

Amazon

Exam Questions AWS-Certified-Machine-Learning-Specialty

AWS Certified Machine Learning - Specialty



NEW QUESTION 1

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

The company is using optical character recognition (OCR) to extract text for data labeling. However, documents are in different structures and formats, and the company is facing challenges with setting up the manual workflows for each document type. Additionally, the company trained a named entity recognition (NER) model for custom entity detection using a small sample size. This model has a very low confidence score and will require retraining with a large dataset. Which solution for text extraction and entity detection will require the LEAST amount of effort?

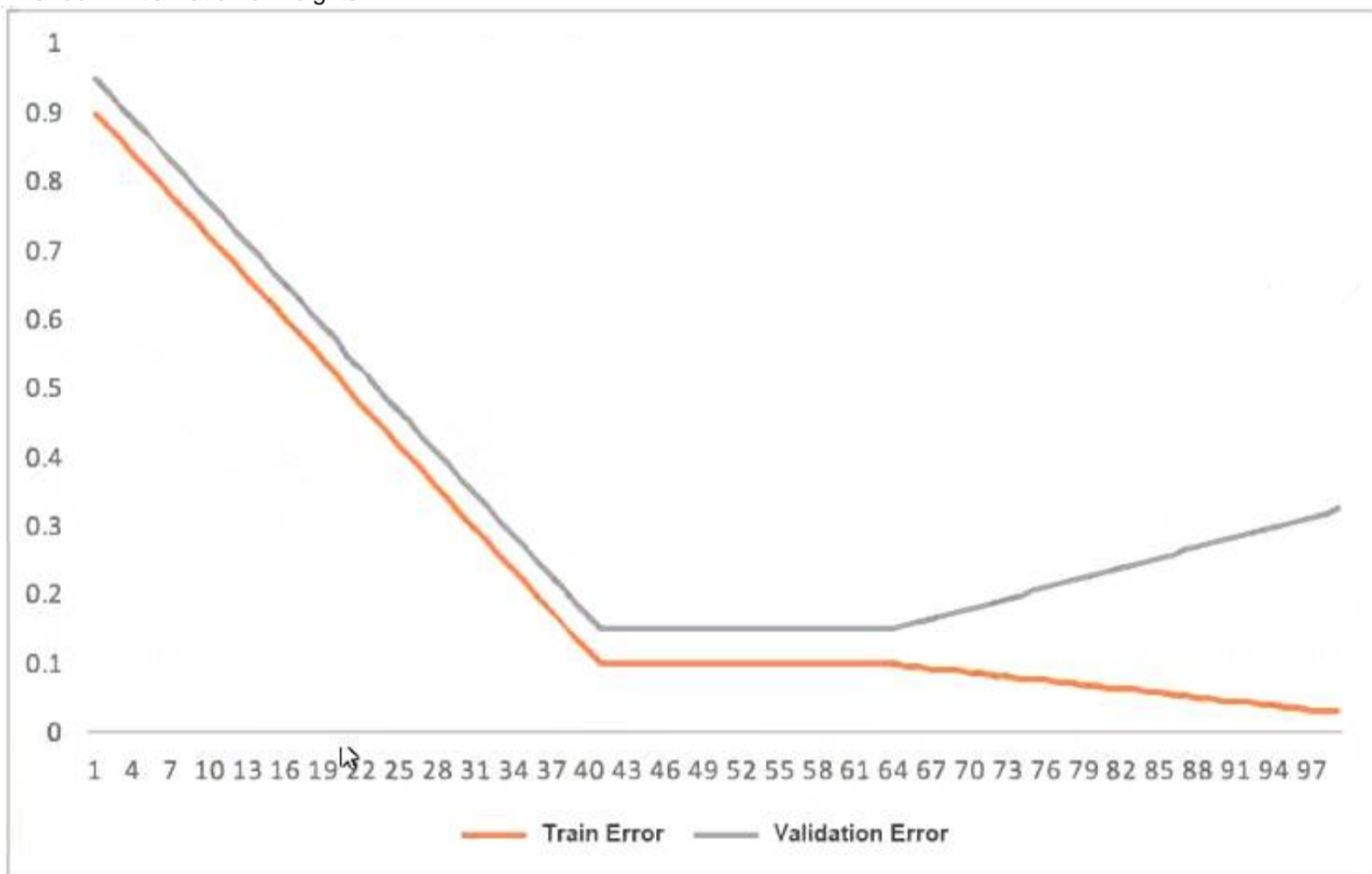
- A. Extract text from receipt images by using Amazon Textract
- B. Use the Amazon SageMaker BlazingText algorithm to train on the text for entities and custom entities.
- C. Extract text from receipt images by using a deep learning OCR model from the AWS Marketplace
- D. Use the NER deep learning model to extract entities.
- E. Extract text from receipt images by using Amazon Textract
- F. Use Amazon Comprehend for entity detection, and use Amazon Comprehend custom entity recognition for custom entity detection.
- G. Extract text from receipt images by using a deep learning OCR model from the AWS Marketplace
- H. Use Amazon Comprehend for entity detection, and use Amazon Comprehend custom entity recognition for custom entity detection.

