Exam Questions AWS-Certified-Machine-Learning-Specialty

AWS Certified Machine Learning - Specialty

https://www.2passeasy.com/dumps/AWS-Certified-Machine-Learning-Specialty/

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 Data Scientist is developing a machine learning model to predict future patient outcomes based on information collected about each patient and their treatment plans. The model should output a continuous value as its prediction. The data available includes labeled outcomes for a set of 4,000 patients. The study was conducted on a group of individuals over the age of 65 who have a particular disease that is known to worsen with age.
Initial models have performed poorly. While reviewing the underlying data, the Data Scientist notices that, out of 4,000 patient observations, there are 450 where the patient age has been input as 0. The other features for these observations appear normal compared to the rest of the sample population.
How should the Data Scientist correct this issue?

A. Drop all records from the dataset where age has been set to 0.
B. Replace the age field value for records with a value of 0 with the mean or median value from the dataset.
C. Drop the age feature from the dataset and train the model using the rest of the features.
D. Use k-means clustering to handle missing features.

Answer: A

NEW QUESTION 3
A data scientist has a dataset of machine part images stored in Amazon Elastic File System (Amazon EFS). The data scientist needs to use Amazon SageMaker to create and train an image classification machine learning model based on this dataset. Because of budget and time constraints, management wants the data scientist to create and train a model with the least number of steps and integration work required.
How should the data scientist meet these requirements?

A. Mount the EFS file system to a SageMaker notebook and run a script that copies the data to an Amazon FSx for Lustre file syste
B. Run the SageMaker training job with the FSx for Lustre file system as the data source.
C. Launch a transient Amazon EMR cluste
D. Configure steps to mount the EFS file system and copy the data to an Amazon S3 bucket by using S3DistC
E. Run the SageMaker training job with Amazon S3 as the data source.
F. Mount the EFS file system to an Amazon EC2 instance and use the AWS CLI to copy the data to an Amazon S3 bucke
G. Run the SageMaker training job with Amazon S3 as the data source.
H. Run a SageMaker training job with an EFS file system as the data source.

Answer: A

NEW QUESTION 4
A company ingests machine learning (ML) data from web advertising clicks into an Amazon S3 data lake. Click data is added to an Amazon Kinesis data stream by using the Kinesis Producer Library (KPL). The data is loaded into the S3 data lake from the data stream by using an Amazon Kinesis Data Firehose delivery stream. As the data volume increases, an ML specialist notices that the rate of data ingested into Amazon S3 is relatively constant. There also is an increasing backlog of data for Kinesis Data Streams and Kinesis Data Firehose to ingest.
Which next step is MOST likely to improve the data ingestion rate into Amazon S3?

A. Increase the number of S3 prefixes for the delivery stream to write to.
B. Decrease the retention period for the data stream.
C. Increase the number of shards for the data stream.
D. Add more consumers using the Kinesis Client Library (KCL).

Answer: C

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

A. Create a workforce with AWS Identity and Access Management (IAM). Build a labeling tool on Amazon EC2 Queue images for labeling by using Amazon Simple Queue Service (Amazon SQS). Write the labeling instructions.
B. Create an Amazon Mechanical Turk workforce and manifest fil
C. Create a labeling job by using the built-in image classification task type in Amazon SageMaker Ground Trut
D. Write the labeling instructions.
E. Create a private workforce and manifest fil
F. Create a labeling job by using the built-in bounding box task type in Amazon SageMaker Ground Trut
G. Write the labeling instructions.
H. Create a workforce with Amazon Cognit
I. Build a labeling web application with AWS Amplif
J. Build a labeling workflow backend using AWS Lambd
K. Write the labeling instructions.

Answer: C

Explanation:
https://docs.aws.amazon.com/sagemaker/latest/dg/sms-workforce-private.html

NEW QUESTION 7
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 8
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 9
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 10
A bank wants to launch a low-rate credit promotion. The bank is located in a town that recently experienced economic hardship. Only some of the bank's customers were affected by the crisis, so the bank's credit team must identify which customers to target with the promotion. However, the credit team wants to make sure that loyal customers' full credit history is considered when the decision is made.
The bank's data science team developed a model that classifies account transactions and understands credit eligibility. The data science team used the XGBoost algorithm to train the model. The team used 7 years of bank transaction historical data for training and hyperparameter tuning over the course of several days.
The accuracy of the model is sufficient, but the credit team is struggling to explain accurately why the model denies credit to some customers. The credit team has almost no skill in data science.
What should the data science team do to address this issue in the MOST operationally efficient manner?

