A2I). Perform a manual review on those words before performing a business validation.
- D. Use Amazon Rekognition's feature to detect text in an image to extract the data from scanned images. Use this information to process the loan applications.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 129

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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).

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 130

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A data scientist must build a custom recommendation model in Amazon SageMaker for an online retail company. Due to the nature of the company's products, customers buy only 4-5 products every 5-10 years. So, the company relies on a steady stream of new customers. When a new customer signs up, the company collects data on the customer's preferences. Below is a sample of the data available to the data scientist.

timestamp	user_id	product_id	preference_1	...	preference_10
2020-03-04	90	25	0.374	...	0.374
2020-03-04	90	61	0	...	0.374
2020-02-21	203	56	1	...	0.098

How should the data scientist split the dataset into a training and test set for this use case?

- A. Shuffle all interaction data. Split off the last 10% of the interaction data for the test set.
- B. Identify the most recent 10% of interactions for each user. Split off these interactions for the test set.
- C. Identify the 10% of users with the least interaction data. Split off all interaction data from these users for the test set.
- D. Randomly select 10% of the users. Split off all interaction data from these users for the test set.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 131

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A financial services company wants to adopt Amazon SageMaker as its default data science environment. The company's data scientists run machine learning (ML) models on confidential financial data. The company is worried about data egress and wants an ML engineer to secure the environment.

Which mechanisms can the ML engineer use to control data egress from SageMaker? (Choose three.)

- A. Connect to SageMaker by using a VPC interface endpoint powered by AWS PrivateLink.
- B. Use SCPs to restrict access to SageMaker.
- C. Disable root access on the SageMaker notebook instances.
- D. Enable network isolation for training jobs and models.
- E. Restrict notebook presigned URLs to specific IPs used by the company.
- F. Protect data with encryption at rest and in transit. Use AWS Key Management Service (AWS KMS) to manage encryption keys.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 132

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company needs to quickly make sense of a large amount of data and gain insight from it. The data is in different formats, the schemas change frequently, and new data sources are added regularly. The company wants to use AWS services to explore multiple data sources, suggest schemas, and enrich and transform the data. The solution should require the least possible coding effort for the data flows and the least possible infrastructure management.

Which combination of AWS services will meet these requirements?

A.

- ⇒ Amazon EMR for data discovery, enrichment, and transformation
- ⇒ Amazon Athena for querying and analyzing the results in Amazon S3 using standard SQL
- ⇒ Amazon QuickSight for reporting and getting insights

B.

- ⇒ Amazon Kinesis Data Analytics for data ingestion
- ⇒ Amazon EMR for data discovery, enrichment, and transformation
- ⇒ Amazon Redshift for querying and analyzing the results in Amazon S3

C.

- ⇒ AWS Glue for data discovery, enrichment, and transformation
- ⇒ Amazon Athena for querying and analyzing the results in Amazon S3 using standard SQL
- ⇒ Amazon QuickSight for reporting and getting insights

D.

- ⇒ AWS Data Pipeline for data transfer
- ⇒ AWS Step Functions for orchestrating AWS Lambda jobs for data discovery, enrichment, and transformation
- ⇒ Amazon Athena for querying and analyzing the results in Amazon S3 using standard SQL
- ⇒ Amazon QuickSight for reporting and getting insights

Show Suggested Answer

Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 133

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 134

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company is building a predictive maintenance model based on machine learning (ML). The data is stored in a fully private Amazon S3 bucket that is encrypted at rest with AWS Key Management Service (AWS KMS) CMKs. An ML specialist must run data preprocessing by using an Amazon SageMaker Processing job that is triggered from code in an Amazon SageMaker notebook. The job should read data from Amazon S3, process it, and upload it back to the same S3 bucket.

The preprocessing code is stored in a container image in Amazon Elastic Container Registry (Amazon ECR). The ML specialist needs to grant permissions to ensure a smooth data preprocessing workflow.

Which set of actions should the ML specialist take to meet these requirements?

- A. Create an IAM role that has permissions to create Amazon SageMaker Processing jobs, S3 read and write access to the relevant S3 bucket, and appropriate KMS and ECR permissions. Attach the role to the SageMaker notebook instance. Create an Amazon SageMaker Processing job from the notebook.
- B. Create an IAM role that has permissions to create Amazon SageMaker Processing jobs. Attach the role to the SageMaker notebook instance. Create an Amazon SageMaker Processing job with an IAM role that has read and write permissions to the relevant S3 bucket, and appropriate KMS and ECR permissions.
- C. Create an IAM role that has permissions to create Amazon SageMaker Processing jobs and to access Amazon ECR. Attach the role to the SageMaker notebook instance. Set up both an S3 endpoint and a KMS endpoint in the default VPC. Create Amazon SageMaker Processing jobs from the notebook.
- D. Create an IAM role that has permissions to create Amazon SageMaker Processing jobs. Attach the role to the SageMaker notebook instance. Set up an S3 endpoint in the default VPC. Create Amazon SageMaker Processing jobs with the access key and secret key of the IAM user with appropriate KMS and ECR permissions.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 135

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A data scientist has been running an Amazon SageMaker notebook instance for a few weeks. During this time, a new version of Jupyter Notebook was released along with additional software updates. The security team mandates that all running SageMaker notebook instances use the latest security and software updates provided by SageMaker.

How can the data scientist meet this requirements?

- A. Call the CreateNotebookInstanceLifecycleConfig API operation
- B. Create a new SageMaker notebook instance and mount the Amazon Elastic Block Store (Amazon EBS) volume from the original instance
- C. Stop and then restart the SageMaker notebook instance
- D. Call the UpdateNotebookInstanceLifecycleConfig API operation

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 136

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A library is developing an automatic book-borrowing system that uses Amazon Rekognition. Images of library members' faces are stored in an Amazon S3 bucket. When members borrow books, the Amazon Rekognition CompareFaces API operation compares real faces against the stored faces in Amazon S3. The library needs to improve security by making sure that images are encrypted at rest. Also, when the images are used with Amazon Rekognition, they need to be encrypted in transit. The library also must ensure that the images are not used to improve Amazon Rekognition as a service. How should a machine learning specialist architect the solution to satisfy these requirements?

- A. Enable server-side encryption on the S3 bucket. Submit an AWS Support ticket to opt out of allowing images to be used for improving the service, and follow the process provided by AWS Support.
- B. Switch to using an Amazon Rekognition collection to store the images. Use the IndexFaces and SearchFacesByImage API operations instead of the CompareFaces API operation.
- C. Switch to using the AWS GovCloud (US) Region for Amazon S3 to store images and for Amazon Rekognition to compare faces. Set up a VPN connection and only call the Amazon Rekognition API operations through the VPN.
- D. Enable client-side encryption on the S3 bucket. Set up a VPN connection and only call the Amazon Rekognition API operations through the VPN.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 137

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 138

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company has set up and deployed its machine learning (ML) model into production with an endpoint using Amazon SageMaker hosting services. The ML team has configured automatic scaling for its SageMaker instances to support workload changes. During testing, the team notices that additional instances are being launched before the new instances are ready. This behavior needs to change as soon as possible.

How can the ML team solve this issue?

- A. Decrease the cooldown period for the scale-in activity. Increase the configured maximum capacity of instances.
- B. Replace the current endpoint with a multi-model endpoint using SageMaker.
- C. Set up Amazon API Gateway and AWS Lambda to trigger the SageMaker inference endpoint.
- D. Increase the cooldown period for the scale-out activity.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 139

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A telecommunications company is developing a mobile app for its customers. The company is using an Amazon SageMaker hosted endpoint for machine learning model inferences.

Developers want to introduce a new version of the model for a limited number of users who subscribed to a preview feature of the app. After the new version of the model is tested as a preview, developers will evaluate its accuracy. If a new version of the model has better accuracy, developers need to be able to gradually release the new version for all users over a fixed period of time.

How can the company implement the testing model with the LEAST amount of operational overhead?

- A. Update the ProductionVariant data type with the new version of the model by using the CreateEndpointConfig operation with the InitialVariantWeight parameter set to 0. Specify the TargetVariant parameter for InvokeEndpoint calls for users who subscribed to the preview feature. When the new version of the model is ready for release, gradually increase InitialVariantWeight until all users have the updated version.
- B. Configure two SageMaker hosted endpoints that serve the different versions of the model. Create an Application Load Balancer (ALB) to route traffic to both endpoints based on the TargetVariant query string parameter. Reconfigure the app to send the TargetVariant query string parameter for users who subscribed to the preview feature. When the new version of the model is ready for release, change the ALB's routing algorithm to weighted until all users have the updated version.
- C. Update the DesiredWeightsAndCapacity data type with the new version of the model by using the UpdateEndpointWeightsAndCapacities operation with the DesiredWeight parameter set to 0. Specify the TargetVariant parameter for InvokeEndpoint calls for users who subscribed to the preview feature. When the new version of the model is ready for release, gradually increase DesiredWeight until all users have the updated version.
- D. Configure two SageMaker hosted endpoints that serve the different versions of the model. Create an Amazon Route 53 record that is configured with a simple routing policy and that points to the current version of the model. Configure the mobile app to use the endpoint URL for users who subscribed to the preview feature and to use the Route 53 record for other users. When the new version of the model is ready for release, add a new model version endpoint to Route 53, and switch the policy to weighted until all users have the updated version.

Show Suggested Answer

Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 140

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company offers an online shopping service to its customers. The company wants to enhance the site's security by requesting additional information when customers access the site from locations that are different from their normal location. The company wants to update the process to call a machine learning (ML) model to determine when additional information should be requested.

The company has several terabytes of data from its existing ecommerce web servers containing the source IP addresses for each request made to the web server. For authenticated requests, the records also contain the login name of the requesting user.

Which approach should an ML specialist take to implement the new security feature in the web application?