Answer: C

NEW QUESTION 2

This graph shows the training and validation loss against the epochs for a neural network. The network being trained is as follows:

- Two dense layers one output neuron
- 100 neurons in each layer
- 100 epochs
- Random initialization of weights



Which technique can be used to improve model performance in terms of accuracy in the validation set?

- A. Early stopping
- B. Random initialization of weights with appropriate seed
- C. Increasing the number of epochs
- D. Adding another layer with the 100 neurons

Answer: D

NEW QUESTION 3

A company is building a new version of a recommendation engine. Machine learning (ML) specialists need to keep adding new data from users to improve personalized recommendations. The ML specialists gather data from the users' interactions on the platform and from sources such as external websites and social media.

The pipeline cleans, transforms, enriches, and compresses terabytes of data daily, and this data is stored in Amazon S3. A set of Python scripts was coded to do the job and is stored in a large Amazon EC2 instance. The whole process takes more than 20 hours to finish, with each script taking at least an hour. The company wants to move the scripts out of Amazon EC2 into a more managed solution that will eliminate the need to maintain servers.

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

- A. Load the data into an Amazon Redshift cluster
- B. Execute the pipeline by using SQS
- C. Store the results in Amazon S3.
- D. Load the data into Amazon DynamoDB
- E. Convert the scripts to an AWS Lambda function
- F. Execute the pipeline by triggering Lambda execution
- G. Store the results in Amazon S3.

- H. Create an AWS Glue job
- I. Convert the scripts to PySpark
- J. Execute the pipeline
- K. Store the results in Amazon S3.
- L. Create a set of individual AWS Lambda functions to execute each of the script
- M. Build a step function by using the AWS Step Functions Data Science SD
- N. Store the results in Amazon S3.

Answer: B

NEW QUESTION 4

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

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

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

Which solution meets these requirements?

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

Answer: C

NEW QUESTION 5

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

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

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

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

Answer: C

NEW QUESTION 6

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

How should the data scientist meet these requirements?

- A. Mount the EFS file system to a SageMaker notebook and run a script that copies the data to an Amazon FSx for Lustre file system
- B. Run the SageMaker training job with the FSx for Lustre file system as the data source.
- C. Launch a transient Amazon EMR cluster
- 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 bucket
- 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 7

An office security agency conducted a successful pilot using 100 cameras installed at key locations within the main office. Images from the cameras were uploaded to Amazon S3 and tagged using Amazon Rekognition, and the results were stored in Amazon ES. The agency is now looking to expand the pilot into a full production system using thousands of video cameras in its office locations globally. The goal is to identify activities performed by non-employees in real time.