A. Use Amazon SageMaker Studio to rebuild the mode
B. Create a notebook that uses the XGBoost training container to perform model trainin
C. Deploy the model at an endpoin
D. Enable Amazon SageMaker Model Monitor to store inference
E. Use the inferences to create Shapley values that help explain model behavio
F. Create a chart that shows features and SHapley Additive explanation (SHAP) values to explain to the credit team how the features affect the model outcomes.
G. Use Amazon SageMaker Studio to rebuild the mode

H. Create a notebook that uses the XGBoost training container to perform model trainin
I. Activate Amazon SageMaker Debugger, and configure it to calculate and collect Shapley value
J. Create a chart that shows features and SHapley Additive explanation (SHAP) values to explain to the credit team how the features affect the model outcomes.
K. Create an Amazon SageMaker notebook instanc
L. Use the notebook instance and the XGBoost library to locally retrain the mode
M. Use the plot_importance() method in the Python XGBoost interface to create a feature importance char
N. Use that chart to explain to the credit team how the features affect the model outcomes.
O. Use Amazon SageMaker Studio to rebuild the mode
P. Create a notebook that uses the XGBoost training container to perform model trainin
Q. Deploy the model at an endpoin
R. Use Amazon SageMakerProcessing to post-analyze the model and create a feature importance explainability chart automatically for the credit team.

Answer: C

NEW QUESTION 10
A Machine Learning Specialist deployed a model that provides product recommendations on a company's website Initially, the model was performing very well and resulted in customers buying more products on average However within the past few months the Specialist has noticed that the effect of product recommendations has diminished and customers are starting to return to their original habits of spending less The Specialist is unsure of what happened, as the model has not changed from its initial deployment over a year ago
Which method should the Specialist try to improve model performance?

A. The model needs to be completely re-engineered because it is unable to handle product inventory changes
B. The model's hyperparameters should be periodically updated to prevent drift
C. The model should be periodically retrained from scratch using the original data while adding a regularization term to handle product inventory changes
D. The model should be periodically retrained using the original training data plus new data as product inventory changes

