Amazon

Exam Questions AWS-Certified-Machine-Learning-Specialty

AWS Certified Machine Learning - Specialty

About Exambible

Your Partner of IT Exam

Found in 1998

Exambible is a company specialized on providing high quality IT exam practice study materials, especially Cisco CCNA, CCDA, CCNP, CCIE, Checkpoint CCSE, CompTIA A+, Network+ certification practice exams and so on. We guarantee that the candidates will not only pass any IT exam at the first attempt but also get profound understanding about the certificates they have got. There are so many alike companies in this industry, however, Exambible has its unique advantages that other companies could not achieve.

Our Advances

* 99.9% Uptime

　　All examinations will be up to date.

* 24/7 Quality Support

　　We will provide service round the clock.

* 100% Pass Rate

　　Our guarantee that you will pass the exam.

* Unique Gurantee

　　If you do not pass the exam at the first time, we will not only arrange FULL REFUND for you, but also provide you another exam of your claim, ABSOLUTELY FREE!

NEW QUESTION 1
A Machine Learning Specialist observes several performance problems with the training portion of a machine learning solution on Amazon SageMaker The solution uses a large training dataset 2 TB in size and is using the SageMaker k-means algorithm The observed issues include the unacceptable length of time it takes before the training job launches and poor I/O throughput while training the model
What should the Specialist do to address the performance issues with the current solution?

A. Use the SageMaker batch transform feature
B. Compress the training data into Apache Parquet format.
C. Ensure that the input mode for the training job is set to Pipe.
D. Copy the training dataset to an Amazon EFS volume mounted on the SageMaker instance.

Answer: B

NEW QUESTION 2
A 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.
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.

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

A. Binarization
B. One-hot encoding
C. Tokenization
D. Normalization transformation

Answer: B

NEW QUESTION 3
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 instanc
B. Ingest tables without sensitive data through an AWS Site-to-Site VPN connection directly into Amazon S3.
C. Create an AWS Glue job to connect to the PostgreSQL DB instanc
D. Ingest all data through an AWS Site- to-Site VPN connection into Amazon S3 while removing sensitive data using a PySpark job.
E. Use AWS Database Migration Service (AWS DMS) with table mapping to select PostgreSQL tables with no sensitive data through an SSL connectio
F. Replicate data directly into Amazon S3.
G. Use PostgreSQL logical replication to replicate all data to PostgreSQL in Amazon EC2 through AWS Direct Connect with a VPN connectio
H. Use AWS Glue to move data from Amazon EC2 to Amazon S3.

Answer: C

NEW QUESTION 4
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 uniqueAmazon Kinesis Video Streams video strea
B. On each stream, use Amazon Rekognition Video and createa stream processor to detect faces from a collection of known employees, and alert when non-employees are detected.
C. Use a proxy server at each local office and for each camera, and stream the RTSP feed to a uniqueAmazon Kinesis Video Streams video strea
D. On each stream, use Amazon Rekognition Image to detectfaces from a collection of known employees and alert when non-employees are detected.
E. Install AWS DeepLens cameras and use the DeepLens_Kinesis_Video module to stream video to Amazon Kinesis Video Streams for each camer
F. On each stream, use Amazon Rekognition Video and create a stream processor to detect faces from a collection on each stream, and alert when nonemployees

are detected.
G. Install AWS DeepLens cameras and use the DeepLens_Kinesis_Video module to stream video to Amazon Kinesis Video Streams for each camer
H. 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.

Answer: C

NEW QUESTION 5
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? (Select 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

Answer: CD

Explanation:
The PCA and K-means algorithms are useful in collection of data using census form.

NEW QUESTION 6
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.
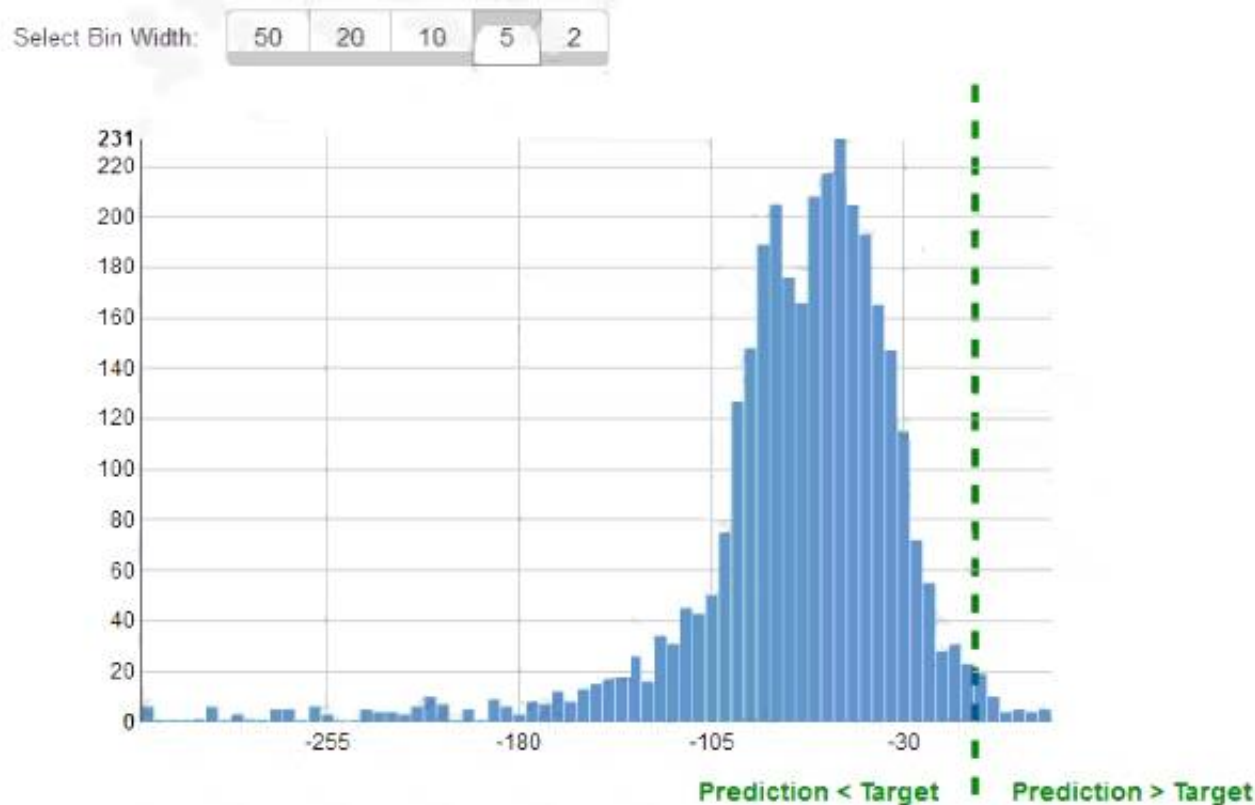
Answer: C