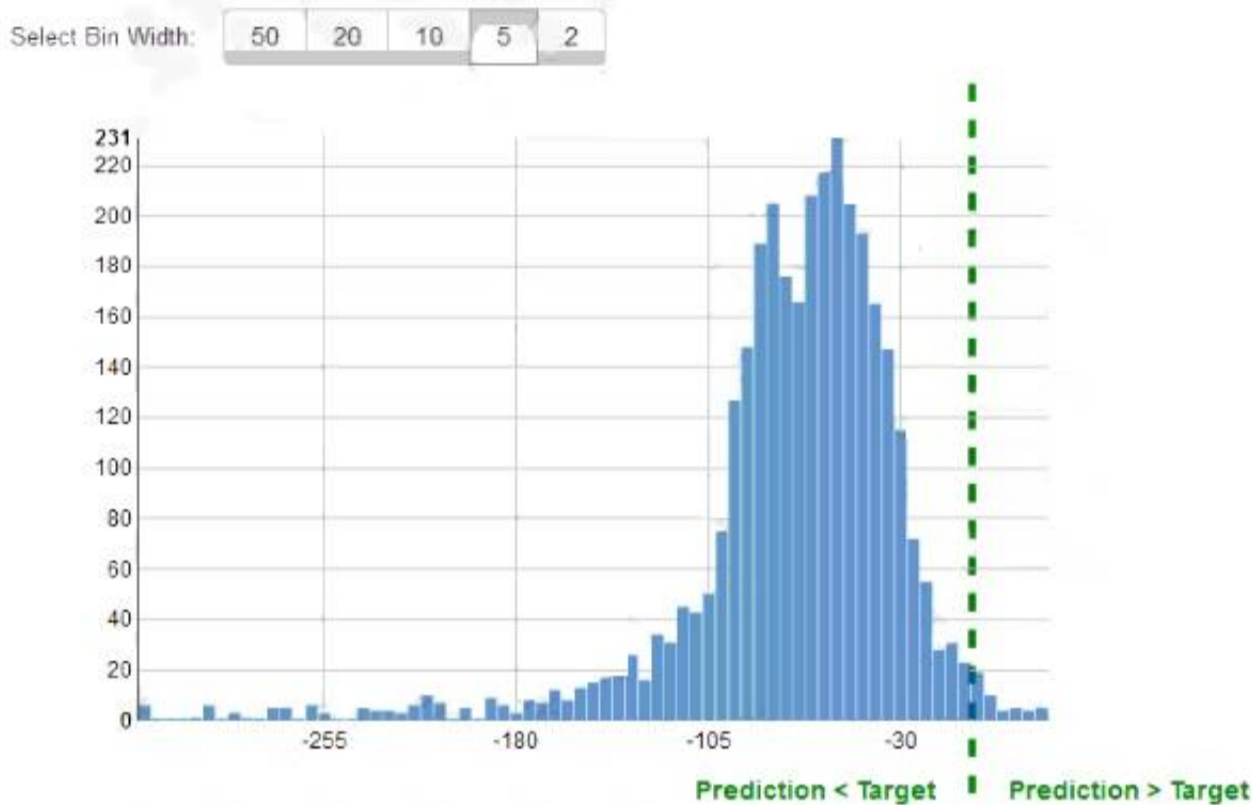
Which solution should the agency consider?

- A. Use a proxy server at each local office and for each camera, and stream the RTSP feed to a unique Amazon Kinesis Video Streams video stream.
- B. On each stream, use Amazon Rekognition Video and create a stream processor to detect faces from a collection of known employees, and alert when non-employees are detected.
- C. Use a proxy server at each local office and for each camera, and stream the RTSP feed to a unique Amazon Kinesis Video Streams video stream.
- D. On each stream, use Amazon Rekognition Image to detect faces 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 camera.
- 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 non-employees are detected.
- G. Install AWS DeepLens cameras and use the DeepLens_Kinesis_Video module to stream video to Amazon Kinesis Video Streams for each camera.
- 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 8

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 9

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 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 Machine Learning Specialist is creating a new natural language processing application that processes a dataset comprised of 1 million sentences. The aim is to

then run Word2Vec to generate embeddings of the sentences and enable different types of predictions

Here is an example from the dataset

"The quck BROWN FOX jumps over the lazy dog "

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

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

Answer: BCF

NEW QUESTION 11

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 14

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 CloudWate
- 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 CloudWate
- 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 CloudWate
- 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 15