- A. Use Amazon SageMaker Ground Truth to label each record as either a successful or failed access attempt. Use Amazon SageMaker to train a binary classification model using the factorization machines (FM) algorithm.
- B. Use Amazon SageMaker to train a model using the IP Insights algorithm. Schedule updates and retraining of the model using new log data nightly.
- C. Use Amazon SageMaker Ground Truth to label each record as either a successful or failed access attempt. Use Amazon SageMaker to train a binary classification model using the IP Insights algorithm.
- D. Use Amazon SageMaker to train a model using the Object2Vec algorithm. Schedule updates and retraining of the model using new log data nightly.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 141

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A retail company wants to combine its customer orders with the product description data from its product catalog. The structure and format of the records in each dataset is different. A data analyst tried to use a spreadsheet to combine the datasets, but the effort resulted in duplicate records and records that were not properly combined. The company needs a solution that it can use to combine similar records from the two datasets and remove any duplicates.

Which solution will meet these requirements?

- A. Use an AWS Lambda function to process the data. Use two arrays to compare equal strings in the fields from the two datasets and remove any duplicates.
- B. Create AWS Glue crawlers for reading and populating the AWS Glue Data Catalog. Call the AWS Glue SearchTables API operation to perform a fuzzy- matching search on the two datasets, and cleanse the data accordingly.
- C. Create AWS Glue crawlers for reading and populating the AWS Glue Data Catalog. Use the FindMatches transform to cleanse the data.
- D. Create an AWS Lake Formation custom transform. Run a transformation for matching products from the Lake Formation console to cleanse the data automatically.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 142

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company provisions Amazon SageMaker notebook instances for its data science team and creates Amazon VPC interface endpoints to ensure communication between the VPC and the notebook instances. All connections to the Amazon SageMaker API are contained entirely and securely using the AWS network.

However, the data science team realizes that individuals outside the VPC can still connect to the notebook instances across the internet.

Which set of actions should the data science team take to fix the issue?

- A. Modify the notebook instances' security group to allow traffic only from the CIDR ranges of the VPC. Apply this security group to all of the notebook instances' VPC interfaces.
- B. Create an IAM policy that allows the `sagemaker:CreatePresignedNotebookInstanceUrl` and `sagemaker:DescribeNotebookInstance` actions from only the VPC endpoints. Apply this policy to all IAM users, groups, and roles used to access the notebook instances.
- C. Add a NAT gateway to the VPC. Convert all of the subnets where the Amazon SageMaker notebook instances are hosted to private subnets. Stop and start all of the notebook instances to reassign only private IP addresses.
- D. Change the network ACL of the subnet the notebook is hosted in to restrict access to anyone outside the VPC.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 143

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company will use Amazon SageMaker to train and host a machine learning (ML) model for a marketing campaign. The majority of data is sensitive customer data. The data must be encrypted at rest. The company wants AWS to maintain the root of trust for the master keys and wants encryption key usage to be logged.

Which implementation will meet these requirements?

- A. Use encryption keys that are stored in AWS Cloud HSM to encrypt the ML data volumes, and to encrypt the model artifacts and data in Amazon S3.
- B. Use SageMaker built-in transient keys to encrypt the ML data volumes. Enable default encryption for new Amazon Elastic Block Store (Amazon EBS) volumes.
- C. Use customer managed keys in AWS Key Management Service (AWS KMS) to encrypt the ML data volumes, and to encrypt the model artifacts and data in Amazon S3.
- D. Use AWS Security Token Service (AWS STS) to create temporary tokens to encrypt the ML storage volumes, and to encrypt the model artifacts and data in Amazon S3.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 144

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A machine learning specialist stores IoT soil sensor data in Amazon DynamoDB table and stores weather event data as JSON files in Amazon S3. The dataset in DynamoDB is 10 GB in size and the dataset in Amazon S3 is 5 GB in size. The specialist wants to train a model on this data to help predict soil moisture levels as a function of weather events using Amazon SageMaker.

Which solution will accomplish the necessary transformation to train the Amazon SageMaker model with the LEAST amount of administrative overhead?

- A. Launch an Amazon EMR cluster. Create an Apache Hive external table for the DynamoDB table and S3 data. Join the Hive tables and write the results out to Amazon S3.
- B. Crawl the data using AWS Glue crawlers. Write an AWS Glue ETL job that merges the two tables and writes the output to an Amazon Redshift cluster.
- C. Enable Amazon DynamoDB Streams on the sensor table. Write an AWS Lambda function that consumes the stream and appends the results to the existing weather files in Amazon S3.
- D. Crawl the data using AWS Glue crawlers. Write an AWS Glue ETL job that merges the two tables and writes the output in CSV format to Amazon S3.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 145

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company sells thousands of products on a public website and wants to automatically identify products with potential durability problems. The company has 1.000 reviews with date, star rating, review text, review summary, and customer email fields, but many reviews are incomplete and have empty fields. Each review has already been labeled with the correct durability result.

A machine learning specialist must train a model to identify reviews expressing concerns over product durability. The first model needs to be trained and ready to review in 2 days.

What is the MOST direct approach to solve this problem within 2 days?

- A. Train a custom classifier by using Amazon Comprehend.
- B. Build a recurrent neural network (RNN) in Amazon SageMaker by using Gluon and Apache MXNet.
- C. Train a built-in BlazingText model using Word2Vec mode in Amazon SageMaker.
- D. Use a built-in seq2seq model in Amazon SageMaker.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 146

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company that runs an online library is implementing a chatbot using Amazon Lex to provide book recommendations based on category. This intent is fulfilled by an AWS Lambda function that queries an Amazon DynamoDB table for a list of book titles, given a particular category. For testing, there are only three categories implemented as the custom slot types: "comedy," "adventure," and "documentary."

A machine learning (ML) specialist notices that sometimes the request cannot be fulfilled because Amazon Lex cannot understand the category spoken by users with utterances such as "funny," "fun," and "humor." The ML specialist needs to fix the problem without changing the Lambda code or data in DynamoDB.

How should the ML specialist fix the problem?

- A. Add the unrecognized words in the enumeration values list as new values in the slot type.
- B. Create a new custom slot type, add the unrecognized words to this slot type as enumeration values, and use this slot type for the slot.
- C. Use the AMAZON.SearchQuery built-in slot types for custom searches in the database.
- D. Add the unrecognized words as synonyms in the custom slot type.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 147

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A manufacturing company uses machine learning (ML) models to detect quality issues. The models use images that are taken of the company's product at the end of each production step. The company has thousands of machines at the production site that generate one image per second on average.

The company ran a successful pilot with a single manufacturing machine. For the pilot, ML specialists used an industrial PC that ran AWS IoT Greengrass with a long-running AWS Lambda function that uploaded the images to Amazon S3. The uploaded images invoked a Lambda function that was written in Python to perform inference by using an Amazon SageMaker endpoint that ran a custom model. The inference results were forwarded back to a web service that was hosted at the production site to prevent faulty products from being shipped.

The company scaled the solution out to all manufacturing machines by installing similarly configured industrial PCs on each production machine. However, latency for predictions increased beyond acceptable limits. Analysis shows that the internet connection is at its capacity limit.

How can the company resolve this issue MOST cost-effectively?

- A. Set up a 10 Gbps AWS Direct Connect connection between the production site and the nearest AWS Region. Use the Direct Connect connection to upload the images. Increase the size of the instances and the number of instances that are used by the SageMaker endpoint.
- B. Extend the long-running Lambda function that runs on AWS IoT Greengrass to compress the images and upload the compressed files to Amazon S3. Decompress the files by using a separate Lambda function that invokes the existing Lambda function to run the inference pipeline.
- C. Use auto scaling for SageMaker. Set up an AWS Direct Connect connection between the production site and the nearest AWS Region. Use the Direct Connect connection to upload the images.
- D. Deploy the Lambda function and the ML models onto the AWS IoT Greengrass core that is running on the industrial PCs that are installed on each machine. Extend the long-running Lambda function that runs on AWS IoT Greengrass to invoke the Lambda function with the captured images and run the inference on the edge component that forwards the results directly to the web service.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 148

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A data scientist is using an Amazon SageMaker notebook instance and needs to securely access data stored in a specific Amazon S3 bucket.

How should the data scientist accomplish this?

- A. Add an S3 bucket policy allowing GetObject, PutObject, and ListBucket permissions to the Amazon SageMaker notebook ARN as principal.
- B. Encrypt the objects in the S3 bucket with a custom AWS Key Management Service (AWS KMS) key that only the notebook owner has access to.
- C. Attach the policy to the IAM role associated with the notebook that allows GetObject, PutObject, and ListBucket operations to the specific S3 bucket.
- D. Use a script in a lifecycle configuration to configure the AWS CLI on the instance with an access key ID and secret.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 149

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company is launching a new product and needs to build a mechanism to monitor comments about the company and its new product on social media. The company needs to be able to evaluate the sentiment expressed in social media posts, and visualize trends and configure alarms based on various thresholds.

The company needs to implement this solution quickly, and wants to minimize the infrastructure and data science resources needed to evaluate the messages.

The company already has a solution in place to collect posts and store them within an Amazon S3 bucket.

What services should the data science team use to deliver this solution?

- A. Train a model in Amazon SageMaker by using the BlazingText algorithm to detect sentiment in the corpus of social media posts. Expose an endpoint that can be called by AWS Lambda. Trigger a Lambda function when posts are added to the S3 bucket to invoke the endpoint and record the sentiment in an Amazon DynamoDB table and in a custom Amazon CloudWatch metric. Use CloudWatch alarms to notify analysts of trends.
- B. Train a model in Amazon SageMaker by using the semantic segmentation algorithm to model the semantic content in the corpus of social media posts. Expose an endpoint that can be called by AWS Lambda. Trigger a Lambda function when objects are added to the S3 bucket to invoke the endpoint and record the sentiment in an Amazon DynamoDB table. Schedule a second Lambda function to query recently added records and send an Amazon Simple Notification Service (Amazon SNS) notification to notify analysts of trends.
- C. Trigger an AWS Lambda function when social media posts are added to the S3 bucket. Call Amazon Comprehend for each post to capture the sentiment in the message and record the sentiment in an Amazon DynamoDB table. Schedule a second Lambda function to query recently added records and send an Amazon Simple Notification Service (Amazon SNS) notification to notify analysts of trends.
- D. Trigger an AWS Lambda function when social media posts are added to the S3 bucket. Call Amazon Comprehend for each post to capture the sentiment in the message and record the sentiment in a custom Amazon CloudWatch metric and in S3. Use CloudWatch alarms to notify analysts of trends.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 150

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A bank wants to launch a low-rate credit promotion. The bank is located in a town that recently experienced economic hardship. Only some of the bank's customers were affected by the crisis, so the bank's credit team must identify which customers to target with the promotion. However, the credit team wants to make sure that loyal customers' full credit history is considered when the decision is made.