Explanation:
If this is a ComputerVision problem augmentation can help and we may consider A an option. However in analyzing customer historic data, there is no easy way to increase randomization in training. If you go deep into modelling and coding. When you build model with tensorflow/pytorch, most of the time the trainloader is already sampling in data in random manner (with shuffle enable). What we usually do to reduce overfitting is by adding dropout.
https://docs.aws.amazon.com/machine-learning/latest/dg/model-fit-underfitting-vs-overfitting.html

NEW QUESTION 7
While reviewing the histogram for residuals on regression evaluation data a Machine Learning Specialist notices that the residuals do not form a zero-centered bell shape as shown What does this mean?

A. The model might have prediction errors over a range of target values.
B. The dataset cannot be accurately represented using the regression model
C. There are too many variables in the model
D. The model is predicting its target values perfectly.

Answer: D

NEW QUESTION 8
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. An XGBoost 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)

Answer: A

NEW QUESTION 9
A Machine Learning Specialist was given a dataset consisting of unlabeled data The Specialist must create a model that can help the team classify the data into different buckets What model should be used to complete this work?

A. K-means clustering
B. Random Cut Forest (RCF)
C. XGBoost
D. BlazingText

Answer: A

NEW QUESTION 10
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

Answer: A

NEW QUESTION 10
A large mobile network operating company is building a machine learning model to predict customers who are likely to unsubscribe from the service. The company plans to offer an incentive for these customers as the cost of churn is far greater than the cost of the incentive.
The model produces the following confusion matrix after evaluating on a test dataset of 100 customers: Based on the model evaluation results, why is this a viable model for production?

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

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

Answer: A

NEW QUESTION 15
A Machine Learning Specialist is using Apache Spark for pre-processing training data As part of the Spark pipeline, the Specialist wants to use Amazon SageMaker for training a model and hosting it Which of the following would the Specialist do to integrate the Spark application with SageMaker? (Select THREE)

A. Download the AWS SDK for the Spark environment
B. Install the SageMaker Spark library in the Spark environment.
C. Use the appropriate estimator from the SageMaker Spark Library to train a model.
D. Compress the training data into a ZIP file and upload it to a pre-defined Amazon S3 bucket.
E. Use the sageMakerMode
F. transform method to get inferences from the model hosted in SageMaker
G. Convert the DataFrame object to a CSV file, and use the CSV file as input for obtaining inferences from SageMaker.

Answer: DEF

NEW QUESTION 19
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 long Amazon SageMaker API calls to Amazon S3. Add code to push a custom metric to Amazon CloudWatc

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

Answer: C

NEW QUESTION 23
A Machine Learning Specialist is preparing data for training on Amazon SageMaker The Specialist is transformed into a numpy .array, which appears to be negatively affecting the speed of the training
What should the Specialist do to optimize the data for training on SageMaker'?

A. Use the SageMaker batch transform feature to transform the training data into a DataFrame
B. Use AWS Glue to compress the data into the Apache Parquet format
C. Transform the dataset into the Recordio protobuf format
D. Use the SageMaker hyperparameter optimization feature to automatically optimize the data

Answer: C

NEW QUESTION 28
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)

Answer: C

NEW QUESTION 32
A manufacturer of car engines collects data from cars as they are being driven The data collected includes timestamp, engine temperature, rotations per minute (RPM), and other sensor readings The company wants to predict when an engine is going to have a problem so it can notify drivers in advance to get engine maintenance The engine data is loaded into a data lake for training
Which is the MOST suitable predictive model that can be deployed into production'?

A. Add labels over time to indicate which engine faults occur at what time in the future to turn this into a supervised learning problem Use a recurrent neural network (RNN) to train the model to recognize when an engine might need maintenance for a certain fault.
B. This data requires an unsupervised learning algorithm Use Amazon SageMaker k-means to cluster the data
C. Add labels over time to indicate which engine faults occur at what time in the future to turn this into a supervised learning problem Use a convolutional neural network (CNN) to train the model to recognize when an engine might need maintenance for a certain fault.
D. This data is already formulated as a time series Use Amazon SageMaker seq2seq to model the time series.

Answer: B

NEW QUESTION 34
An e-commerce company needs a customized training model to classify images of its shirts and pants products The company needs a proof of concept in 2 to 3 days with good accuracy Which compute choice should the Machine Learning Specialist select to train and achieve good accuracy on the model quickly?

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

Answer: C

NEW QUESTION 39
An Machine Learning Specialist discover the following statistics while experimenting on a model.

Experiment 1
Baseline model
Train error = 5%
Test error = 16%

Experiment 2
The Specialist added more layers and neurons to the model and received the following results:
Train error = 5.2%
Test error = 15.7%

Experiment 3
The Specialist reverted back to the original number of neurons from Experiment 1 and implemented regularization in the neural network, which yielded the following results:
Train error = 4.7%
Test error = 9.5%

What can the Specialist from the experiments?