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 17

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 21

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 22

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%

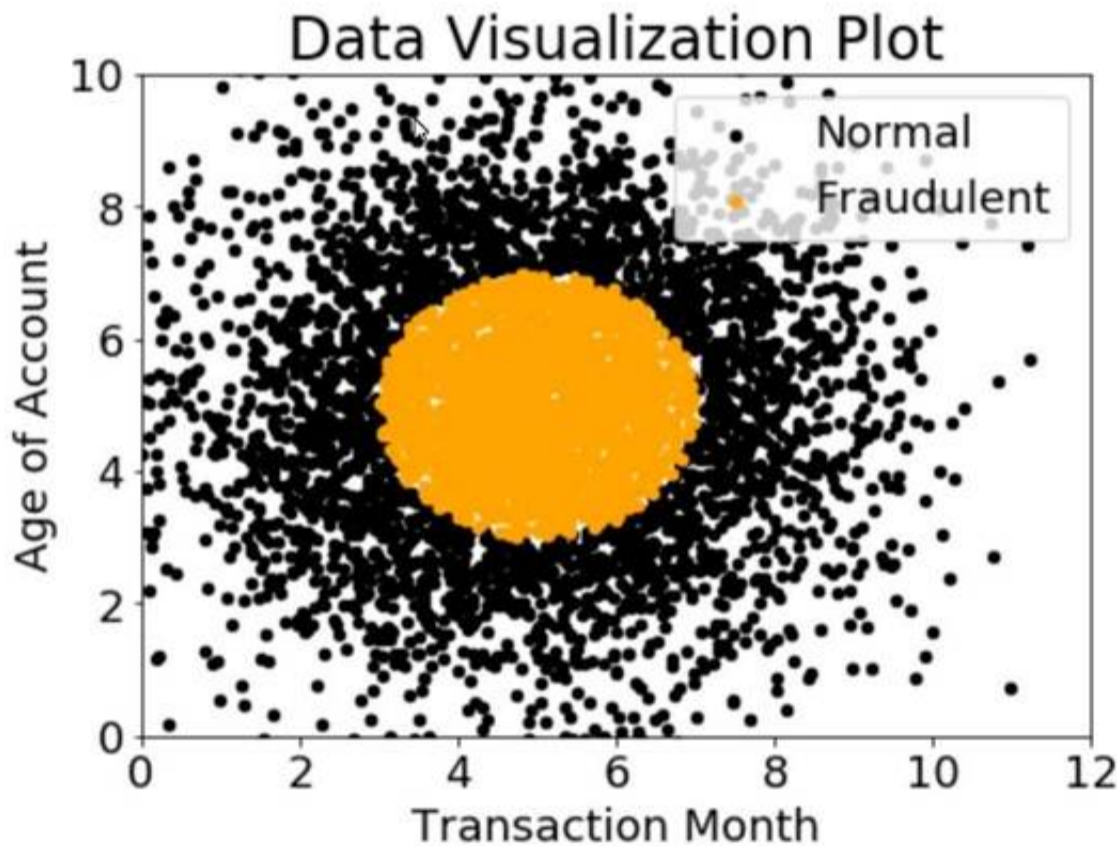
What can the Specialist from the experiments?

- A. The model In Experiment 1 had a high variance error that 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 that 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 Experiment 3 by regularization Experiment 2 shows that random noise cannot be reduced by increasing layers and neurons in the model

Answer: C

NEW QUESTION 27

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 32

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

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

- A. AWS Secrets Manager
- B. AWS CodeStar
- C. Amazon ECR
- D. Amazon ECS
- E. Amazon S3

Answer: CE

NEW QUESTION 36

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

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

- A. Use Amazon Comprehend for the part-of-speech tagging, key phrase extraction, and classification tasks.
- B. Use an NLP library in Amazon SageMaker for the part-of-speech tagging
- C. Use Amazon Comprehend for the key phrase extraction
- 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 phrase 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 phrase extraction task
- H. Use AWS Deep Learning Containers with Amazon SageMaker to build the custom classifier.

Answer: B

NEW QUESTION 40

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

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

Which approach will maximize transcription accuracy during the development phase?

- A. Use a voice-driven Amazon Lex bot to perform the ASR customization
- B. Create customer slots within the bot that specifically identify each of the required product names
- 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 customization
- 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 value
- 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 customization
- 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 model
- J. Analyze the transcripts and update the training dataset with a manually corrected version of transcripts where product names are not being transcribed correctly
- K. Create an updated custom language model.

Answer: A

NEW QUESTION 45

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

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

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

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

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

Answer: A

NEW QUESTION 48

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 to ask questions using written and spoken interfaces.

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 Polly
- E. Amazon Comprehend
- F. Amazon Transcribe

Answer: BEF

NEW QUESTION 50

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 51

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 use cases while ensuring the data is only accessible to specific IAM users. Which storage option provides the most processing flexibility and will allow access control with IAM?