The bank's data science team developed a model that classifies account transactions and understands credit eligibility. The data science team used the XGBoost algorithm to train the model. The team used 7 years of bank transaction historical data for training and hyperparameter tuning over the course of several days.

The accuracy of the model is sufficient, but the credit team is struggling to explain accurately why the model denies credit to some customers. The credit team has almost no skill in data science.

What should the data science team do to address this issue in the MOST operationally efficient manner?

- A. Use Amazon SageMaker Studio to rebuild the model. Create a notebook that uses the XGBoost training container to perform model training. Deploy the model at an endpoint. Enable Amazon SageMaker Model Monitor to store inferences. Use the inferences to create Shapley values that help explain model behavior. Create a chart that shows features and SHapley Additive exPlanations (SHAP) values to explain to the credit team how the features affect the model outcomes.
- B. Use Amazon SageMaker Studio to rebuild the model. Create a notebook that uses the XGBoost training container to perform model training. Activate Amazon SageMaker Debugger, and configure it to calculate and collect Shapley values. Create a chart that shows features and SHapley Additive exPlanations (SHAP) values to explain to the credit team how the features affect the model outcomes.
- C. Create an Amazon SageMaker notebook instance. Use the notebook instance and the XGBoost library to locally retrain the model. Use the `plot_importance()` method in the Python XGBoost interface to create a feature importance chart. Use that chart to explain to the credit team how the features affect the model outcomes.
- D. Use Amazon SageMaker Studio to rebuild the model. Create a notebook that uses the XGBoost training container to perform model training. Deploy the model at an endpoint. Use Amazon SageMaker Processing to post-analyze the model and create a feature importance explainability chart automatically for the credit team.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 151

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A data science team is planning to build a natural language processing (NLP) application. The application's text preprocessing stage will include part-of-speech tagging and key phrase extraction. The preprocessed text will be input to a custom classification algorithm that the data science team has already written and trained using Apache MXNet.

Which solution can the team build MOST quickly to meet these requirements?

- A. Use Amazon Comprehend for the part-of-speech tagging, key phrase extraction, and classification tasks.
- B. Use an NLP library in Amazon SageMaker for the part-of-speech tagging. Use Amazon Comprehend for the key phrase extraction. Use AWS Deep Learning Containers with Amazon SageMaker to build the custom classifier.
- C. Use Amazon Comprehend for the part-of-speech tagging and key phrase extraction tasks. Use Amazon SageMaker built-in Latent Dirichlet Allocation (LDA) algorithm to build the custom classifier.
- D. Use Amazon Comprehend for the part-of-speech tagging and key phrase extraction tasks. Use AWS Deep Learning Containers with Amazon SageMaker to build the custom classifier.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 152

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A machine learning (ML) specialist must develop a classification model for a financial services company. A domain expert provides the dataset, which is tabular with 10,000 rows and 1,020 features. During exploratory data analysis, the specialist finds no missing values and a small percentage of duplicate rows. There are correlation scores of > 0.9 for 200 feature pairs. The mean value of each feature is similar to its 50th percentile.

Which feature engineering strategy should the ML specialist use with Amazon SageMaker?

- A. Apply dimensionality reduction by using the principal component analysis (PCA) algorithm.
- B. Drop the features with low correlation scores by using a Jupyter notebook.
- C. Apply anomaly detection by using the Random Cut Forest (RCF) algorithm.
- D. Concatenate the features with high correlation scores by using a Jupyter notebook.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 153

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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 100,000 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

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 154

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A machine learning specialist needs to analyze comments on a news website with users across the globe. The specialist must find the most discussed topics in the comments that are in either English or Spanish.

What steps could be used to accomplish this task? (Choose two.)

- A. Use an Amazon SageMaker BlazingText algorithm to find the topics independently from language. Proceed with the analysis.
- B. Use an Amazon SageMaker seq2seq algorithm to translate from Spanish to English, if necessary. Use a SageMaker Latent Dirichlet Allocation (LDA) algorithm to find the topics.
- C. Use Amazon Translate to translate from Spanish to English, if necessary. Use Amazon Comprehend topic modeling to find the topics.
- D. Use Amazon Translate to translate from Spanish to English, if necessary. Use Amazon Lex to extract topics from the content.
- E. Use Amazon Translate to translate from Spanish to English, if necessary. Use Amazon SageMaker Neural Topic Model (NTM) to find the topics.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 155

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A machine learning (ML) specialist is administering a production Amazon SageMaker endpoint with model monitoring configured. Amazon SageMaker Model Monitor detects violations on the SageMaker endpoint, so the ML specialist retrains the model with the latest dataset. This dataset is statistically representative of the current production traffic. The ML specialist notices that even after deploying the new SageMaker model and running the first monitoring job, the SageMaker endpoint still has violations.

What should the ML specialist do to resolve the violations?

- A. Manually trigger the monitoring job to re-evaluate the SageMaker endpoint traffic sample.
- B. Run the Model Monitor baseline job again on the new training set. Configure Model Monitor to use the new baseline.
- C. Delete the endpoint and recreate it with the original configuration.
- D. Retrain the model again by using a combination of the original training set and the new training set.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 156

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company supplies wholesale clothing to thousands of retail stores. A data scientist must create a model that predicts the daily sales volume for each item for each store. The data scientist discovers that more than half of the stores have been in business for less than 6 months. Sales data is highly consistent from week to week. Daily data from the database has been aggregated weekly, and weeks with no sales are omitted from the current dataset. Five years (100 MB) of sales data is available in Amazon S3.

Which factors will adversely impact the performance of the forecast model to be developed, and which actions should the data scientist take to mitigate them?
(Choose two.)

- A. Detecting seasonality for the majority of stores will be an issue. Request categorical data to relate new stores with similar stores that have more historical data.
- B. The sales data does not have enough variance. Request external sales data from other industries to improve the model's ability to generalize.
- C. Sales data is aggregated by week. Request daily sales data from the source database to enable building a daily model.
- D. The sales data is missing zero entries for item sales. Request that item sales data from the source database include zero entries to enable building the model.
- E. Only 100 MB of sales data is available in Amazon S3. Request 10 years of sales data, which would provide 200 MB of training data for the model.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 157

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

An ecommerce company is automating the categorization of its products based on images. A data scientist has trained a computer vision model using the Amazon SageMaker image classification algorithm. The images for each product are classified according to specific product lines. The accuracy of the model is too low when categorizing new products. All of the product images have the same dimensions and are stored within an Amazon S3 bucket. The company wants to improve the model so it can be used for new products as soon as possible.

Which steps would improve the accuracy of the solution? (Choose three.)

- A. Use the SageMaker semantic segmentation algorithm to train a new model to achieve improved accuracy.
- B. Use the Amazon Rekognition DetectLabels API to classify the products in the dataset.
- C. Augment the images in the dataset. Use open source libraries to crop, resize, flip, rotate, and adjust the brightness and contrast of the images.
- D. Use a SageMaker notebook to implement the normalization of pixels and scaling of the images. Store the new dataset in Amazon S3.
- E. Use Amazon Rekognition Custom Labels to train a new model.
- F. Check whether there are class imbalances in the product categories, and apply oversampling or undersampling as required. Store the new dataset in Amazon S3.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 158

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A data scientist is training a text classification model by using the Amazon SageMaker built-in BlazingText algorithm. There are 5 classes in the dataset, with 300 samples for category A, 292 samples for category B, 240 samples for category C, 258 samples for category D, and 310 samples for category E.

The data scientist shuffles the data and splits off 10% for testing. After training the model, the data scientist generates confusion matrices for the training and test sets.

Training data confusion matrix

		Predicted class					Total
		A	B	C	D	E	
True class	A	270	0	0	0	0	270
	B	1	260	0	0	2	263
	C	0	0	111	100	5	216
	D	4	3	132	92	1	232
	E	0	0	2	3	274	279
	Total	275	263	245	195	282	1260

Test data confusion matrix

		Predicted class					Total
		A	B	C	D	E	
True class	A	9	1	0	0	0	10
	B	2	25	0	2	0	29
	C	10	2	11	10	1	34
	D	1	0	12	14	0	27
	E	9	1	4	1	25	40
	Total	31	29	27	27	26	140

What could the data scientist conclude from these results?

- A. Classes C and D are too similar.
- B. The dataset is too small for holdout cross-validation.
- C. The data distribution is skewed.
- D. The model is overfitting for classes B and E.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 159

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company that manufactures mobile devices wants to determine and calibrate the appropriate sales price for its devices. The company is collecting the relevant data and is determining data features that it can use to train machine learning (ML) models. There are more than 1,000 features, and the company wants to determine the primary features that contribute to the sales price.

Which techniques should the company use for feature selection? (Choose three.)

- A. Data scaling with standardization and normalization
- B. Correlation plot with heat maps
- C. Data binning
- D. Univariate selection
- E. Feature importance with a tree-based classifier
- F. Data augmentation

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 160

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A power company wants to forecast future energy consumption for its customers in residential properties and commercial business properties. Historical power consumption data for the last 10 years is available. A team of data scientists who performed the initial data analysis and feature selection will include the historical power consumption data and data such as weather, number of individuals on the property, and public holidays.

The data scientists are using Amazon Forecast to generate the forecasts.

Which algorithm in Forecast should the data scientists use to meet these requirements?

- A. Autoregressive Integrated Moving Average (AIRMA)
- B. Exponential Smoothing (ETS)
- C. Convolutional Neural Network - Quantile Regression (CNN-QR)
- D. Prophet

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 161

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company wants to use automatic speech recognition (ASR) to transcribe messages that are less than 60 seconds long from a voicemail-style application. The company requires the correct identification of 200 unique product names, some of which have unique spellings or pronunciations.