A. The model In Experiment 1 had a high variance error lhat was reduced in Experiment 3 by regularization Experiment 2 shows that there is minimal bias error in Experiment 1
B. The model in Experiment 1 had a high bias error that was reduced in Experiment 3 by regularization Experiment 2 shows that there is minimal variance error in Experiment 1
C. The model in Experiment 1 had a high bias error and a high variance error that were reduced in Experiment 3 by regularization Experiment 2 shows thai high bias cannot be reduced by increasing layers and neurons in the model
D. The model in Experiment 1 had a high random noise error that was reduced in Expenment 3 by regularization Expenment 2 shows that random noise cannot be reduced by increasing layers and neurons in the model

Answer: C

NEW QUESTION 42
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 phase 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 phase extraction, and classification tasks.
B. Use an NLP library in Amazon SageMaker for the part-of-speech taggin
C. Use Amazon Comprehend for the key phase extractio
D. Use AWS Deep Learning Containers with Amazon SageMaker to build the custom classifier.
E. Use Amazon Comprehend for the part-of-speech tagging and key phase extraction task
F. Use Amazon SageMaker built-in Latent Dirichlet Allocation (LDA) algorithm to build the custom classifier.
G. Use Amazon Comprehend for the part-of-speech tagging and key phase extraction task
H. Use AWS Deep Learning Containers with Amazon SageMaker to build the custom classifier.

Answer: B

NEW QUESTION 43
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 customizatio
B. Create customer slots within the bot that specifically identify each of the required product name
C. Use the Amazon Lex synonym mechanism to provide additional variations of each product name as mis-transcriptions are identified in development.
D. Use Amazon Transcribe to perform the ASR customizatio
E. 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 valu
F. Use this updated custom vocabulary file in all future transcription tasks.
G. Create a custom vocabulary file containing each product name with phonetic pronunciations, and use it with Amazon Transcribe to perform the ASR customizatio
H. 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.
I. Use the audio transcripts to create a training dataset and build an Amazon Transcribe custom language mode
J. Analyze the transcripts and update the training dataset with a manually corrected version of transcripts where product names are not being transcribed correctl
K. Create an updated custom language model.

Answer: A

NEW QUESTION 44
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? (Select TWO)

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

Answer: BC

Explanation:
https://aws.amazon.com/jp/emr/?whats-new-cards.sort-by=item.additionalFields.postDateTime&whats-new-car

NEW QUESTION 46
A large company has developed a B1 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 interlaces
Which combination of services can be used to build this conversational interface? (Select THREE)

A. Alexa for Business
B. Amazon Connect
C. Amazon Lex

D. Amazon Poly
E. Amazon Comprehend
F. Amazon Transcribe

Answer: BEF

NEW QUESTION 51
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 mode
B. Use an Amazon Kinesis Data Stream to process the incoming webtrafficdat
C. Attach a preprocessing AWS Lambda function to perform data enrichment by calling the RCF modelto calculate the anomaly score for each record.
D. Use historic web traffic data to train an anomaly detection model using the Amazon SageMaker built-inXGBoost mode
E. Use an Amazon Kinesis Data Stream to process the incoming web traffic dat
F. Attach apreprocessing AWS Lambda function to perform data enrichment by calling the XGBoost model to calculate the anomaly score for each record.
G. Collect the streaming data using Amazon Kinesis Data Firehos
H. Map the delivery stream as an inputsource for Amazon Kinesis Data Analytic
I. Write a SQL query to run in real time against the streaming datawith the k-Nearest Neighbors (kNN) SQL extension to calculate anomaly scores for each record using a tumbling window.
J. Collect the streaming data using Amazon Kinesis Data Firehos
K. Map the delivery stream as an inputsource for Amazon Kinesis Data Analytic
L. Write a SQL query to run in real time against the streaming datawith the Amazon Random Cut Forest (RCF) SQL extension to calculate anomaly scores for each record using a sliding window.

Answer: D

NEW QUESTION 53
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.

Answer: AC

NEW QUESTION 56
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

Answer: A

NEW QUESTION 60
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? (Select TWO.)

A. AWS CloudTrail
B. AWS Health
C. AWS Trusted Advisor
D. Amazon CloudWatch
E. AWS Config

Answer: AD

NEW QUESTION 65
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 bucke
B. Create an ECS cluster that is based on an AWS Deep Learning Containers imag
C. Write the code to perform the feature engineerin
D. Train a logistic regression model for predicting the price, pointing to the bucket with the datase
E. Wait for the training job to complet
F. Perform the inferences.
G. Create an Amazon SageMaker notebook with a new IAM role that is associated with the noteboo
H. Pull the dataset from the S3 bucke
I. Explore different combinations of feature engineering transformations,regression algorithms, and hyperparameter
J. Compare all the results in the notebook, and deploy the most accurate configuration in an endpoint for predictions.
K. Create an IAM role with access to Amazon S3, Amazon SageMaker, and AWS Lambd
L. Create a training job with the SageMaker built-in XGBoost model pointing to the bucket with the datase
M. Specify the price as the target featur
N. Wait for the job to complet
O. Load the model artifact to a Lambda function for inference on prices of new houses.
P. Create an IAM role for Amazon SageMaker with access to the S3 bucke
Q. Create a SageMaker AutoML job with SageMaker Autopilot pointing to the bucket with the datase
R. Specify the price as the target attribut
S. Wait for the job to complet
T. Deploy the best model for predictions.

Answer: A

NEW QUESTION 68
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)

Answer: B

NEW QUESTION 73
Example Corp has an annual sale event from October to December. The company has sequential sales data from the past 15 years and wants to use Amazon ML to predict the sales for this year's upcoming event. Which method should Example Corp use to split the data into a training dataset and evaluation dataset?

A. Pre-split the data before uploading to Amazon S3
B. Have Amazon ML split the data randomly.
C. Have Amazon ML split the data sequentially.
D. Perform custom cross-validation on the data

Answer: C

NEW QUESTION 78
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 OuickSight 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 data as it is generated by Amazon SageMaker
D. Send Amazon CloudWatch Logs that were generated by Amazon SageMaker lo Amazon ES and use Kibana to query and visualize the log data.