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

Answer: C

NEW QUESTION 54

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

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

Answer: D

NEW QUESTION 56

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: D

NEW QUESTION 61

Amazon Connect has recently been tolled out across a company as a contact call center. The solution has been configured to store voice call recordings on Amazon S3.

The content of the voice calls are being analyzed for the incidents being discussed by the call operators. Amazon Transcribe is being used to convert the audio to text, and the output is stored on Amazon S3.

Which approach will provide the information required for further analysis?

- A. Use Amazon Comprehend with the transcribed files to build the key topics
- B. Use Amazon Translate with the transcribed files to train and build a model for the key topics
- C. Use the AWS Deep Learning AMI with Gluon Semantic Segmentation on the transcribed files to train and build a model for the key topics
- D. Use the Amazon SageMaker k-Nearest-Neighbors (kNN) algorithm on the transcribed files to generate a word embeddings dictionary for the key topics

Answer: B

NEW QUESTION 66

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 69

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

Property_ID	Wall_Color
1000	Red
1001	White
1002	Green

The specialist chose a model that needs numerical input data.

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

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

Answer: AD

NEW QUESTION 71

A library is developing an automatic book-borrowing system that uses Amazon Rekognition. Images of library members' faces are stored in an Amazon S3 bucket. When members borrow books, the Amazon Rekognition CompareFaces API operation compares real faces against the stored faces in Amazon S3.

The library needs to improve security by making sure that images are encrypted at rest. Also, when the images are used with Amazon Rekognition, they need to be encrypted in transit. The library also must ensure that the images are not used to improve Amazon Rekognition as a service.

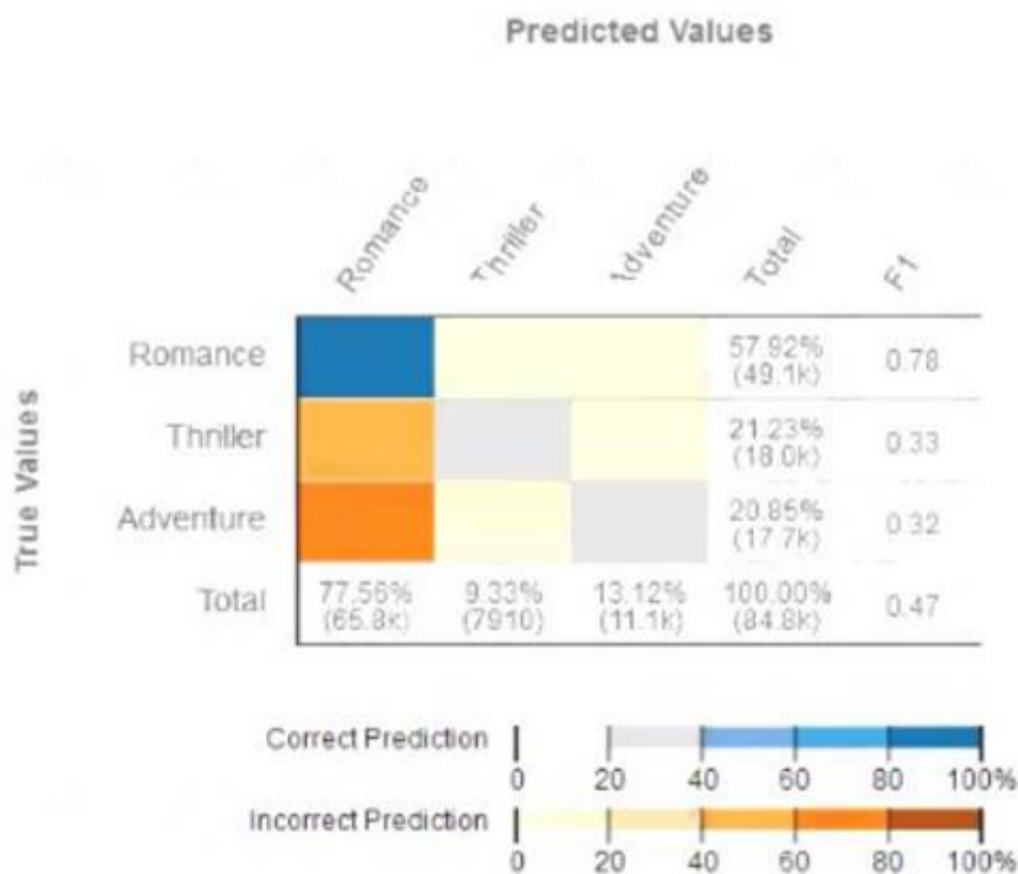
How should a machine learning specialist architect the solution to satisfy these requirements?