The company has 4,000 words of Amazon SageMaker Ground Truth voicemail transcripts it can use to customize the chosen ASR model. The company needs to ensure that everyone can update their customizations multiple times each hour.

Which approach will maximize transcription accuracy during the development phase?

- A. Use a voice-driven Amazon Lex bot to perform the ASR customization. Create customer slots within the bot that specifically identify each of the required product names. Use the Amazon Lex synonym mechanism to provide additional variations of each product name as mis-transcriptions are identified in development.
- B. Use Amazon Transcribe to perform the ASR customization. Analyze the word confidence scores in the transcript, and automatically create or update a custom vocabulary file with any word that has a confidence score below an acceptable threshold value. Use this updated custom vocabulary file in all future transcription tasks.
- C. Create a custom vocabulary file containing each product name with phonetic pronunciations, and use it with Amazon Transcribe to perform the ASR customization. Analyze the transcripts and manually update the custom vocabulary file to include updated or additional entries for those names that are not being correctly identified.
- D. Use the audio transcripts to create a training dataset and build an Amazon Transcribe custom language model. Analyze the transcripts and update the training dataset with a manually corrected version of transcripts where product names are not being transcribed correctly. Create an updated custom language model.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 162

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 163

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A machine learning specialist is developing a regression model to predict rental rates from rental listings. A variable named Wall_Color represents the most prominent exterior wall color of the property. The following is the sample data, excluding all other variables:

Property_ID	Wall_Color
1000	Red
1001	White
1002	Green

The specialist chose a model that needs numerical input data.

Which feature engineering approaches should the specialist use to allow the regression model to learn from the Wall_Color data? (Choose two.)

- A. Apply integer transformation and set Red = 1, White = 5, and Green = 10.
- B. Add new columns that store one-hot representation of colors.
- C. Replace the color name string by its length.
- D. Create three columns to encode the color in RGB format.
- E. Replace each color name by its training set frequency.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 164

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A data scientist is working on a public sector project for an urban traffic system. While studying the traffic patterns, it is clear to the data scientist that the traffic behavior at each light is correlated, subject to a small stochastic error term. The data scientist must model the traffic behavior to analyze the traffic patterns and reduce congestion.

How will the data scientist MOST effectively model the problem?

- A. The data scientist should obtain a correlated equilibrium policy by formulating this problem as a multi-agent reinforcement learning problem.
- B. The data scientist should obtain the optimal equilibrium policy by formulating this problem as a single-agent reinforcement learning problem.
- C. Rather than finding an equilibrium policy, the data scientist should obtain accurate predictors of traffic flow by using historical data through a supervised learning approach.
- D. Rather than finding an equilibrium policy, the data scientist should obtain accurate predictors of traffic flow by using unlabeled simulated data representing the new traffic patterns in the city and applying an unsupervised learning approach.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 165

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A data scientist is using the Amazon SageMaker Neural Topic Model (NTM) algorithm to build a model that recommends tags from blog posts. The raw blog post data is stored in an Amazon S3 bucket in JSON format. During model evaluation, the data scientist discovered that the model recommends certain stopwords such as "a," "an," and "the" as tags to certain blog posts, along with a few rare words that are present only in certain blog entries. After a few iterations of tag review with the content team, the data scientist notices that the rare words are unusual but feasible. The data scientist also must ensure that the tag recommendations of the generated model do not include the stopwords.

What should the data scientist do to meet these requirements?

- A. Use the Amazon Comprehend entity recognition API operations. Remove the detected words from the blog post data. Replace the blog post data source in the S3 bucket.
- B. Run the SageMaker built-in principal component analysis (PCA) algorithm with the blog post data from the S3 bucket as the data source. Replace the blog post data in the S3 bucket with the results of the training job.
- C. Use the SageMaker built-in Object Detection algorithm instead of the NTM algorithm for the training job to process the blog post data.
- D. Remove the stopwords from the blog post data by using the CountVectorizer function in the scikit-learn library. Replace the blog post data in the S3 bucket with the results of the vectorizer.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 166

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company wants to create a data repository in the AWS Cloud for machine learning (ML) projects. The company wants to use AWS to perform complete ML lifecycles and wants to use Amazon S3 for the data storage. All of the company's data currently resides on premises and is 40 TB in size.

The company wants a solution that can transfer and automatically update data between the on-premises object storage and Amazon S3. The solution must support encryption, scheduling, monitoring, and data integrity validation.

Which solution meets these requirements?

- A. Use the S3 sync command to compare the source S3 bucket and the destination S3 bucket. Determine which source files do not exist in the destination S3 bucket and which source files were modified.
- B. Use AWS Transfer for FTPS to transfer the files from the on-premises storage to Amazon S3.
- C. Use AWS DataSync to make an initial copy of the entire dataset. Schedule subsequent incremental transfers of changing data until the final cutover from on premises to AWS.
- D. Use S3 Batch Operations to pull data periodically from the on-premises storage. Enable S3 Versioning on the S3 bucket to protect against accidental overwrites.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 167

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 168

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 169

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company is building a new version of a recommendation engine. Machine learning (ML) specialists need to keep adding new data from users to improve personalized recommendations. The ML specialists gather data from the users' interactions on the platform and from sources such as external websites and social media. The pipeline cleans, transforms, enriches, and compresses terabytes of data daily, and this data is stored in Amazon S3. A set of Python scripts was coded to do the job and is stored in a large Amazon EC2 instance. The whole process takes more than 20 hours to finish, with each script taking at least an hour. The company wants to move the scripts out of Amazon EC2 into a more managed solution that will eliminate the need to maintain servers.

Which approach will address all of these requirements with the LEAST development effort?

- A. Load the data into an Amazon Redshift cluster. Execute the pipeline by using SQL. Store the results in Amazon S3.
- B. Load the data into Amazon DynamoDB. Convert the scripts to an AWS Lambda function. Execute the pipeline by triggering Lambda executions. Store the results in Amazon S3.
- C. Create an AWS Glue job. Convert the scripts to PySpark. Execute the pipeline. Store the results in Amazon S3.
- D. Create a set of individual AWS Lambda functions to execute each of the scripts. Build a step function by using the AWS Step Functions Data Science SDK. Store the results in Amazon S3.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 170

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 171

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 172

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A data scientist has a dataset of machine part images stored in Amazon Elastic File System (Amazon EFS). The data scientist needs to use Amazon SageMaker to create and train an image classification machine learning model based on this dataset. Because of budget and time constraints, management wants the data scientist to create and train a model with the least number of steps and integration work required.

How should the data scientist meet these requirements?

- A. Mount the EFS file system to a SageMaker notebook and run a script that copies the data to an Amazon FSx for Lustre file system. Run the SageMaker training job with the FSx for Lustre file system as the data source.
- B. Launch a transient Amazon EMR cluster. Configure steps to mount the EFS file system and copy the data to an Amazon S3 bucket by using S3DistCp. Run the SageMaker training job with Amazon S3 as the data source.
- C. Mount the EFS file system to an Amazon EC2 instance and use the AWS CLI to copy the data to an Amazon S3 bucket. Run the SageMaker training job with Amazon S3 as the data source.
- D. Run a SageMaker training job with an EFS file system as the data source.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 173

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A retail company uses a machine learning (ML) model for daily sales forecasting. The company's brand manager reports that the model has provided inaccurate results for the past 3 weeks.

At the end of each day, an AWS Glue job consolidates the input data that is used for the forecasting with the actual daily sales data and the predictions of the model. The AWS Glue job stores the data in Amazon S3. The company's ML team is using an Amazon SageMaker Studio notebook to gain an understanding about the source of the model's inaccuracies.

What should the ML team do on the SageMaker Studio notebook to visualize the model's degradation MOST accurately?

- A. Create a histogram of the daily sales over the last 3 weeks. In addition, create a histogram of the daily sales from before that period.
- B. Create a histogram of the model errors over the last 3 weeks. In addition, create a histogram of the model errors from before that period.
- C. Create a line chart with the weekly mean absolute error (MAE) of the model.
- D. Create a scatter plot of daily sales versus model error for the last 3 weeks. In addition, create a scatter plot of daily sales versus model error from before that period.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 174

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

An ecommerce company sends a weekly email newsletter to all of its customers. Management has hired a team of writers to create additional targeted content. A data scientist needs to identify five customer segments based on age, income, and location. The customers' current segmentation is unknown. The data scientist previously built an XGBoost model to predict the likelihood of a customer responding to an email based on age, income, and location.

Why does the XGBoost model NOT meet the current requirements, and how can this be fixed?

- A. The XGBoost model provides a true/false binary output. Apply principal component analysis (PCA) with five feature dimensions to predict a segment.
- B. The XGBoost model provides a true/false binary output. Increase the number of classes the XGBoost model predicts to five classes to predict a segment.
- C. The XGBoost model is a supervised machine learning algorithm. Train a k-Nearest-Neighbors (kNN) model with $K = 5$ on the same dataset to predict a segment.
- D. The XGBoost model is a supervised machine learning algorithm. Train a k-means model with $K = 5$ on the same dataset to predict a segment.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 175

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A global financial company is using machine learning to automate its loan approval process. The company has a dataset of customer information. The dataset contains some categorical fields, such as customer location by city and housing status. The dataset also includes financial fields in different units, such as account balances in US dollars and monthly interest in US cents.

The company's data scientists are using a gradient boosting regression model to infer the credit score for each customer. The model has a training accuracy of 99% and a testing accuracy of 75%. The data scientists want to improve the model's testing accuracy.

Which process will improve the testing accuracy the MOST?

- A. Use a one-hot encoder for the categorical fields in the dataset. Perform standardization on the financial fields in the dataset. Apply L1 regularization to the data.
- B. Use tokenization of the categorical fields in the dataset. Perform binning on the financial fields in the dataset. Remove the outliers in the data by using the z- score.
- C. Use a label encoder for the categorical fields in the dataset. Perform L1 regularization on the financial fields in the dataset. Apply L2 regularization to the data.
- D. Use a logarithm transformation on the categorical fields in the dataset. Perform binning on the financial fields in the dataset. Use imputation to populate missing values in the dataset.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 176

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A machine learning (ML) specialist needs to extract embedding vectors from a text series. The goal is to provide a ready-to-ingest feature space for a data scientist to develop downstream ML predictive models. The text consists of curated sentences in English. Many sentences use similar words but in different contexts. There are questions and answers among the sentences, and the embedding space must differentiate between them.