Answer: B

NEW QUESTION 82
A Machine Learning Specialist is training a model to identify the make and model of vehicles in images The Specialist wants to use transfer learning and an existing model trained on images of general objects The Specialist collated a large custom dataset of pictures containing different vehicle makes and models

A. Initialize the model with random weights in all layers including the last fully connected layer
B. Initialize the model with pre-trained weights in all layers and replace the last fully connected layer.
C. Initialize the model with random weights in all layers and replace the last fully connected layer
D. Initialize the model with pre-trained weights in all layers including the last fully connected layer

Answer: D

NEW QUESTION 84
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 quenes
D. Use AWS Lambda to transform the data and Amazon Kinesis Data Analytics to run queries

Answer: D

NEW QUESTION 86
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 tex
D. Use the Amazon SageMaker Neural Topic Model (NTM) and Object Detection algorithms to tag data into distinct categories/classes.
E. Use the AWS Deep Learning AMI and Amazon EC2 GPU instances to create custom models for audio transcription and topic modeling, and use object detection to tag data into distinct categories/classes.

Answer: A

NEW QUESTION 89
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 metric
B. Run the script as an AWS Glue job.
C. Add an Amazon SageMaker Experiments tracker to the script to capture key metric
D. Run the script as an AWS Glue job.
E. Add an Amazon SageMaker Debugger hook to the script to capture key parameter
F. Run the script as a SageMaker processing job.
G. Add an Amazon SageMaker Experiments tracker to the script to capture key parameter
H. Run the script as a SageMaker processing job.

Answer: B

NEW QUESTION 92
An Amazon SageMaker notebook instance is launched into Amazon VPC The SageMaker notebook references data contained in an Amazon S3 bucket in another account The bucket is encrypted using SSE-KMS The instance returns an access denied error when trying to access data in Amazon S3.
Which of the following are required to access the bucket and avoid the access denied error? (Select THREE)

A. An AWS KMS key policy that allows access to the customer master key (CMK)
B. A SageMaker notebook security group that allows access to Amazon S3
C. An 1AM role that allows access to the specific S3 bucket
D. A permissive S3 bucket policy
E. An S3 bucket owner that matches the notebook owner
F. A SegaMaker notebook subnet ACL that allow traffic to Amazon S3.

Answer: ACF

NEW QUESTION 95
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

Answer: DE

NEW QUESTION 100
A data engineer at a bank is evaluating a new tabular dataset that includes customer data. The data engineer will use the customer data to create a new model to predict customer behavior. After creating a correlation matrix for the variables, the data engineer notices that many of the 100 features are highly correlated with each other.
Which steps should the data engineer take to address this issue? (Choose two.)

A. Use a linear-based algorithm to train the model.
B. Apply principal component analysis (PCA).
C. Remove a portion of highly correlated features from the dataset.
D. Apply min-max feature scaling to the dataset.
E. Apply one-hot encoding category-based variables.