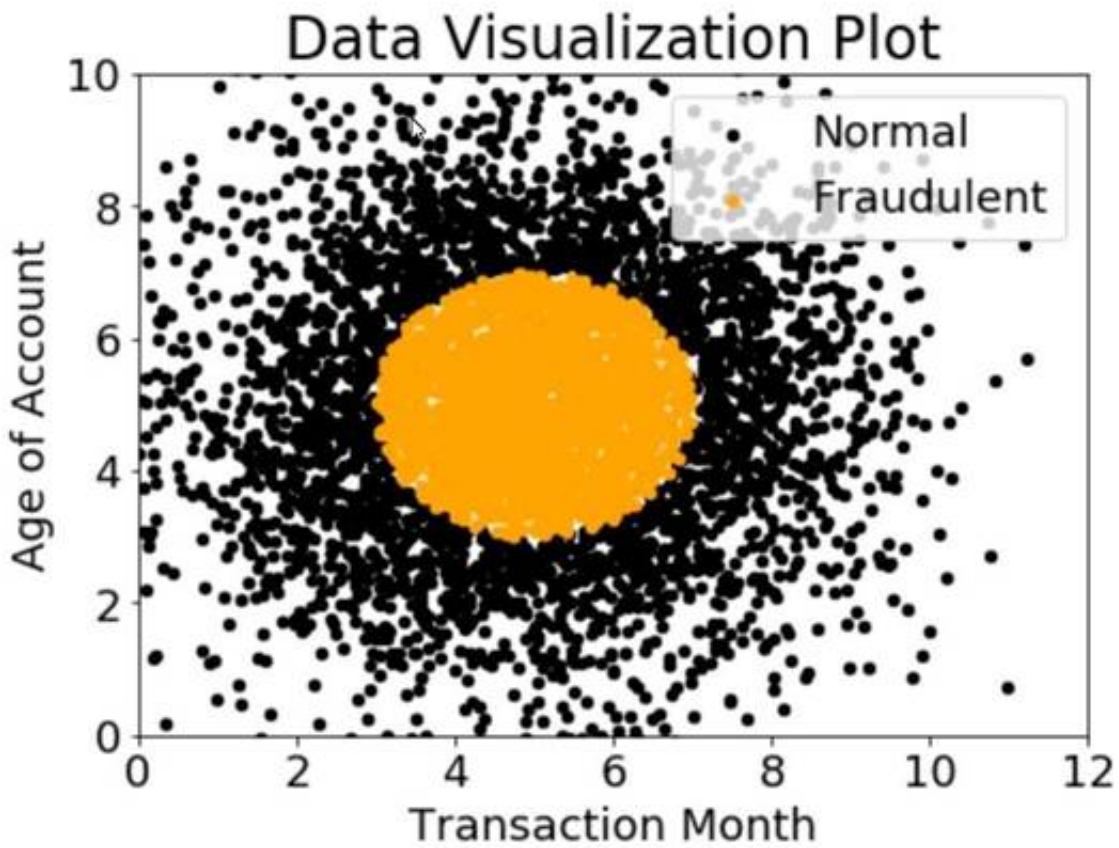
Answer: D

NEW QUESTION 12
A company is setting up an Amazon SageMaker environment. The corporate data security policy does not allow communication over the internet.
How can the company enable the Amazon SageMaker service without enabling direct internet access to Amazon SageMaker notebook instances?

A. Create a NAT gateway within the corporate VPC.
B. Route Amazon SageMaker traffic through an on-premises network.
C. Create Amazon SageMaker VPC interface endpoints within the corporate VPC.
D. Create VPC peering with Amazon VPC hosting Amazon SageMaker.

Answer: A

NEW QUESTION 13
A company wants to classify user behavior as either fraudulent or normal. Based on internal research, a Machine Learning Specialist would like to build a binary classifier based on two features: age of account and transaction month. The class distribution for these features is illustrated in the figure provided.



Based on this information which model would have the HIGHEST accuracy?

A. Long short-term memory (LSTM) model with scaled exponential linear unit (SELL))
B. Logistic regression
C. Support vector machine (SVM) with non-linear kernel
D. Single perceptron with tanh activation function

Answer: C

NEW QUESTION 15
A company needs to quickly make sense of a large amount of data and gain insight from it. The data is in different formats, the schemas change frequently, and new data sources are added regularly. The company wants to use AWS services to explore multiple data sources, suggest schemas, and enrich and transform the

data. The solution should require the least possible coding effort for the data flows and the least possible infrastructure management.
Which combination of AWS services will meet these requirements?

A. Amazon EMR for data discovery, enrichment, and transformationAmazon Athena for querying and analyzing the results in Amazon S3 using standard SQL Amazon QuickSight for reporting and getting insights
B. Amazon Kinesis Data Analytics for data ingestionAmazon EMR for data discovery, enrichment, and transformation Amazon Redshift for querying and analyzing the results in Amazon S3
C. AWS Glue for data discovery, enrichment, and transformationAmazon Athena for querying and analyzing the results in Amazon S3 using standard SQL Amazon QuickSight for reporting and getting insights
D. AWS Data Pipeline for data transferAWS Step Functions for orchestrating AWS Lambda jobs for data discovery, enrichment, and transformationAmazon Athena for querying and analyzing the results in Amazon S3 using standard SQL Amazon QuickSight for reporting and getting insights

Answer: A

NEW QUESTION 17
A financial services company wants to adopt Amazon SageMaker as its default data science environment. The company's data scientists run machine learning (ML) models on confidential financial data. The company is worried about data egress and wants an ML engineer to secure the environment.
Which mechanisms can the ML engineer use to control data egress from SageMaker? (Choose three.)

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

Answer: BDE

Explanation:
https://aws.amazon.com/blogs/machine-learning/millennium-management-secure-machine-learning-using-amaz

NEW QUESTION 22
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 24
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 25
A trucking company is collecting live image data from its fleet of trucks across the globe. The data is growing rapidly and approximately 100 GB of new data is generated every day. The company wants to explore machine learning uses cases while ensuring the data is only accessible to specific IAM users.
Which storage option provides the most processing flexibility and will allow access control with IAM?

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

Answer: C

NEW QUESTION 30
A Machine Learning Specialist is building a convolutional neural network (CNN) that will classify 10 types of animals. The Specialist has built a series of layers in a neural network that will take an input image of an animal, pass it through a series of convolutional and pooling layers, and then finally pass it through a dense and fully connected layer with 10 nodes The Specialist would like to get an output from the neural network that is a probability distribution of how likely it is that the input image belongs to each of the 10 classes
Which function will produce the desired output?