Which options can produce the required embedding vectors that capture word context and sequential QA information? (Choose two.)

- A. Amazon SageMaker seq2seq algorithm
- B. Amazon SageMaker BlazingText algorithm in Skip-gram mode
- C. Amazon SageMaker Object2Vec algorithm
- D. Amazon SageMaker BlazingText algorithm in continuous bag-of-words (CBOW) mode
- E. Combination of the Amazon SageMaker BlazingText algorithm in Batch Skip-gram mode with a custom recurrent neural network (RNN)

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 177

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A retail company wants to update its customer support system. The company wants to implement automatic routing of customer claims to different queues to prioritize the claims by category.

Currently, an operator manually performs the category assignment and routing. After the operator classifies and routes the claim, the company stores the claim's record in a central database. The claim's record includes the claim's category.

The company has no data science team or experience in the field of machine learning (ML). The company's small development team needs a solution that requires no ML expertise.

Which solution meets these requirements?

- A. Export the database to a .csv file with two columns: claim_label and claim_text. Use the Amazon SageMaker Object2Vec algorithm and the .csv file to train a model. Use SageMaker to deploy the model to an inference endpoint. Develop a service in the application to use the inference endpoint to process incoming claims, predict the labels, and route the claims to the appropriate queue.
- B. Export the database to a .csv file with one column: claim_text. Use the Amazon SageMaker Latent Dirichlet Allocation (LDA) algorithm and the .csv file to train a model. Use the LDA algorithm to detect labels automatically. Use SageMaker to deploy the model to an inference endpoint. Develop a service in the application to use the inference endpoint to process incoming claims, predict the labels, and route the claims to the appropriate queue.
- C. Use Amazon Textract to process the database and automatically detect two columns: claim_label and claim_text. Use Amazon Comprehend custom classification and the extracted information to train the custom classifier. Develop a service in the application to use the Amazon Comprehend API to process incoming claims, predict the labels, and route the claims to the appropriate queue.
- D. Export the database to a .csv file with two columns: claim_label and claim_text. Use Amazon Comprehend custom classification and the .csv file to train the custom classifier. Develop a service in the application to use the Amazon Comprehend API to process incoming claims, predict the labels, and route the claims to the appropriate queue.

Show Suggested Answer

Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 178

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A machine learning (ML) specialist is using Amazon SageMaker hyperparameter optimization (HPO) to improve a model's accuracy. The learning rate parameter is specified in the following HPO configuration:

```
{
  "Name": "learning_rate",
  "MaxValue" : "0.0001",
  "MinValue": "0.1"
}
```

During the results analysis, the ML specialist determines that most of the training jobs had a learning rate between 0.01 and 0.1. The best result had a learning rate of less than 0.01. Training jobs need to run regularly over a changing dataset. The ML specialist needs to find a tuning mechanism that uses different learning rates more evenly from the provided range between MinValue and MaxValue.

Which solution provides the MOST accurate result?

A. Modify the HPO configuration as follows:

```
{
  "Name": "learning_rate",
  "MaxValue" : "0.0001",
  "MinValue": "0.1"
  "ScalingType": "ReverseLogarithmic"
}
```

Select the most accurate hyperparameter

configuration from this HPO job.

B. Run three different HPO jobs that use different learning rates from the following intervals for MinValue and MaxValue while using the same number of training jobs for each HPO job: $\Rightarrow [0.01, 0.1] \Rightarrow [0.001, 0.01] \Rightarrow [0.0001, 0.001]$ Select the most accurate hyperparameter configuration from these three HPO jobs.

C. Modify the HPO configuration as follows:

```
{
  "Name": "learning_rate",
  "MaxValue" : "0.0001",
  "MinValue": "0.1"
  "ScalingType": "Logarithmic"
}
```

Select the most accurate hyperparameter configuration

from this training job.

D. Run three different HPO jobs that use different learning rates from the following intervals for MinValue and MaxValue. Divide the number of training jobs for each HPO job by three: $\Rightarrow [0.01, 0.1] \Rightarrow [0.001, 0.01] [0.0001, 0.001]$ • Select the most accurate hyperparameter configuration from these three HPO jobs.

Show Suggested Answer

Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 179

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A manufacturing company wants to use machine learning (ML) to automate quality control in its facilities. The facilities are in remote locations and have limited internet connectivity. The company has 20 TB of training data that consists of labeled images of defective product parts. The training data is in the corporate on-premises data center.

The company will use this data to train a model for real-time defect detection in new parts as the parts move on a conveyor belt in the facilities. The company needs a solution that minimizes costs for compute infrastructure and that maximizes the scalability of resources for training. The solution also must facilitate the company's use of an ML model in the low-connectivity environments.

Which solution will meet these requirements?

- A. Move the training data to an Amazon S3 bucket. Train and evaluate the model by using Amazon SageMaker. Optimize the model by using SageMaker Neo. Deploy the model on a SageMaker hosting services endpoint.
- B. Train and evaluate the model on premises. Upload the model to an Amazon S3 bucket. Deploy the model on an Amazon SageMaker hosting services endpoint.
- C. Move the training data to an Amazon S3 bucket. Train and evaluate the model by using Amazon SageMaker. Optimize the model by using SageMaker Neo. Set up an edge device in the manufacturing facilities with AWS IoT Greengrass. Deploy the model on the edge device.
- D. Train the model on premises. Upload the model to an Amazon S3 bucket. Set up an edge device in the manufacturing facilities with AWS IoT Greengrass. Deploy the model on the edge device.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 180

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 181

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A real-estate company is launching a new product that predicts the prices of new houses. The historical data for the properties and prices is stored in .csv format in an Amazon S3 bucket. The data has a header, some categorical fields, and some missing values. The company's data scientists have used Python with a common open-source library to fill the missing values with zeros. The data scientists have dropped all of the categorical fields and have trained a model by using the open-source linear regression algorithm with the default parameters.

The accuracy of the predictions with the current model is below 50%. The company wants to improve the model performance and launch the new product as soon as possible.

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

- A. Create a service-linked role for Amazon Elastic Container Service (Amazon ECS) with access to the S3 bucket. Create an ECS cluster that is based on an AWS Deep Learning Containers image. Write the code to perform the feature engineering. Train a logistic regression model for predicting the price, pointing to the bucket with the dataset. Wait for the training job to complete. Perform the inferences.
- B. Create an Amazon SageMaker notebook with a new IAM role that is associated with the notebook. Pull the dataset from the S3 bucket. Explore different combinations of feature engineering transformations, regression algorithms, and hyperparameters. Compare all the results in the notebook, and deploy the most accurate configuration in an endpoint for predictions.
- C. Create an IAM role with access to Amazon S3, Amazon SageMaker, and AWS Lambda. Create a training job with the SageMaker built-in XGBoost model pointing to the bucket with the dataset. Specify the price as the target feature. Wait for the job to complete. Load the model artifact to a Lambda function for inference on prices of new houses.
- D. Create an IAM role for Amazon SageMaker with access to the S3 bucket. Create a SageMaker AutoML job with SageMaker Autopilot pointing to the bucket with the dataset. Specify the price as the target attribute. Wait for the job to complete. Deploy the best model for predictions.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 182

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A data scientist is reviewing customer comments about a company's products. The data scientist needs to present an initial exploratory analysis by using charts and a word cloud. The data scientist must use feature engineering techniques to prepare this analysis before starting a natural language processing (NLP) model.

Which combination of feature engineering techniques should the data scientist use to meet these requirements? (Choose two.)

- A. Named entity recognition
- B. Coreference
- C. Stemming
- D. Term frequency-inverse document frequency (TF-IDF)
- E. Sentiment analysis

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 183

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A data scientist is evaluating a GluonTS on Amazon SageMaker DeepAR model. The evaluation metrics on the test set indicate that the coverage score is 0.489 and 0.889 at the 0.5 and 0.9 quantiles, respectively.

What can the data scientist reasonably conclude about the distributional forecast related to the test set?

- A. The coverage scores indicate that the distributional forecast is poorly calibrated. These scores should be approximately equal to each other at all quantiles.
- B. The coverage scores indicate that the distributional forecast is poorly calibrated. These scores should peak at the median and be lower at the tails.
- C. The coverage scores indicate that the distributional forecast is correctly calibrated. These scores should always fall below the quantile itself.
- D. The coverage scores indicate that the distributional forecast is correctly calibrated. These scores should be approximately equal to the quantile itself.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 184

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

An energy company has wind turbines, weather stations, and solar panels that generate telemetry data. The company wants to perform predictive maintenance on these devices. The devices are in various locations and have unstable internet connectivity.

A team of data scientists is using the telemetry data to perform machine learning (ML) to conduct anomaly detection and predict maintenance before the devices start to deteriorate. The team needs a scalable, secure, high-velocity data ingestion mechanism. The team has decided to use Amazon S3 as the data storage location.

Which approach meets these requirements?

- A. Ingest the data by using an HTTP API call to a web server that is hosted on Amazon EC2. Set up EC2 instances in an Auto Scaling configuration behind an Elastic Load Balancer to load the data into Amazon S3.
- B. Ingest the data over Message Queuing Telemetry Transport (MQTT) to AWS IoT Core. Set up a rule in AWS IoT Core to use Amazon Kinesis Data Firehose to send data to an Amazon Kinesis data stream that is configured to write to an S3 bucket.
- C. Ingest the data over Message Queuing Telemetry Transport (MQTT) to AWS IoT Core. Set up a rule in AWS IoT Core to direct all MQTT data to an Amazon Kinesis Data Firehose delivery stream that is configured to write to an S3 bucket.
- D. Ingest the data over Message Queuing Telemetry Transport (MQTT) to Amazon Kinesis data stream that is configured to write to an S3 bucket.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 185

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A retail company collects customer comments about its products from social media, the company website, and customer call logs. A team of data scientists and engineers wants to find common topics and determine which products the customers are referring to in their comments. The team is using natural language processing (NLP) to build a model to help with this classification.