- A. Enable server-side encryption on the S3 bucket
- 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 and SearchFacesByImage 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 bucket
- H. Set up a VPN connection and only call the Amazon Rekognition API operations through the VPN.

Answer: B

NEW QUESTION 73

Given the following confusion matrix for a movie classification model, what is the true class frequency for Romance and the predicted class frequency for Adventure?



- A. The true class frequency for Romance is 77.56% and the predicted class frequency for Adventure is 20.85%
- B. The true class frequency for Romance is 57.92% and the predicted class frequency for Adventure is 13.12%
- C. The true class frequency for Romance is 0.78 and the predicted class frequency for Adventure is (0.47 - 0.32).
- D. The true class frequency for Romance is 77.56% * 0.78 and the predicted class frequency for Adventure is 20.85% + 0.32

Answer: B

Explanation:

<https://docs.aws.amazon.com/machine-learning/latest/dg/multiclass-model-insights.html>

NEW QUESTION 78

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 79

A Machine Learning Specialist uploads a dataset to an Amazon S3 bucket protected with server-side encryption using AWS KMS.

How should the ML Specialist define the Amazon SageMaker notebook instance so it can read the same dataset from Amazon S3?

- A. Define security group(s) to allow all HTTP inbound/outbound traffic and assign those security group(s) to the Amazon SageMaker notebook instance.
- B. Configure the Amazon SageMaker notebook instance to have access to the 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 dataset
- 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 84

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

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

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

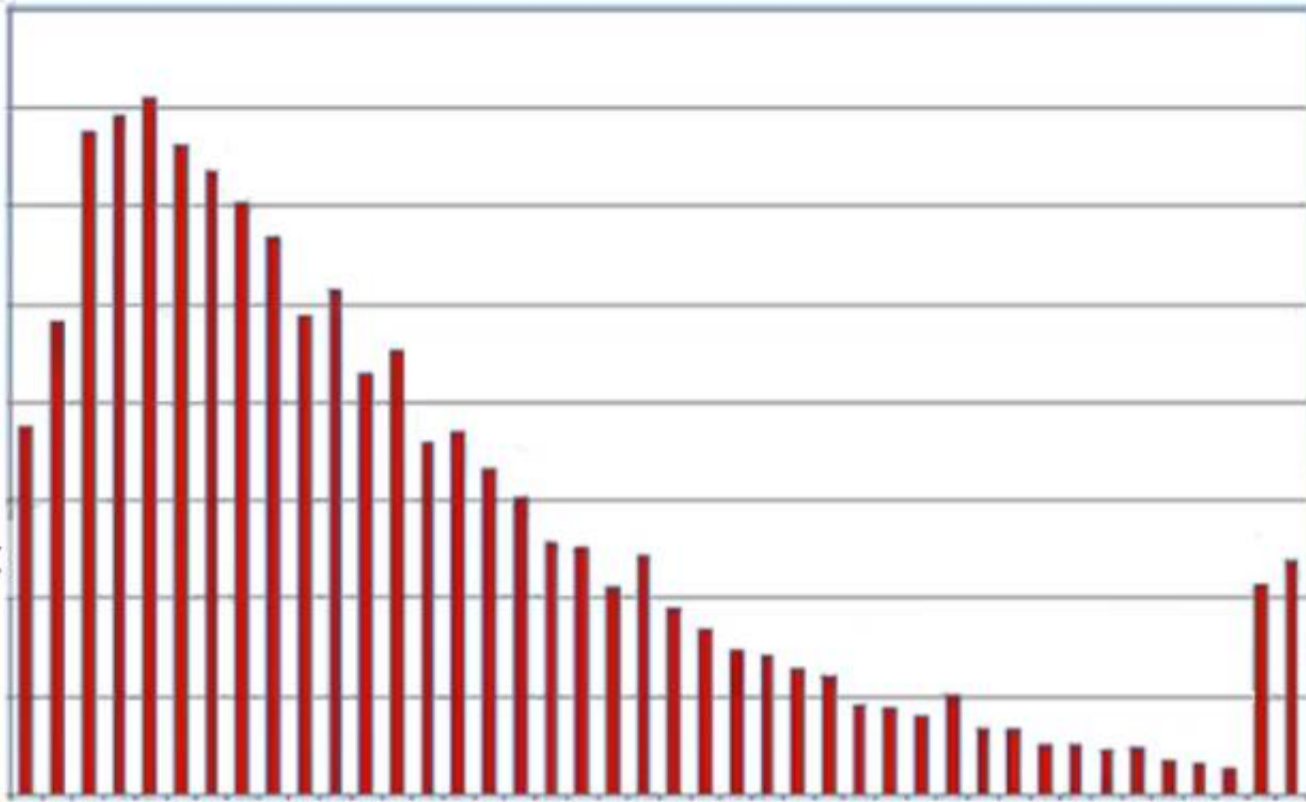
How can the company resolve this issue MOST cost-effectively?