A. Dropout
B. Smooth L1 loss

C. Softmax
D. Rectified linear units (ReLU)

Answer: C

NEW QUESTION 32
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 35
A company that promotes healthy sleep patterns by providing cloud-connected devices currently hosts a sleep tracking application on AWS. The application collects device usage information from device users. The company's Data Science team is building a machine learning model to predict if and when a user will stop utilizing the company's devices. Predictions from this model are used by a downstream application that determines the best approach for contacting users.
The Data Science team is building multiple versions of the machine learning model to evaluate each version against the company's business goals. To measure long-term effectiveness, the team wants to run multiple versions of the model in parallel for long periods of time, with the ability to control the portion of inferences served by the models.
Which solution satisfies these requirements with MINIMAL effort?

A. Build and host multiple models in Amazon SageMake
B. Create multiple Amazon SageMaker endpoints, one for each mode
C. Programmatically control invoking different models for inference at the applicationlayer.
D. Build and host multiple models in Amazon SageMake
E. Create an Amazon SageMaker endpoint configuration with multiple production variant
F. Programmatically control the portion of the inferences served by the multiple models by updating the endpoint configuration.
G. Build and host multiple models in Amazon SageMaker Neo to take into account different types of medical device
H. Programmatically control which model is invoked for inference based on the medical device type.
I. Build and host multiple models in Amazon SageMake
J. Create a single endpoint that accesses multiple model
K. Use Amazon SageMaker batch transform to control invoking the different models through the single endpoint.

Answer: B

Explanation:
A/B testing with Amazon SageMaker is required in the Exam. In A/B testing, you test different variants of your models and compare how each variant performs. Amazon SageMaker enables you to test multiple models or model versions behind the `same endpoint` using `production variants`. Each production variant identifies a machine learning (ML) model and the resources deployed for hosting the model. To test multiple models by `distributing traffic` between them, specify the `percentage of the traffic` that gets routed to each model by specifying the `weight` for each `production variant` in the endpoint configuration.
https://docs.aws.amazon.com/sagemaker/latest/dg/model-ab-testing.html#model-testing-target-variant

NEW QUESTION 40
A company is running a machine learning prediction service that generates 100 TB of predictions every day A Machine Learning Specialist must generate a visualization of the daily precision-recall curve from the predictions, and forward a read-only version to the Business team.
Which solution requires the LEAST coding effort?

A. Run a daily Amazon EMR workflow to generate precision-recall data, and save the results in Amazon S3 Give the Business team read-only access to S3
B. Generate daily precision-recall data in Amazon QuickSight, and publish the results in a dashboard shared with the Business team
C. Run a daily Amazon EMR workflow to generate precision-recall data, and save the results in Amazon S3 Visualize the arrays in Amazon QuickSight, and publish them in a dashboard shared with the Business team
D. Generate daily precision-recall data in Amazon ES, and publish the results in a dashboard shared with the Business team.

Answer: C

NEW QUESTION 44
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 47
A Machine Learning Specialist needs to create a data repository to hold a large amount of time-based training data for a new model. In the source system, new files are added every hour Throughout a single 24-hour period, the volume of hourly updates will change significantly. The Specialist always wants to train on the last 24 hours of the data
Which type of data repository is the MOST cost-effective solution?

A. An Amazon EBS-backed Amazon EC2 instance with hourly directories
B. An Amazon RDS database with hourly table partitions
C. An Amazon S3 data lake with hourly object prefixes
D. An Amazon EMR cluster with hourly hive partitions on Amazon EBS volumes

Answer: C

NEW QUESTION 48
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 53
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 58
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 61
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 66
A company supplies wholesale clothing to thousands of retail stores. A data scientist must create a model that predicts the daily sales volume for each item for each store. The data scientist discovers that more than half of the stores have been in business for less than 6 months. Sales data is highly consistent from week to week. Daily data from the database has been aggregated weekly, and weeks with no sales are omitted from the current dataset. Five years (100 MB) of sales data is available in Amazon S3.
Which factors will adversely impact the performance of the forecast model to be developed, and which actions should the data scientist take to mitigate them? (Choose two.)

A. Detecting seasonality for the majority of stores will be an issu
B. Request categorical data to relate new stores with similar stores that have more historical data.
C. The sales data does not have enough varianc
D. Request external sales data from other industries to improve the model's ability to generalize.
E. Sales data is aggregated by wee
F. Request daily sales data from the source database to enable building a daily model.
G. The sales data is missing zero entries for item sale

H. Request that item sales data from the source database include zero entries to enable building the model.
I. Only 100 MB of sales data is available in Amazon S3. Request 10 years of sales data, which would provide 200 MB of training data for the model.