Each product can be classified into multiple categories that the company defines. These categories are related but are not mutually exclusive. For example, if there is mention of "Sample Yogurt" in the document of customer comments, then "Sample Yogurt" should be classified as "yogurt," "snack," and "dairy product."

The team is using Amazon Comprehend to train the model and must complete the project as soon as possible.

Which functionality of Amazon Comprehend should the team use to meet these requirements?

- A. Custom classification with multi-class mode
- B. Custom classification with multi-label mode
- C. Custom entity recognition
- D. Built-in models

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 186

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A data engineer is using AWS Glue to create optimized, secure datasets in Amazon S3. The data science team wants the ability to access the ETL scripts directly from Amazon SageMaker notebooks within a VPC. After this setup is complete, the data science team wants the ability to run the AWS Glue job and invoke the SageMaker training job.

Which combination of steps should the data engineer take to meet these requirements? (Choose three.)

- A. Create a SageMaker development endpoint in the data science team's VPC.
- B. Create an AWS Glue development endpoint in the data science team's VPC.
- C. Create SageMaker notebooks by using the AWS Glue development endpoint.
- D. Create SageMaker notebooks by using the SageMaker console.
- E. Attach a decryption policy to the SageMaker notebooks.
- F. Create an IAM policy and an IAM role for the SageMaker notebooks.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 187

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A data engineer needs to provide a team of data scientists with the appropriate dataset to run machine learning training jobs. The data will be stored in Amazon S3. The data engineer is obtaining the data from an Amazon Redshift database and is using join queries to extract a single tabular dataset. A portion of the schema is as follows:

TransactionTimestamp (Timestamp)

CardName (Varchar)

CardNo (Varchar)

The data engineer must provide the data so that any row with a CardNo value of NULL is removed. Also, the TransactionTimestamp column must be separated into a TransactionDate column and a TransactionTime column. Finally, the CardName column must be renamed to NameOnCard.

The data will be extracted on a monthly basis and will be loaded into an S3 bucket. The solution must minimize the effort that is needed to set up infrastructure for the ingestion and transformation. The solution also must be automated and must minimize the load on the Amazon Redshift cluster.

Which solution meets these requirements?

- A. Set up an Amazon EMR cluster. Create an Apache Spark job to read the data from the Amazon Redshift cluster and transform the data. Load the data into the S3 bucket. Schedule the job to run monthly.
- B. Set up an Amazon EC2 instance with a SQL client tool, such as SQL Workbench/J, to query the data from the Amazon Redshift cluster directly. Export the resulting dataset into a file. Upload the file into the S3 bucket. Perform these tasks monthly.
- C. Set up an AWS Glue job that has the Amazon Redshift cluster as the source and the S3 bucket as the destination. Use the built-in transforms Filter, Map, and RenameField to perform the required transformations. Schedule the job to run monthly.
- D. Use Amazon Redshift Spectrum to run a query that writes the data directly to the S3 bucket. Create an AWS Lambda function to run the query monthly.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 188

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A machine learning (ML) specialist wants to bring a custom training algorithm to Amazon SageMaker. The ML specialist implements the algorithm in a Docker container that is supported by SageMaker.

How should the ML specialist package the Docker container so that SageMaker can launch the training correctly?

- A. Specify the server argument in the ENTRYPOINT instruction in the Dockerfile.
- B. Specify the training program in the ENTRYPOINT instruction in the Dockerfile.
- C. Include the path to the training data in the docker build command when packaging the container.
- D. Use a COPY instruction in the Dockerfile to copy the training program to the /opt/ml/train directory.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 189

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

An ecommerce company wants to use machine learning (ML) to monitor fraudulent transactions on its website. The company is using Amazon SageMaker to research, train, deploy, and monitor the ML models.

The historical transactions data is in a .csv file that is stored in Amazon S3. The data contains features such as the user's IP address, navigation time, average time on each page, and the number of clicks for each session. There is no label in the data to indicate if a transaction is anomalous.

Which models should the company use in combination to detect anomalous transactions? (Choose two.)

- A. IP Insights
- B. K-nearest neighbors (k-NN)
- C. Linear learner with a logistic function
- D. Random Cut Forest (RCF)
- E. XGBoost

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 190

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A healthcare company is using an Amazon SageMaker notebook instance to develop machine learning (ML) models. The company's data scientists will need to be able to access datasets stored in Amazon S3 to train the models. Due to regulatory requirements, access to the data from instances and services used for training must not be transmitted over the internet.

Which combination of steps should an ML specialist take to provide this access? (Choose two.)

- A. Configure the SageMaker notebook instance to be launched with a VPC attached and internet access disabled.
- B. Create and configure a VPN tunnel between SageMaker and Amazon S3.
- C. Create and configure an S3 VPC endpoint Attach it to the VPC.
- D. Create an S3 bucket policy that allows traffic from the VPC and denies traffic from the internet.
- E. Deploy AWS Transit Gateway Attach the S3 bucket and the SageMaker instance to the gateway.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 191

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 192

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A newspaper publisher has a table of customer data that consists of several numerical and categorical features, such as age and education history, as well as subscription status. The company wants to build a targeted marketing model for predicting the subscription status based on the table data.

Which Amazon SageMaker built-in algorithm should be used to model the targeted marketing?

- A. Random Cut Forest (RCF)
- B. XGBoost
- C. Neural Topic Model (NTM)
- D. DeepAR forecasting

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 193

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company will use Amazon SageMaker to train and host a machine learning model for a marketing campaign. The data must be encrypted at rest. Most of the data is sensitive customer data. The company wants AWS to maintain the root of trust for the encryption keys and wants key usage to be logged.

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

- A. Use AWS Security Token Service (AWS STS) to create temporary tokens to encrypt the storage volumes for all SageMaker instances and to encrypt the model artifacts and data in Amazon S3.
- B. Use customer managed keys in AWS Key Management Service (AWS KMS) to encrypt the storage volumes for all SageMaker instances and to encrypt the model artifacts and data in Amazon S3.
- C. Use encryption keys stored in AWS CloudHSM to encrypt the storage volumes for all SageMaker instances and to encrypt the model artifacts and data in Amazon S3.
- D. Use SageMaker built-in transient keys to encrypt the storage volumes for all SageMaker instances. Enable default encryption for new Amazon Elastic Block Store (Amazon EBS) volumes.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 194

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A data scientist is working on a model to predict a company's required inventory stock levels. All historical data is stored in .csv files in the company's data lake on Amazon S3. The dataset consists of approximately 500 GB of data. The data scientist wants to use SQL to explore the data before training the model. The company wants to minimize costs.

Which option meets these requirements with the LEAST operational overhead?

- A. Create an Amazon EMR cluster. Create external tables in the Apache Hive metastore, referencing the data that is stored in the S3 bucket. Explore the data from the Hive console.
- B. Use AWS Glue to crawl the S3 bucket and create tables in the AWS Glue Data Catalog. Use Amazon Athena to explore the data.
- C. Create an Amazon Redshift cluster. Use the COPY command to ingest the data from Amazon S3. Explore the data from the Amazon Redshift query editor GUI.
- D. Create an Amazon Redshift cluster. Create external tables in an external schema, referencing the S3 bucket that contains the data. Explore the data from the Amazon Redshift query editor GUI.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 195

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A geospatial analysis company processes thousands of new satellite images each day to produce vessel detection data for commercial shipping. The company stores the training data in Amazon S3. The training data incrementally increases in size with new images each day.

The company has configured an Amazon SageMaker training job to use a single ml.p2.xlarge instance with File input mode to train the built-in Object Detection algorithm. The training process was successful last month but is now failing because of a lack of storage. Aside from the addition of training data, nothing has changed in the model training process.

A machine learning (ML) specialist needs to change the training configuration to fix the problem. The solution must optimize performance and must minimize the cost of training.

Which solution will meet these requirements?

- A. Modify the training configuration to use two ml.p2.xlarge instances.
- B. Modify the training configuration to use Pipe input mode.
- C. Modify the training configuration to use a single ml.p3.2xlarge instance.
- D. Modify the training configuration to use Amazon Elastic File System (Amazon EFS) instead of Amazon S3 to store the input training data.

Show Suggested Answer

Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 196

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company is using Amazon SageMaker to build a machine learning (ML) model to predict customer churn based on customer call transcripts. Audio files from customer calls are located in an on-premises VoIP system that has petabytes of recorded calls. The on-premises infrastructure has high-velocity networking and connects to the company's AWS infrastructure through a VPN connection over a 100 Mbps connection.

The company has an algorithm for transcribing customer calls that requires GPUs for inference. The company wants to store these transcriptions in an Amazon S3 bucket in the AWS Cloud for model development.

Which solution should an ML specialist use to deliver the transcriptions to the S3 bucket as quickly as possible?

- A. Order and use an AWS Snowball Edge Compute Optimized device with an NVIDIA Tesla module to run the transcription algorithm. Use AWS DataSync to send the resulting transcriptions to the transcription S3 bucket.
- B. Order and use an AWS Snowcone device with Amazon EC2 Inf1 instances to run the transcription algorithm. Use AWS DataSync to send the resulting transcriptions to the transcription S3 bucket.
- C. Order and use AWS Outposts to run the transcription algorithm on GPU-based Amazon EC2 instances. Store the resulting transcriptions in the transcription S3 bucket.
- D. Use AWS DataSync to ingest the audio files to Amazon S3. Create an AWS Lambda function to run the transcription algorithm on the audio files when they are uploaded to Amazon S3. Configure the function to write the resulting transcriptions to the transcription S3 bucket.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 197

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company has a podcast platform that has thousands of users. The company has implemented an anomaly detection algorithm to detect low podcast engagement based on a 10-minute running window of user events such as listening, pausing, and exiting the podcast. A machine learning (ML) specialist is designing the data ingestion of these events with the knowledge that the event payload needs some small transformations before inference.

How should the ML specialist design the data ingestion to meet these requirements with the LEAST operational overhead?