Answer: BD

NEW QUESTION 104

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      1
Actual    0 99,966 | 34
          1    877 | 123
```
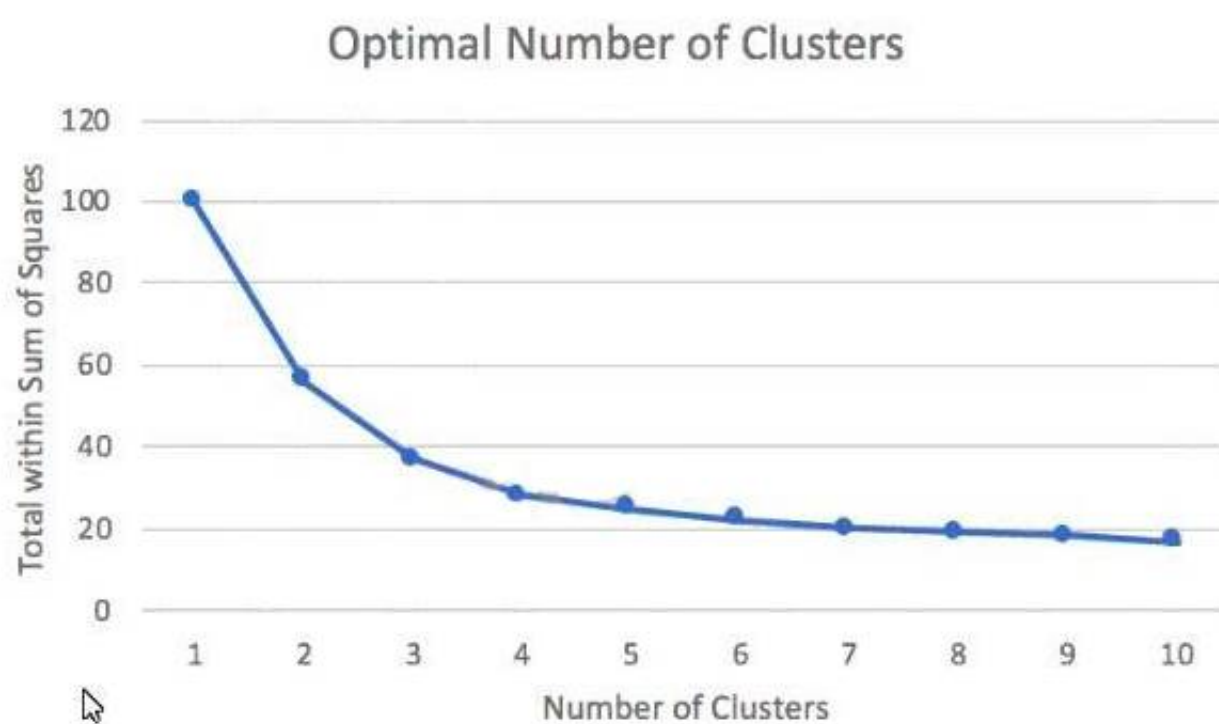
Which combination of steps should the Data Scientist take to reduce the number of false negative predictions by the model? (Choose two.)

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.

Answer: BD

NEW QUESTION 106

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

Answer: C

NEW QUESTION 108

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

Answer: C

NEW QUESTION 113

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 bucke
B. 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.
C. Switch to using an Amazon Rekognition collection to store the image
D. Use the IndexFaces andSearchFacesByImage API operations instead of the CompareFaces API operation.
E. Switch to using the AWS GovCloud (US) Region for Amazon S3 to store images and for Amazon Rekognition to compare face
F. Set up a VPN connection and only call the Amazon Rekognition API operations through the VPN.
G. Enable client-side encryption on the S3 bucke
H. Set up a VPN connection and only call the Amazon Rekognition API operations through the VPN.

Answer: B

NEW QUESTION 114
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					
		A	B	C	D	E	Total
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					
		A	B	C	D	E	Total
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 form 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.

Answer: B

NEW QUESTION 115
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

Answer: C

NEW QUESTION 116
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 VP
B. Place the Amazon SageMaker endpoint and S3 buckets within the same VPC.
C. Associate the Amazon SageMaker notebook with a private subnet in a VP
D. Use 1AM policies to grant access to Amazon S3 and Amazon SageMaker.
E. Associate the Amazon SageMaker notebook with a private subnet in a VP
F. Ensure the VPC has S3 VPC endpoints and Amazon SageMaker VPC endpoints attached to it.
G. Associate the Amazon SageMaker notebook with a private subnet in a VP
H. Ensure the VPC has a NAT gateway and an associated security group allowing only outbound connections to Amazon S3 and Amazon SageMaker

Answer: D

NEW QUESTION 121
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 post
B. Expose an endpoint that can be called by AWS Lambd
C. 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 metri
D. Use CloudWatch alarms to notify analysts of trends.
E. Train a model in Amazon SageMaker by using the semantic segmentation algorithm to model the semantic content in the corpus of social media post
F. Expose an endpoint that can be called by AWS Lambd
G. Trigger a Lambda function when objects are added to the S3 bucket to invoke the endpoint and record the sentiment in an Amazon DynamoDB tabl
H. 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.
I. Trigger an AWS Lambda function when social media posts are added to the S3 bucke
J. Call Amazon Comprehend for each post to capture the sentiment in the message and record the sentiment in an Amazon DynamoDB tabl
K. 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.
L. Trigger an AWS Lambda function when social media posts are added to the S3 bucke
M. 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.

Answer: A

NEW QUESTION 122
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) tothe Amazon SageMaker notebook instance.
B. onfigure the Amazon SageMaker notebook instance to have access to the VP
C. Grant permission in the KMS key policy to the notebook's KMS role.
D. Assign an IAM role to the Amazon SageMaker notebook with S3 read access to the datase
E. Grant permission in the KMS key policy to that role.
F. Assign the same KMS key used to encrypt data in Amazon S3 to the Amazon SageMaker notebook instance.

Answer: D

NEW QUESTION 124
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 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 bucke
B. Determine which source files do not exist in the destination S3 bucket and which source files were modified.
C. Use AWS Transfer for FTPS to transfer the files from the on-premises storage to Amazon S3.
D. Use AWS DataSync to make an initial copy of the entire datase
E. Schedule subsequent incremental transfers of changing data until the final cutover from on premises to AWS.
F. Use S3 Batch Operations to pull data periodically from the on-premises storag
G. Enable S3 Versioning on the S3 bucket to protect against accidental overwrites.

Answer: C

Explanation:
Configure DataSync to make an initial copy of your entire dataset, and schedule subsequent incremental transfers of changing data until the final cut-over from on-premises to AWS.

NEW QUESTION 127
A manufacturing company asks its Machine Learning Specialist to develop a model that classifies defective parts into one of eight defect types. The company has provided roughly 100000 images per defect type for training During the initial training of the image classification model the Specialist notices that the validation accuracy is 80%, while the training accuracy is 90% It is known that human-level performance for this type of image classification is around 90%
What should the Specialist consider to fix this issue1?

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

Answer: D

NEW QUESTION 130
A company has an ecommerce website with a product recommendation engine built in TensorFlow. The recommendation engine endpoint is hosted by Amazon SageMaker. Three compute-optimized instances support the expected peak load of the website.
Response times on the product recommendation page are increasing at the beginning of each month. Some users are encountering errors. The website receives the majority of its traffic between 8 AM and 6 PM on weekdays in a single time zone.
Which of the following options are the MOST effective in solving the issue while keeping costs to a minimum? (Choose two.)

A. Configure the endpoint to use Amazon Elastic Inference (EI) accelerators.
B. Create a new endpoint configuration with two production variants.
C. Configure the endpoint to automatically scale with the InvocationsPerInstance metric.
D. Deploy a second instance pool to support a blue/green deployment of models.
E. Reconfigure the endpoint to use burstable instances.

Answer: BD

NEW QUESTION 135
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.

Answer: B

NEW QUESTION 139
While working on a neural network project, a Machine Learning Specialist discovers thai some features in the data have very high magnitude resulting in this data being weighted more in the cost function What should the Specialist do to ensure better convergence during backpropagation?

A. Dimensionality reduction
B. Data normalization
C. Model regulanzation
D. Data augmentation for the minority class

Answer: D