Answer: AB

NEW QUESTION 69
A Machine Learning Specialist is working for a credit card processing company and receives an unbalanced dataset containing credit card transactions. It contains 99,000 valid transactions and 1,000 fraudulent transactions The Specialist is asked to score a model that was run against the dataset The Specialist has been advised that identifying valid transactions is equally as important as identifying fraudulent transactions
What metric is BEST suited to score the model?

A. Precision
B. Recall
C. Area Under the ROC Curve (AUC)
D. Root Mean Square Error (RMSE)

Answer: A

NEW QUESTION 72
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 73
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 78
A company is building a line-counting application for use in a quick-service restaurant. The company wants to use video cameras pointed at the line of customers at a given register to measure how many people are in line and deliver notifications to managers if the line grows too long. The restaurant locations have limited bandwidth for connections to external services and cannot accommodate multiple video streams without impacting other operations.
Which solution should a machine learning specialist implement to meet these requirements?

A. Install cameras compatible with Amazon Kinesis Video Streams to stream the data to AWS over the restaurant's existing internet connectio
B. Write an AWS Lambda function to take an image and send it to Amazon Rekognition to count the number of faces in the imag
C. Send an Amazon Simple Notification Service (Amazon SNS) notification if the line is too long.
D. Deploy AWS DeepLens cameras in the restaurant to capture vide
E. Enable Amazon Rekognition on the AWS DeepLens device, and use it to trigger a local AWS Lambda function when a person is recognize
F. Use the Lambda function to send an Amazon Simple Notification Service (Amazon SNS) notification if the line is too long.
G. Build a custom model in Amazon SageMaker to recognize the number of people in an imag
H. Install cameras compatible with Amazon Kinesis Video Streams in the restauran
I. Write an AWS Lambda function to take an imag
J. Use the SageMaker endpoint to call the model to count peopl
K. Send an Amazon Simple Notification Service (Amazon SNS) notification if the line is too long.
L. Build a custom model in Amazon SageMaker to recognize the number of people in an imag
M. Deploy AWS DeepLens cameras in the restauran
N. Deploy the model to the camera
O. Deploy an AWS Lambda function to the cameras to use the model to count people and send an Amazon Simple Notification Service (Amazon SNS) notification if the line is too long.

Answer: A

NEW QUESTION 81
A retail chain has been ingesting purchasing records from its network of 20,000 stores to Amazon S3 using Amazon Kinesis Data Firehose To support training an improved machine learning model, training records will require new but simple transformations, and some attributes will be combined The model needs lo be retrained daily
Given the large number of stores and the legacy data ingestion, which change will require the LEAST amount of development effort?

A. Require that the stores to switch to capturing their data locally on AWS Storage Gateway for loading into Amazon S3 then use AWS Glue to do the transformation
B. Deploy an Amazon EMR cluster running Apache Spark with the transformation logic, and have the cluster run each day on the accumulating records in Amazon S3, outputting new/transformed records to Amazon S3
C. Spin up a fleet of Amazon EC2 instances with the transformation logic, have them transform the data records accumulating on Amazon S3, and output the transformed records to Amazon S3.
D. Insert an Amazon Kinesis Data Analytics stream downstream of the Kinesis Data Firehouse stream that transforms raw record attributes into simple transformed values using SQL.

Answer: D

NEW QUESTION 82
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 87
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 91
A Machine Learning Specialist is using an Amazon SageMaker notebook instance in a private subnet of a corporate VPC. The ML Specialist has important data stored on the Amazon SageMaker notebook instance's Amazon EBS volume, and needs to take a snapshot of that EBS volume. However the ML Specialist

cannot find the Amazon SageMaker notebook instance's EBS volume or Amazon EC2 instance within the VPC.
Why is the ML Specialist not seeing the instance visible in the VPC?

A. Amazon SageMaker notebook instances are based on the EC2 instances within the customer account, but they run outside of VPCs.
B. Amazon SageMaker notebook instances are based on the Amazon ECS service within customer accounts.
C. Amazon SageMaker notebook instances are based on EC2 instances running within AWS serviceaccounts.
D. Amazon SageMaker notebook instances are based on AWS ECS instances running within AWS service accounts.