- A. Ingest event data by using a GraphQLAPI in AWS AppSync. Store the data in an Amazon DynamoDB table. Use DynamoDB Streams to call an AWS Lambda function to transform the most recent 10 minutes of data before inference.
- B. Ingest event data by using Amazon Kinesis Data Streams. Store the data in Amazon S3 by using Amazon Kinesis Data Firehose. Use AWS Glue to transform the most recent 10 minutes of data before inference.
- C. Ingest event data by using Amazon Kinesis Data Streams. Use an Amazon Kinesis Data Analytics for Apache Flink application to transform the most recent 10 minutes of data before inference.
- D. Ingest event data by using Amazon Managed Streaming for Apache Kafka (Amazon MSK). Use an AWS Lambda function to transform the most recent 10 minutes of data before inference.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 198

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company wants to predict the classification of documents that are created from an application. New documents are saved to an Amazon S3 bucket every 3 seconds. The company has developed three versions of a machine learning (ML) model within Amazon SageMaker to classify document text. The company wants to deploy these three versions to predict the classification of each document.

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

- A. Configure an S3 event notification that invokes an AWS Lambda function when new documents are created. Configure the Lambda function to create three SageMaker batch transform jobs, one batch transform job for each model for each document.
- B. Deploy all the models to a single SageMaker endpoint. Treat each model as a production variant. Configure an S3 event notification that invokes an AWS Lambda function when new documents are created. Configure the Lambda function to call each production variant and return the results of each model.
- C. Deploy each model to its own SageMaker endpoint. Configure an S3 event notification that invokes an AWS Lambda function when new documents are created. Configure the Lambda function to call each endpoint and return the results of each model.
- D. Deploy each model to its own SageMaker endpoint. Create three AWS Lambda functions. Configure each Lambda function to call a different endpoint and return the results. Configure three S3 event notifications to invoke the Lambda functions when new documents are created.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 199

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A manufacturing company needs to identify returned smartphones that have been damaged by moisture. The company has an automated process that produces 2,000 diagnostic values for each phone. The database contains more than five million phone evaluations. The evaluation process is consistent, and there are no missing values in the data. A machine learning (ML) specialist has trained an Amazon SageMaker linear learner ML model to classify phones as moisture damaged or not moisture damaged by using all available features. The model's F1 score is 0.6.

Which changes in model training would MOST likely improve the model's F1 score? (Choose two.)

- A. Continue to use the SageMaker linear learner algorithm. Reduce the number of features with the SageMaker principal component analysis (PCA) algorithm.
- B. Continue to use the SageMaker linear learner algorithm. Reduce the number of features with the scikit-learn multi-dimensional scaling (MDS) algorithm.
- C. Continue to use the SageMaker linear learner algorithm. Set the predictor type to regressor.
- D. Use the SageMaker k-means algorithm with k of less than 1,000 to train the model.
- E. Use the SageMaker k-nearest neighbors (k-NN) algorithm. Set a dimension reduction target of less than 1,000 to train the model.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 200

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company is building a machine learning (ML) model to classify images of plants. An ML specialist has trained the model using the Amazon SageMaker built-in Image Classification algorithm. The model is hosted using a SageMaker endpoint on an ml.m5.xlarge instance for real-time inference. When used by researchers in the field, the inference has greater latency than is acceptable. The latency gets worse when multiple researchers perform inference at the same time on their devices. Using Amazon CloudWatch metrics, the ML specialist notices that the ModelLatency metric shows a high value and is responsible for most of the response latency.

The ML specialist needs to fix the performance issue so that researchers can experience less latency when performing inference from their devices.

Which action should the ML specialist take to meet this requirement?

- A. Change the endpoint instance to an ml.t3 burstable instance with the same vCPU number as the ml.m5.xlarge instance has.
- B. Attach an Amazon Elastic Inference ml.eia2.medium accelerator to the endpoint instance.
- C. Enable Amazon SageMaker Autopilot to automatically tune performance of the model.
- D. Change the endpoint instance to use a memory optimized ML instance.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 201

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

An automotive company is using computer vision in its autonomous cars. The company has trained its models successfully by using transfer learning from a convolutional neural network (CNN). The models are trained with PyTorch through the use of the Amazon SageMaker SDK. The company wants to reduce the time that is required for performing inferences, given the low latency that is required for self-driving.

Which solution should the company use to evaluate and improve the performance of the models?

- A. Use Amazon CloudWatch algorithm metrics for visibility into the SageMaker training weights, gradients, biases, and activation outputs. Compute the filter ranks based on this information. Apply pruning to remove the low-ranking filters. Set the new weights. Run a new training job with the pruned model.
- B. Use SageMaker Debugger for visibility into the training weights, gradients, biases, and activation outputs. Adjust the model hyperparameters, and look for lower inference times. Run a new training job.
- C. Use SageMaker Debugger for visibility into the training weights, gradients, biases, and activation outputs. Compute the filter ranks based on this information. Apply pruning to remove the low-ranking filters. Set the new weights. Run a new training job with the pruned model.
- D. Use SageMaker Model Monitor for visibility into the ModelLatency metric and OverheadLatency metric of the model after the model is deployed. Adjust the model hyperparameters, and look for lower inference times. Run a new training job.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 202

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company's machine learning (ML) specialist is designing a scalable data storage solution for Amazon SageMaker. The company has an existing TensorFlow-based model that uses a `train.py` script. The model relies on static training data that is currently stored in TFRecord format.

What should the ML specialist do to provide the training data to SageMaker with the LEAST development overhead?

- A. Put the TFRecord data into an Amazon S3 bucket. Use AWS Glue or AWS Lambda to reformat the data to protobuf format and store the data in a second S3 bucket. Point the SageMaker training invocation to the second S3 bucket.
- B. Rewrite the `train.py` script to add a section that converts TFRecord data to protobuf format. Point the SageMaker training invocation to the local path of the data. Ingest the protobuf data instead of the TFRecord data.
- C. Use SageMaker script mode, and use `train.py` unchanged. Point the SageMaker training invocation to the local path of the data without reformatting the training data.
- D. Use SageMaker script mode, and use `train.py` unchanged. Put the TFRecord data into an Amazon S3 bucket. Point the SageMaker training invocation to the S3 bucket without reformatting the training data.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 203

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

An ecommerce company wants to train a large image classification model with 10,000 classes. The company runs multiple model training iterations and needs to minimize operational overhead and cost. The company also needs to avoid loss of work and model retraining.

Which solution will meet these requirements?

- A. Create the training jobs as AWS Batch jobs that use Amazon EC2 Spot Instances in a managed compute environment.
- B. Use Amazon EC2 Spot Instances to run the training jobs. Use a Spot Instance interruption notice to save a snapshot of the model to Amazon S3 before an instance is terminated.
- C. Use AWS Lambda to run the training jobs. Save model weights to Amazon S3.
- D. Use managed spot training in Amazon SageMaker. Launch the training jobs with checkpointing enabled.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 204

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A retail company uses a machine learning (ML) model for daily sales forecasting. The model has provided inaccurate results for the past 3 weeks. At the end of each day, an AWS Glue job consolidates the input data that is used for the forecasting with the actual daily sales data and the predictions of the model. The AWS Glue job stores the data in Amazon S3.

The company's ML team determines that the inaccuracies are occurring because of a change in the value distributions of the model features. The ML team must implement a solution that will detect when this type of change occurs in the future.

Which solution will meet these requirements with the LEAST amount of operational overhead?

- A. Use Amazon SageMaker Model Monitor to create a data quality baseline. Confirm that the `emit_metrics` option is set to Enabled in the baseline constraints file. Set up an Amazon CloudWatch alarm for the metric.
- B. Use Amazon SageMaker Model Monitor to create a model quality baseline. Confirm that the `emit_metrics` option is set to Enabled in the baseline constraints file. Set up an Amazon CloudWatch alarm for the metric.
- C. Use Amazon SageMaker Debugger to create rules to capture feature values. Set up an Amazon CloudWatch alarm for the rules.
- D. Use Amazon CloudWatch to monitor Amazon SageMaker endpoints. Analyze logs in Amazon CloudWatch Logs to check for data drift.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 205

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A machine learning (ML) specialist has prepared and used a custom container image with Amazon SageMaker to train an image classification model. The ML specialist is performing hyperparameter optimization (HPO) with this custom container image to produce a higher quality image classifier.

The ML specialist needs to determine whether HPO with the SageMaker built-in image classification algorithm will produce a better model than the model produced by HPO with the custom container image. All ML experiments and HPO jobs must be invoked from scripts inside SageMaker Studio notebooks.

How can the ML specialist meet these requirements in the LEAST amount of time?

- A. Prepare a custom HPO script that runs multiple training jobs in SageMaker Studio in local mode to tune the model of the custom container image. Use the automatic model tuning capability of SageMaker with early stopping enabled to tune the model of the built-in image classification algorithm. Select the model with the best objective metric value.
- B. Use SageMaker Autopilot to tune the model of the custom container image. Use the automatic model tuning capability of SageMaker with early stopping enabled to tune the model of the built-in image classification algorithm. Compare the objective metric values of the resulting models of the SageMaker AutopilotAutoML job and the automatic model tuning job. Select the model with the best objective metric value.
- C. Use SageMaker Experiments to run and manage multiple training jobs and tune the model of the custom container image. Use the automatic model tuning capability of SageMaker to tune the model of the built-in image classification algorithm. Select the model with the best objective metric value.
- D. Use the automatic model tuning capability of SageMaker to tune the models of the custom container image and the built-in image classification algorithm at the same time. Select the model with the best objective metric value.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 206

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company wants to deliver digital car management services to its customers. The company plans to analyze data to predict the likelihood of users changing cars. The company has 10 TB of data that is stored in an Amazon Redshift cluster. The company's data engineering team is using Amazon SageMaker Studio for data analysis and model development. Only a subset of the data is relevant for developing the machine learning models. The data engineering team needs a secure and cost-effective way to export the data to a data repository in Amazon S3 for model development.

Which solutions will meet these requirements? (Choose two.)