NEW QUESTION 141
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 cod
B. Tag the Docker image with the registry hostname and upload it to Amazon ECR.
C. Serialize the trained model so the format is compressed for deploymen
D. Tag the Docker image with the registry hostname and upload it to Amazon S3.
E. Serialize the trained model so the format is compressed for deploymen
F. Build the image and upload it to Docker Hub.
G. Build the Docker image with the inference cod
H. Configure Docker Hub and upload the image to AmazonECR.

Answer: D

NEW QUESTION 143
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 outpu
B. Apply principal component analysis (PCA) with five feature dimensions to predict a segment.
C. The XGBoost model provides a true/false binary outpu
D. Increase the number of classes the XGBoost model predicts to five classes to predict a segment.
E. The XGBoost model is a supervised machine learning algorith
F. Train a k-Nearest-Neighbors (kNN) model with K = 5 on the same dataset to predict a segment.
G. The XGBoost model is a supervised machine learning algorith
H. Train a k-means model with K = 5 on the same dataset to predict a segment.

Answer: C

NEW QUESTION 146
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 instanc
C. Attach Amazon Elastic Inference to the instance.
D. Redeploy the model on a P3dn instance.
E. Deploy the model onto an Amazon Elastic Container Service (Amazon ECS) cluster using a P3 instance.

Answer: B

Explanation:
https://aws.amazon.com/machine-learning/elastic-inference/

NEW QUESTION 148
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.

Answer: C

NEW QUESTION 152
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

Answer: B

NEW QUESTION 153
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 Regio
B. Use the Direct Connect connection to upload the image
C. Increase the size of the instances and the number of instances that are used by the SageMaker endpoint.
D. 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.
E. Use auto scaling for SageMake
F. Set up an AWS Direct Connect connection between the production site and the nearest AWS Regio
G. Use the Direct Connect connection to upload the images.
H. 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 machin
I. 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.

Answer: D

NEW QUESTION 157
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.

Answer: C

Explanation:
https://docs.aws.amazon.com/sagemaker/latest/dg/seq-2-seq-howitworks.html
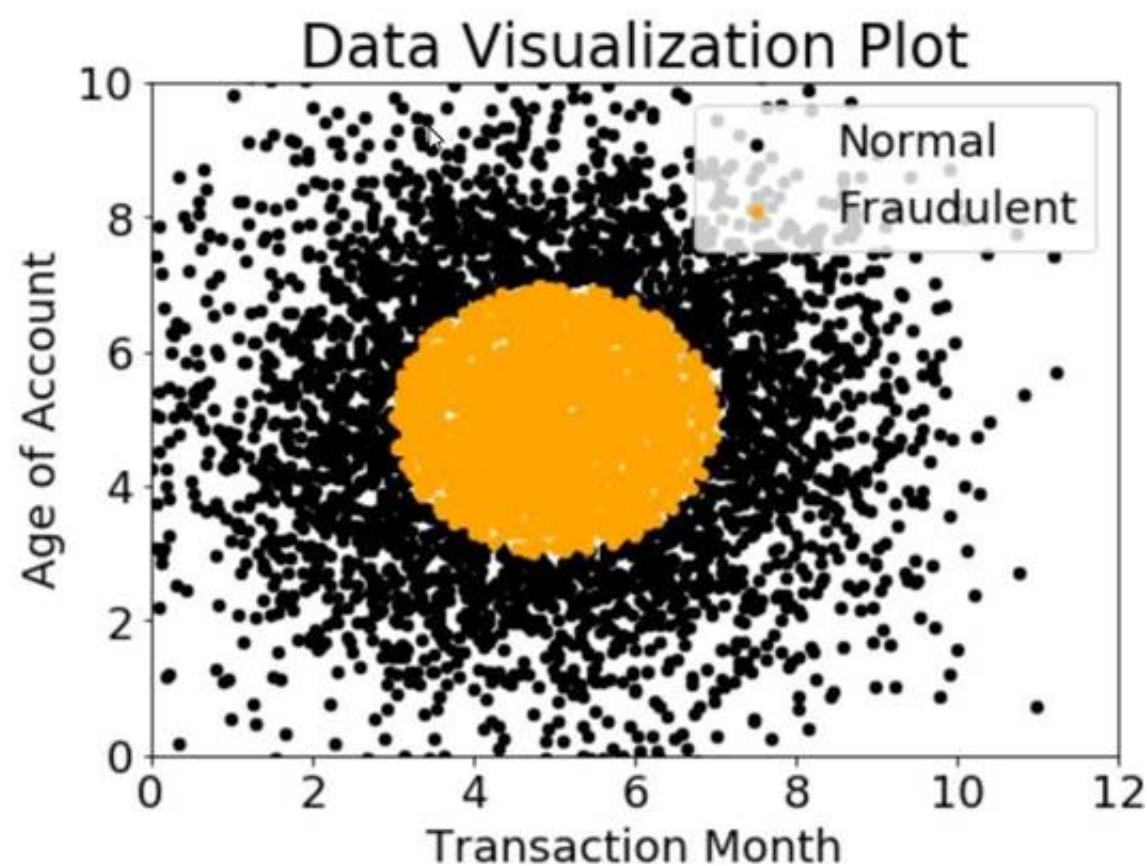

NEW QUESTION 158
A company provisions Amazon SageMaker notebook instances for its data science team and creates Amazon VPC interface endpoints to ensure communication between the VPC and the notebook instances. All connections to the Amazon SageMaker API are contained entirely and securely using the 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 VP
B. Apply this security group to all of the notebook instances' VPC interfaces.
C. Create an IAM policy that allows the sagemaker:CreatePresignedNotebookInstanceUrl and sagemaker:DescribeNotebookInstance actions from only the VPC endpoint
D. Apply this policy to all IAM users, groups, and roles used to access the notebook instances.
E. Add a NAT gateway to the VP
F. Convert all of the subnets where the Amazon SageMaker notebook instances are hosted to private subnet
G. Stop and start all of the notebook instances to reassign only private IP addresses.
H. Change the network ACL of the subnet the notebook is hosted in to restrict access to anyone outside the VPC.

Answer: B

NEW QUESTION 159
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

Answer: C

NEW QUESTION 162
A company is running an Amazon SageMaker training job that will access data stored in its Amazon S3 bucket A compliance policy requires that the data never be transmitted across the internet How should the company set up the job?

A. Launch the notebook instances in a public subnet and access the data through the public S3 endpoint
B. Launch the notebook instances in a private subnet and access the data through a NAT gateway
C. Launch the notebook instances in a public subnet and access the data through a NAT gateway
D. Launch the notebook instances in a private subnet and access the data through an S3 VPC endpoint.

Answer: D

NEW QUESTION 165
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 VP
B. Use the SageMaker DeepAR algorithm to randomize the credit card numbers.
C. 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.
D. Use an Amazon SageMaker launch configuration to encrypt the data once it is copied to the SageMaker instance in a VP
E. Use the SageMaker principal component analysis (PCA) algorithm to reduce the length of the credit card numbers.
F. 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.

Answer: D

NEW QUESTION 166
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	...	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 dat
B. Split off the last 10% of the interaction data for the test set.
C. Identify the most recent 10% of interactions for each use
D. Split off these interactions for the test set.
E. Identify the 10% of users with the least interaction dat
F. Split off all interaction data from these users for the test set.
G. Randomly select 10% of the user
H. Split off all interaction data from these users for the test set.

Answer: B

Explanation:
https://aws.amazon.com/blogs/machine-learning/building-a-customized-recommender-system-in-amazon-sagem

NEW QUESTION 167
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

Answer: D

NEW QUESTION 170
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 lime 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 toserialize 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.

Answer: B

Explanation:
https://medium.com/searce/convert-csv-json-files-to-apache-parquet-using-aws-glue-a760d177b45f https://github.com/ecloudvalley/Building-a-Data-Lake-with-AWS-Glue-and-Amazon-S3

NEW QUESTION 174
A Machine Learning Specialist is working with a large company to leverage machine learning within its products. The company wants to group its customers into categories based on which customers will and will not churn within the next 6 months. The company has labeled the data available to the Specialist.
Which machine learning model type should the Specialist use to accomplish this task?

A. Linear regression
B. Classification
C. Clustering
D. Reinforcement learning

Answer: B

Explanation:
The goal of classification is to determine to which class or category a data point (customer in our case) belongs to. For classification problems, data scientists would use historical data with predefined target variables AKA labels (churner/non-churner) – answers that need to be predicted – to train an algorithm. With classification, businesses can answer the following questions:

- Will this customer churn or not?
- Will a customer renew their subscription?
- Will a user downgrade a pricing plan?
- Are there any signs of unusual customer behavior?

NEW QUESTION 179
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 trfese requirements?

A. Use an AWS Glue crawler to crawl S3 data, an AWS Lambda function to trigger an AWS Glue ETL job, and an AWS Glue Data catalog to search and discover metadata.
B. Use an AWS Glue crawler to crawl S3 data, an AWS Lambda function to trigger an AWS Batch job, and an external Apache Hive metastore to search and discover metadata.
C. Use an AWS Glue crawler to crawl S3 data, an Amazon CloudWatch alarm to trigger an AWS Batch job, and an AWS Glue Data Catalog to search and discover metadata.
D. Use an AWS Glue crawler to crawl S3 data, an Amazon CloudWatch alarm to trigger an AWS Glue ETL job, and an external Apache Hive metastore to search and discover metadata.

Answer: A

NEW QUESTION 180
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? (Select 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 thisinstance as a Jupyter notebook in Amazon SageMaker.

Answer: AD

NEW QUESTION 183
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 se
C. Configure Model Monitor to use the new baseline.
D. Delete the endpoint and recreate it with the original configuration.
E. Retrain the model again by using a combination of the original training set and the new training set.

Answer: B

NEW QUESTION 188
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.

Answer: A

NEW QUESTION 190
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

Answer: B

NEW QUESTION 193
A Data Scientist is building a model to predict customer churn using a dataset of 100 continuous numerical features. The Marketing team has not provided any insight about which features are relevant for churn prediction. The Marketing team wants to interpret the model and see the direct impact of relevant features on the model outcome. While training a logistic regression model, the Data Scientist observes that there is a wide gap between the training and validation set accuracy.
Which methods can the Data Scientist use to improve the model performance and satisfy the Marketing team's needs? (Choose two.)

A. Add L1 regularization to the classifier
B. Add features to the dataset
C. Perform recursive feature elimination
D. Perform t-distributed stochastic neighbor embedding (t-SNE)
E. Perform linear discriminant analysis

Answer: BE

NEW QUESTION 197
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.

Answer: C

NEW QUESTION 200
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.

Answer: B

NEW QUESTION 201
When submitting Amazon SageMaker training jobs using one of the built-in algorithms, which common parameters MUST be specified? (Select THREE.)

A. The training channel identifying the location of training data on an Amazon S3 bucket.
B. The validation channel identifying the location of validation data on an Amazon S3 bucket.
C. The 1AM role that Amazon SageMaker can assume to perform tasks on behalf of the users.
D. Hyperparameters in a JSON array as documented for the algorithm used.
E. The Amazon EC2 instance class specifying whether training will be run using CPU or GPU.
F. The output path specifying where on an Amazon S3 bucket the trained model will persist.

Answer: CEF

NEW QUESTION 203
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

Answer: B

Explanation:
The Amazon SageMaker DeepAR forecasting algorithm is a supervised learning algorithm for forecasting scalar (one-dimensional) time series using recurrent

neural networks (RNN). Classical forecasting methods, such as autoregressive integrated moving average (ARIMA) or exponential smoothing (ETS), fit a single model to each individual time series. They then use that model to extrapolate the time series into the future.

NEW QUESTION 205

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 thecustomer 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

Answer: C

NEW QUESTION 208

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 featur
B. When the new version of the model is ready for release, gradually increase InitialVariantWeight until all users have the updated version.
C. Configure two SageMaker hosted endpoints that serve the different versions of the mode
D. Create an Application Load Balancer (ALB) to route traffic to both endpoints based on the TargetVariant query string paramete
E. Reconfigure the app to send the TargetVariant query string parameter for users who subscribed to the preview featur
F. 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.
G. 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 featur
H. When the new version of the model is ready for release, gradually increase DesiredWeight until all users have the updated version.
I. Configure two SageMaker hosted endpoints that serve the different versions of the mode
J. Create an Amazon Route 53 record that is configured with a simple routing policy and that points to the current version of the mode
K. 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 user
L. 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.

Answer: D

NEW QUESTION 213

An employee found a video clip with audio on a company's social media feed. The language used in the video is Spanish. English is the employee's first language, and they do not understand Spanish. The employee wants to do a sentiment analysis.
What combination of services is the MOST efficient to accomplish the task?

A. Amazon Transcribe, Amazon Translate, and Amazon Comprehend
B. Amazon Transcribe, Amazon Comprehend, and Amazon SageMaker seq2seq
C. Amazon Transcribe, Amazon Translate, and Amazon SageMaker Neural Topic Model (NTM)
D. Amazon Transcribe, Amazon Translate, and Amazon SageMaker BlazingText

Answer: A

NEW QUESTION 215

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

Answer: C

NEW QUESTION 218

A Machine Learning Specialist is assigned a TensorFlow project using Amazon SageMaker for training, and needs to continue working for an extended period with no Wi-Fi access.
Which approach should the Specialist use to continue working?

A. Install Python 3 and boto3 on their laptop and continue the code development using that environment.

B. Download the TensorFlow Docker container used in Amazon SageMaker from GitHub to their local environment, and use the Amazon SageMaker Python SDK to test the code.
C. Download TensorFlow from tensorflow.org to emulate the TensorFlow kernel in the SageMaker environment.
D. Download the SageMaker notebook to their local environment then install Jupyter Notebooks on their laptop and continue the development in a local notebook.

Answer: D

NEW QUESTION 220
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.

Answer: D

NEW QUESTION 223
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

Answer: B

NEW QUESTION 224