Answer: C

NEW QUESTION 94
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.

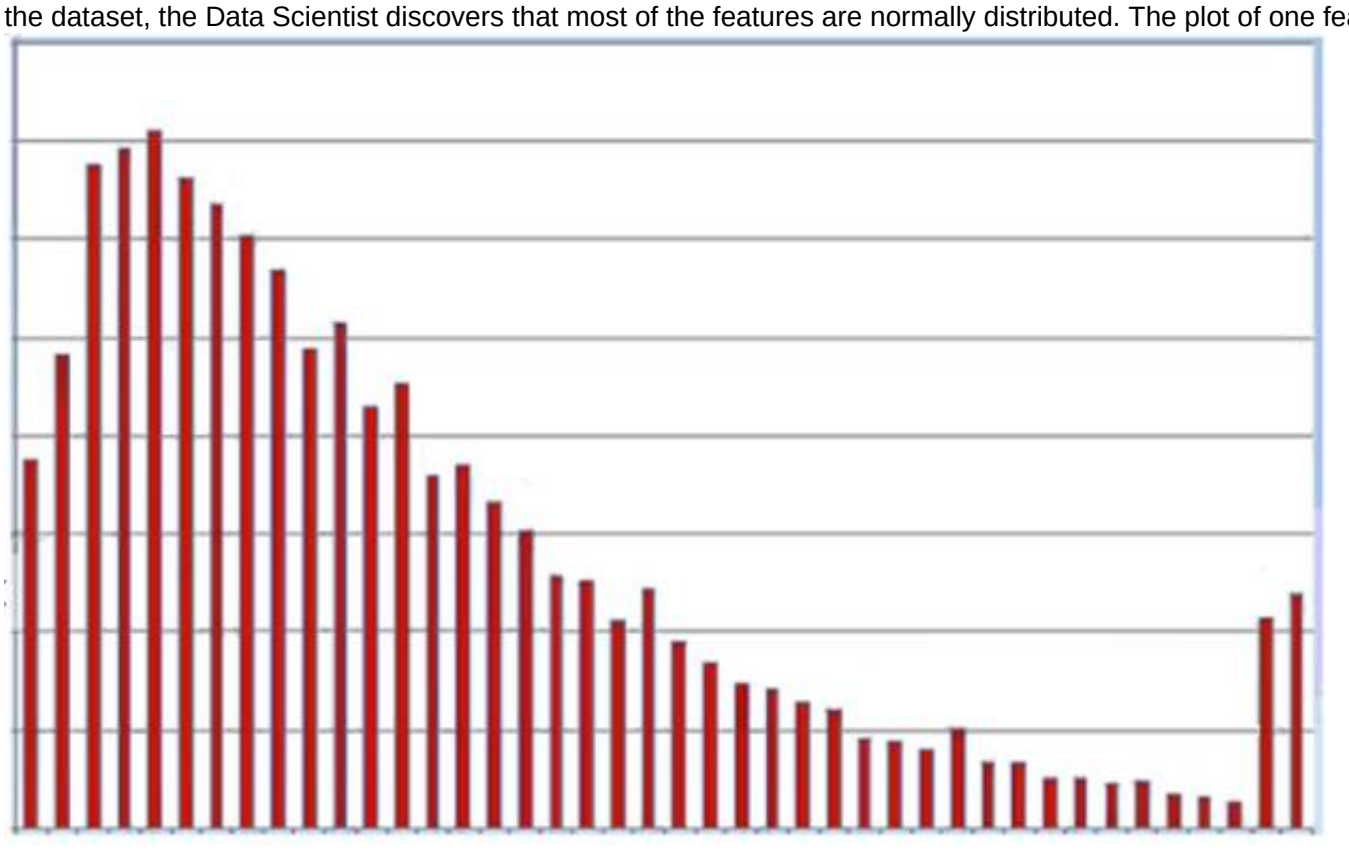
Answer: D

NEW QUESTION 99
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 100
A Data Scientist is building a linear regression model and will use resulting p-values to evaluate the statistical significance of each coefficient. Upon inspection of the dataset, the Data Scientist discovers that most of the features are normally distributed. The plot of one feature in the dataset is shown in the graphic.



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

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

Answer: A

NEW QUESTION 104
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 105
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 107
A Machine Learning Specialist works for a credit card processing company and needs to predict which transactions may be fraudulent in near-real time.
Specifically, the Specialist must train a model that returns the probability that a given transaction may fraudulent.
How should the Specialist frame this business problem?