- A. Launch multiple medium-sized instances in a distributed SageMaker Processing job. Use the prebuilt Docker images for Apache Spark to query and plot the relevant data and to export the relevant data from Amazon Redshift to Amazon S3.
- B. Launch multiple medium-sized notebook instances with a PySpark kernel in distributed mode. Download the data from Amazon Redshift to the notebook cluster. Query and plot the relevant data. Export the relevant data from the notebook cluster to Amazon S3.
- C. Use AWS Secrets Manager to store the Amazon Redshift credentials. From a SageMaker Studio notebook, use the stored credentials to connect to Amazon Redshift with a Python adapter. Use the Python client to query the relevant data and to export the relevant data from Amazon Redshift to Amazon S3.
- D. Use AWS Secrets Manager to store the Amazon Redshift credentials. Launch a SageMaker extra-large notebook instance with block storage that is slightly larger than 10 TB. Use the stored credentials to connect to Amazon Redshift with a Python adapter. Download, query, and plot the relevant data. Export the relevant data from the local notebook drive to Amazon S3.
- E. Use SageMaker Data Wrangler to query and plot the relevant data and to export the relevant data from Amazon Redshift to Amazon S3.

Show Suggested Answer

Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 207

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company is building an application that can predict spam email messages based on email text. The company can generate a few thousand human-labeled datasets that contain a list of email messages and a label of "spam" or "not spam" for each email message. A machine learning (ML) specialist wants to use transfer learning with a Bidirectional Encoder Representations from Transformers (BERT) model that is trained on English Wikipedia text data.

What should the ML specialist do to initialize the model to fine-tune the model with the custom data?

- A. Initialize the model with pretrained weights in all layers except the last fully connected layer.
- B. Initialize the model with pretrained weights in all layers. Stack a classifier on top of the first output position. Train the classifier with the labeled data.
- C. Initialize the model with random weights in all layers. Replace the last fully connected layer with a classifier. Train the classifier with the labeled data.
- D. Initialize the model with pretrained weights in all layers. Replace the last fully connected layer with a classifier. Train the classifier with the labeled data.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 208

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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:

- Call transcription in multiple languages
- Categorization of calls based on the transcript
- Detection of the main customer issues in the calls
- 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.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 209

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 210

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A retail company wants to build a recommendation system for the company's website. The system needs to provide recommendations for existing users and needs to base those recommendations on each user's past browsing history. The system also must filter out any items that the user previously purchased.

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

- A. Train a model by using a user-based collaborative filtering algorithm on Amazon SageMaker. Host the model on a SageMaker real-time endpoint. Configure an Amazon API Gateway API and an AWS Lambda function to handle real-time inference requests that the web application sends. Exclude the items that the user previously purchased from the results before sending the results back to the web application.
- B. Use an Amazon Personalize PERSONALIZED_RANKING recipe to train a model. Create a real-time filter to exclude items that the user previously purchased. Create and deploy a campaign on Amazon Personalize. Use the GetPersonalizedRanking API operation to get the real-time recommendations.
- C. Use an Amazon Personalize USER_PERSONALIZATION recipe to train a model. Create a real-time filter to exclude items that the user previously purchased. Create and deploy a campaign on Amazon Personalize. Use the GetRecommendations API operation to get the real-time recommendations.
- D. Train a neural collaborative filtering model on Amazon SageMaker by using GPU instances. Host the model on a SageMaker real-time endpoint. Configure an Amazon API Gateway API and an AWS Lambda function to handle real-time inference requests that the web application sends. Exclude the items that the user previously purchased from the results before sending the results back to the web application.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 211

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A bank wants to use a machine learning (ML) model to predict if users will default on credit card payments. The training data consists of 30,000 labeled records and is evenly balanced between two categories. For the model, an ML specialist selects the Amazon SageMaker built-in XGBoost algorithm and configures a SageMaker automatic hyperparameter optimization job with the Bayesian method. The ML specialist uses the validation accuracy as the objective metric.

When the bank implements the solution with this model, the prediction accuracy is 75%. The bank has given the ML specialist 1 day to improve the model in production.

Which approach is the FASTEST way to improve the model's accuracy?

- A. Run a SageMaker incremental training based on the best candidate from the current model's tuning job. Monitor the same metric that was used as the objective metric in the previous tuning, and look for improvements.
- B. Set the Area Under the ROC Curve (AUC) as the objective metric for a new SageMaker automatic hyperparameter tuning job. Use the same maximum training jobs parameter that was used in the previous tuning job.
- C. Run a SageMaker warm start hyperparameter tuning job based on the current model's tuning job. Use the same objective metric that was used in the previous tuning.
- D. Set the F1 score as the objective metric for a new SageMaker automatic hyperparameter tuning job. Double the maximum training jobs parameter that was used in the previous tuning job.

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 212

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

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.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 213

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company is building a pipeline that periodically retrains its machine learning (ML) models by using new streaming data from devices. The company's data engineering team wants to build a data ingestion system that has high throughput, durable storage, and scalability. The company can tolerate up to 5 minutes of latency for data ingestion. The company needs a solution that can apply basic data transformation during the ingestion process.

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

- A. Configure the devices to send streaming data to an Amazon Kinesis data stream. Configure an Amazon Kinesis Data Firehose delivery stream to automatically consume the Kinesis data stream, transform the data with an AWS Lambda function, and save the output into an Amazon S3 bucket.
- B. Configure the devices to send streaming data to an Amazon S3 bucket. Configure an AWS Lambda function that is invoked by S3 event notifications to transform the data and load the data into an Amazon Kinesis data stream. Configure an Amazon Kinesis Data Firehose delivery stream to automatically consume the Kinesis data stream and load the output back into the S3 bucket.
- C. Configure the devices to send streaming data to an Amazon S3 bucket. Configure an AWS Glue job that is invoked by S3 event notifications to read the data, transform the data, and load the output into a new S3 bucket.
- D. Configure the devices to send streaming data to an Amazon Kinesis Data Firehose delivery stream. Configure an AWS Glue job that connects to the delivery stream to transform the data and load the output into an Amazon S3 bucket.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 214

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A retail company is ingesting purchasing records from its network of 20,000 stores to Amazon S3 by using Amazon Kinesis Data Firehose. The company uses a small, server-based application in each store to send the data to AWS over the internet. The company uses this data to train a machine learning model that is retrained each day. The company's data science team has identified existing attributes on these records that could be combined to create an improved model.

Which change will create the required transformed records with the LEAST operational overhead?

- A. Create an AWS Lambda function that can transform the incoming records. Enable data transformation on the ingestion Kinesis Data Firehose delivery stream. Use the Lambda function as the invocation target.
- B. Deploy an Amazon EMR cluster that runs Apache Spark and includes the transformation logic. Use Amazon EventBridge (Amazon CloudWatch Events) to schedule an AWS Lambda function to launch the cluster each day and transform the records that accumulate in Amazon S3. Deliver the transformed records to Amazon S3.
- C. Deploy an Amazon S3 File Gateway in the stores. Update the in-store software to deliver data to the S3 File Gateway. Use a scheduled daily AWS Glue job to transform the data that the S3 File Gateway delivers to Amazon S3.
- D. Launch a fleet of Amazon EC2 instances that include the transformation logic. Configure the EC2 instances with a daily cron job to transform the records that accumulate in Amazon S3. Deliver the transformed records to Amazon S3.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 215

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A sports broadcasting company is planning to introduce subtitles in multiple languages for a live broadcast. The commentary is in English. The company needs the transcriptions to appear on screen in French or Spanish, depending on the broadcasting country. The transcriptions must be able to capture domain-specific terminology, names, and locations based on the commentary context. The company needs a solution that can support options to provide tuning data.

Which combination of AWS services and features will meet these requirements with the LEAST operational overhead? (Choose two.)

- A. Amazon Transcribe with custom vocabularies
- B. Amazon Transcribe with custom language models
- C. Amazon SageMaker Seq2Seq
- D. Amazon SageMaker with Hugging Face Speech2Text
- E. Amazon Translate

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 216

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A data scientist at a retail company is forecasting sales for a product over the next 3 months. After preliminary analysis, the data scientist identifies that sales are seasonal and that holidays affect sales. The data scientist also determines that sales of the product are correlated with sales of other products in the same category.

The data scientist needs to train a sales forecasting model that incorporates this information.

Which solution will meet this requirement with the LEAST development effort?

- A. Use Amazon Forecast with Holidays featurization and the built-in autoregressive integrated moving average (ARIMA) algorithm to train the model.
- B. Use Amazon Forecast with Holidays featurization and the built-in DeepAR+ algorithm to train the model.
- C. Use Amazon SageMaker Processing to enrich the data with holiday information. Train the model by using the SageMaker DeepAR built-in algorithm.
- D. Use Amazon SageMaker Processing to enrich the data with holiday information. Train the model by using the Gluon Time Series (GluonTS) toolkit.

Show Suggested Answer



Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 217

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company is building a predictive maintenance model for its warehouse equipment. The model must predict the probability of failure of all machines in the warehouse. The company has collected 10,000 event samples within 3 months. The event samples include 100 failure cases that are evenly distributed across 50 different machine types.

How should the company prepare the data for the model to improve the model's accuracy?

- A. Adjust the class weight to account for each machine type.
- B. Oversample the failure cases by using the Synthetic Minority Oversampling Technique (SMOTE).
- C. Undersample the non-failure events. Stratify the non-failure events by machine type.
- D. Undersample the non-failure events by using the Synthetic Minority Oversampling Technique (SMOTE).

Show Suggested Answer





Exam question from Amazon's AWS Certified Machine Learning - Specialty

Question #: 218

Topic #: 1

[\[All AWS Certified Machine Learning - Specialty Questions\]](#)

A company stores its documents in Amazon S3 with no predefined product categories. A data scientist needs to build a machine learning model to categorize the documents for all the company's products.

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

- A. Build a custom clustering model. Create a Dockerfile and build a Docker image. Register the Docker image in Amazon Elastic Container Registry (Amazon ECR). Use the custom image in Amazon SageMaker to generate a trained model.
- B. Tokenize the data and transform the data into tabular data. Train an Amazon SageMaker k-means model to generate the product categories.
- C. Train an Amazon SageMaker Neural Topic Model (NTM) model to generate the product categories.
- D. Train an Amazon SageMaker Blazing Text model to generate the product categories.

Show Suggested Answer