A Machine Learning Specialist is building a supervised model that will evaluate customers' satisfaction with their mobile phone service based on recent usage The model's output should infer whether or not a customer is likely to switch to a competitor in the next 30 days
Which of the following modeling techniques should the Specialist use1?

A. Time-series prediction
B. Anomaly detection
C. Binary classification
D. Regression

Answer: D

NEW QUESTION 225
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 (-Distributed Stochastic Neighbor Embedding (I-SNE) to visualize the large number of input variables in an easier-to-read dimension.
C. A scatter plot showing (he performance of the objective metric over each training iteration
D. A scatter plot showing the correlation between maximum tree depth and the objective metric.

Answer: D

NEW QUESTION 230
An interactive online dictionary wants to add a widget that displays words used in similar contexts. A Machine Learning Specialist is asked to provide word features for the downstream nearest neighbor model powering the widget.
What should the Specialist do to meet these requirements?

A. Create one-hot word encoding vectors.
B. Produce a set of synonyms for every word using Amazon Mechanical Turk.
C. Create word embedding factors that store edit distance with every other word.
D. Download word embedding's pre-trained on a large corpus.

Answer: D

NEW QUESTION 231
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 and 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

Answer: C

NEW QUESTION 232
A Machine Learning Specialist is building a model to predict future employment rates based on a wide range of economic factors While exploring the data, the Specialist notices that the magnitude of the input features vary greatly The Specialist does not want variables with a larger magnitude to dominate the model
What should the Specialist do to prepare the data for model training'?

A. Apply quantile binning to group the data into categorical bins to keep any relationships in the data by replacing the magnitude with distribution
B. Apply the Cartesian product transformation to create new combinations of fields that are independent of the magnitude
C. Apply normalization to ensure each field will have a mean of 0 and a variance of 1 to remove any significant magnitude
D. Apply the orthogonal sparse Diagram (OSB) transformation to apply a fixed-size sliding window to generate new features of a similar magnitude.

Answer: C

NEW QUESTION 235
A retail company is selling products through a global online marketplace. The company wants to use machine learning (ML) to analyze customer feedback and identify specific areas for improvement. A developer has built a tool that collects customer reviews from the online marketplace and stores them in an Amazon S3 bucket. This process yields a dataset of 40 reviews. A data scientist building the ML models must identify additional sources of data to increase the size of the dataset.
Which data sources should the data scientist use to augment the dataset of reviews? (Choose three.)

A. Emails exchanged by customers and the company's customer service agents
B. Social media posts containing the name of the company or its products
C. A publicly available collection of news articles
D. A publicly available collection of customer reviews
E. Product sales revenue figures for the company
F. Instruction manuals for the company's products

Answer: BDF

NEW QUESTION 237
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.

Answer: B

NEW QUESTION 238
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%r 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.

Answer: A

NEW QUESTION 240
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 operation
B. Remove the detected words from the blog post dat
C. Replace the blog post data source in the S3 bucket.
D. Run the SageMaker built-in principal component analysis (PCA) algorithm with the blog post data from the S3 bucket as the data sourc
E. Replace the blog post data in the S3 bucket with the results of the training job.
F. Use the SageMaker built-in Object Detection algorithm instead of the NTM algorithm for the training job to process the blog post data.
G. Remove the stopwords from the blog post data by using the Count Vectorizer function in the scikit-learnlibrar
H. Replace the blog post data in the S3 bucket with the results of the vectorizer.

Answer: D

NEW QUESTION 245
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

Answer: B

NEW QUESTION 249
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.

Answer: B

NEW QUESTION 253
For the given confusion matrix, what is the recall and precision of the model?

A. Recall = 0.92 Precision = 0.84
B. Recall = 0.84 Precision = 0.8
C. Recall = 0.92 Precision = 0.8
D. Recall = 0.8 Precision = 0.92

Answer: C

NEW QUESTION 258
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 999.1 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? (Select TWO.)

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

Answer: CD

NEW QUESTION 259
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 yearof data with a predictor_type of regressor.
D. Use the Amazon SageMaker Linear Learner algorithm on the single time series consisting of the full yearof data with a predictor_type of classifier.

Answer: C

NEW QUESTION 260
A Machine Learning Specialist is working with multiple data sources containing billions of records that need to be joined. What feature engineering and model

development approach should the Specialist take with a dataset this large?

A. Use an Amazon SageMaker notebook for both feature engineering and model development
B. Use an Amazon SageMaker notebook for feature engineering and Amazon ML for model development
C. Use Amazon EMR for feature engineering and Amazon SageMaker SDK for model development
D. Use Amazon ML for both feature engineering and model development.

Answer: B

NEW QUESTION 264
......

development approach should the Specialist take with a dataset this large?

A. Use an Amazon SageMaker notebook for both feature engineering and model development
B. Use an Amazon SageMaker notebook for feature engineering and Amazon ML for model development
C. Use Amazon EMR for feature engineering and Amazon SageMaker SDK for model development
D. Use Amazon ML for both feature engineering and model development.

Relate Links

100% Pass Your AWS-Certified-Machine-Learning-Specialty Exam with Exambible Prep Materials

https://www.exambible.com/AWS-Certified-Machine-Learning-Specialty-exam/

Contact us

We are proud of our high-quality customer service, which serves you around the clock 24/7.

Viste - https://www.exambible.com/

Relate Links

100% Pass Your AWS-Certified-Machine-Learning-Specialty Exam with Exambible Prep Materials

https://www.exambible.com/AWS-Certified-Machine-Learning-Specialty-exam/