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

Answer: C

NEW QUESTION 109
A health care company is planning to use neural networks to classify their X-ray images into normal and abnormal classes. The labeled data is divided into a training set of 1,000 images and a test set of 200 images. The initial training of a neural network model with 50 hidden layers yielded 99% accuracy on the training set, but only 55% accuracy on the test set.
What changes should the Specialist consider to solve this issue? (Choose three.)

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

Answer: ADE

NEW QUESTION 114
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 119
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 122
A bank's Machine Learning team is developing an approach for credit card fraud detection The company has a large dataset of historical data labeled as fraudulent The goal is to build a model to take the information from new transactions and predict whether each transaction is fraudulent or not
Which built-in Amazon SageMaker machine learning algorithm should be used for modeling this problem?

A. Seq2seq
B. XGBoost
C. K-means
D. Random Cut Forest (RCF)

Answer: C

NEW QUESTION 123
A machine learning specialist works for a fruit processing company and needs to build a system that categorizes apples into three types. The specialist has collected a dataset that contains 150 images for each type of apple and applied transfer learning on a neural network that was pretrained on ImageNet with this dataset.
The company requires at least 85% accuracy to make use of the model.
After an exhaustive grid search, the optimal hyperparameters produced the following: 68% accuracy on the training set 67% accuracy on the validation set
What can the machine learning specialist do to improve the system’s accuracy?

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

Answer: B

NEW QUESTION 126
Which of the following metrics should a Machine Learning Specialist generally use to compare/evaluate machine learning classification models against each other?

A. Recall
B. Misclassification rate
C. Mean absolute percentage error (MAPE)
D. Area Under the ROC Curve (AUC)

Answer: D

NEW QUESTION 129
A machine learning specialist stores IoT soil sensor data in Amazon DynamoDB table and stores weather event data as JSON files in Amazon S3. The dataset in DynamoDB is 10 GB in size and the dataset in Amazon S3 is 5 GB in size. The specialist wants to train a model on this data to help predict soil moisture levels as a function of weather events using Amazon SageMaker.
Which solution will accomplish the necessary transformation to train the Amazon SageMaker model with the LEAST amount of administrative overhead?

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

Answer: C

NEW QUESTION 131
A company has video feeds and images of a subway train station. The company wants to create a deep learning model that will alert the station manager if any passenger crosses the yellow safety line when there is no train in the station. The alert will be based on the video feeds. The company wants the model to detect the yellow line, the passengers who cross the yellow line, and the trains in the video feeds. This task requires labeling. The video data must remain confidential.
A data scientist creates a bounding box to label the sample data and uses an object detection model. However, the object detection model cannot clearly demarcate the yellow line, the passengers who cross the yellow line, and the trains.
Which labeling approach will help the company improve this model?

A. Use Amazon Rekognition Custom Labels to label the dataset and create a custom Amazon Rekognition object detection mode
B. Create a private workforc
C. Use Amazon Augmented AI (Amazon A2I) to review the low-confidence predictions and retrain the custom Amazon Rekognition model.
D. Use an Amazon SageMaker Ground Truth object detection labeling tas
E. Use Amazon Mechanical Turk as the labeling workforce.
F. Use Amazon Rekognition Custom Labels to label the dataset and create a custom Amazon Rekognition object detection mode
G. Create a workforce with a third-party AWS Marketplace vendo
H. Use Amazon Augmented AI (Amazon A2I) to review the low-confidence predictions and retrain the custom Amazon Rekognition model.
I. Use an Amazon SageMaker Ground Truth semantic segmentation labeling tas
J. Use a private workforce as the labeling workforce.

Answer: B

NEW QUESTION 133
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 138
An insurance company is developing a new device for vehicles that uses a camera to observe drivers' behavior and alert them when they appear distracted The company created approximately 10,000 training images in a controlled environment that a Machine Learning Specialist will use to train and evaluate machine learning models
During the model evaluation the Specialist notices that the training error rate diminishes faster as the number of epochs increases and the model is not accurately inferring on the unseen test images
Which of the following should be used to resolve this issue? (Select TWO)

A. Add vanishing gradient to the model
B. Perform data augmentation on the training data
C. Make the neural network architecture complex.
D. Use gradient checking in the model
E. Add L2 regularization to the model

Answer: BD

NEW QUESTION 142
During mini-batch training of a neural network for a classification problem, a Data Scientist notices that training accuracy oscillates What is the MOST likely cause of this issue?

A. The class distribution in the dataset is imbalanced
B. Dataset shuffling is disabled
C. The batch size is too big
D. The learning rate is very high

Answer: B

NEW QUESTION 144
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 146
A Machine Learning Specialist must build out a process to query a dataset on Amazon S3 using Amazon Athena The dataset contains more than 800.000 records stored as plaintext CSV files Each record contains 200 columns and is approximately 1 5 MB in size Most queries will span 5 to 10 columns only
How should the Machine Learning Specialist transform the dataset to minimize query runtime?

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

Answer: A

Explanation:
Using compressions will reduce the amount of data scanned by Amazon Athena, and also reduce your S3 bucket storage. It’s a Win-Win for your AWS bill.
Supported formats: GZIP, LZO, SNAPPY (Parquet) and ZLIB.

NEW QUESTION 148
A Machine Learning Specialist is designing a scalable data storage solution for Amazon SageMaker. There is an existing TensorFlow-based model implemented as a train.py script that relies on static training data that is currently stored as TFRecords.
Which method of providing training data to Amazon SageMaker would meet the business requirements with the LEAST development overhead?

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

Answer: B

Explanation:
https://github.com/aws-samples/amazon-sagemaker-script-mode/blob/master/tf-horovod-inference-pipeline/train

NEW QUESTION 150
A company is using Amazon Textract to extract textual data from thousands of scanned text-heavy legal documents daily. The company uses this information to process loan applications automatically. Some of the documents fail business validation and are returned to human reviewers, who investigate the errors. This activity increases the time to process the loan applications.
What should the company do to reduce the processing time of loan applications?

A. Configure Amazon Textract to route low-confidence predictions to Amazon SageMaker Ground Truth.Perform a manual review on those words before performing a business validation.
B. Use an Amazon Textract synchronous operation instead of an asynchronous operation.
C. Configure Amazon Textract to route low-confidence predictions to Amazon Augmented AI (AmazonA2I). Perform a manual review on those words before performing a business validation.
D. Use Amazon Rekognition's feature to detect text in an image to extract the data from scanned images.Use this information to process the loan applications.

Answer: C

NEW QUESTION 155
A Data Science team is designing a dataset repository where it will store a large amount of training data commonly used in its machine learning models. As Data Scientists may create an arbitrary number of new datasets every day the solution has to scale automatically and be cost-effective. Also, it must be possible to explore the data using SQL.
Which storage scheme is MOST adapted to this scenario?

A. Store datasets as files in Amazon S3.
B. Store datasets as files in an Amazon EBS volume attached to an Amazon EC2 instance.
C. Store datasets as tables in a multi-node Amazon Redshift cluster.
D. Store datasets as global tables in Amazon DynamoDB.

Answer: A

NEW QUESTION 157
A Machine Learning Specialist is assigned to a Fraud Detection team and must tune an XGBoost model, which is working appropriately for test data. However, with unknown data, it is not working as expected. The existing parameters are provided as follows.

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

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

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

Answer: B

NEW QUESTION 158
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 160
A Machine Learning Specialist is planning to create a long-running Amazon EMR cluster. The EMR cluster will have 1 master node, 10 core nodes, and 20 task nodes. To save on costs, the Specialist will use Spot Instances in the EMR cluster.
Which nodes should the Specialist launch on Spot Instances?

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

Answer: A

NEW QUESTION 162
.....

THANKS FOR TRYING THE DEMO OF OUR PRODUCT

Visit Our Site to Purchase the Full Set of Actual AWS-Certified-Machine-Learning-Specialty Exam Questions With Answers.

We Also Provide Practice Exam Software That Simulates Real Exam Environment And Has Many Self-Assessment Features. Order the AWS-Certified-Machine-Learning-Specialty Product From:

https://www.2passeasy.com/dumps/AWS-Certified-Machine-Learning-Specialty/

Money Back Guarantee

AWS-Certified-Machine-Learning-Specialty Practice Exam Features:

* AWS-Certified-Machine-Learning-Specialty Questions and Answers Updated Frequently

* AWS-Certified-Machine-Learning-Specialty Practice Questions Verified by Expert Senior Certified Staff

* AWS-Certified-Machine-Learning-Specialty Most Realistic Questions that Guarantee you a Pass on Your FirstTry

* AWS-Certified-Machine-Learning-Specialty Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year