- A. Set up a 10 Gbps AWS Direct Connect connection between the production site and the nearest AWS Region
- 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 SageMaker
- F. Set up an AWS Direct Connect connection between the production site and the nearest AWS Region
- 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 machine
- 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 86

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 87

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 91

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 93

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 94

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

What should the ML specialist do to resolve the violations?

- A. Manually trigger the monitoring job to re-evaluate the SageMaker endpoint traffic sample.
- B. Run the Model Monitor baseline job again on the new training set.
- 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 98

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

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

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

- A. Update the ProductionVariant data type with the new version of the model by using the CreateEndpointConfig operation with the InitialVariantWeight parameter set to 0. Specify the TargetVariant parameter for InvokeEndpoint calls for users who subscribed to the preview feature.
- 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 model.
- D. Create an Application Load Balancer (ALB) to route traffic to both endpoints based on the TargetVariant query string parameter.
- E. Reconfigure the app to send the TargetVariant query string parameter for users who subscribed to the preview feature.
- 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 feature.
- 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 model.
- J. Create an Amazon Route 53 record that is configured with a simple routing policy and that points to the current version of the model.
- 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 users.
- 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 99

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 102

A company has raw user and transaction data stored in Amazon S3, a MySQL database, and Amazon RedShift. A Data Scientist needs to perform an analysis by joining the three datasets from Amazon S3, MySQL, and Amazon RedShift, and then calculating the average of a few selected columns from the joined data. Which AWS service should the Data Scientist use?

- A. Amazon Athena
- B. Amazon Redshift Spectrum
- C. AWS Glue
- D. Amazon QuickSight

Answer: A

NEW QUESTION 106

A Machine Learning Specialist is using Amazon SageMaker to host a model for a highly available customer-facing application.

The Specialist has trained a new version of the model, validated it with historical data, and now wants to deploy it to production. To limit any risk of a negative customer experience, the Specialist wants to be able to monitor the model and roll it back, if needed.

What is the SIMPLEST approach with the LEAST risk to deploy the model and roll it back, if needed?

- A. Create a SageMaker endpoint and configuration for the new model version.

- B. Redirect production traffic to the new endpoint by updating the client configuratio
- C. Revert traffic to the last version if the model does not perform as expected.
- D. Create a SageMaker endpoint and configuration for the new model versio
- E. Redirect production traffic to the new endpoint by using a load balancer Revert traffic to the last version if the model does not perform as expected.
- F. Update the existing SageMaker endpoint to use a new configuration that is weighted to send 5% of the traffic to the new varian
- G. Revert traffic to the last version by resetting the weights if the model does not perform as expected.
- H. Update the existing SageMaker endpoint to use a new configuration that is weighted to send 100% of the traffic to the new variant Revert traffic to the last version by resetting the weights if the model does not perform as expected.

Answer: A

NEW QUESTION 109

A Machine Learning Specialist is working with a large cybersecurity company that manages security events in real time for companies around the world The cybersecurity company wants to design a solution that will allow it to use machine learning to score malicious events as anomalies on the data as it is being ingested The company also wants be able to save the results in its data lake for later processing and analysis What is the MOST efficient way to accomplish these tasks'?

- A. Ingest the data using Amazon Kinesis Data Firehose, and use Amazon Kinesis Data Analytics Random Cut Forest (RCF) for anomaly detection Then use Kinesis Data Firehose to stream the results to Amazon S3
- B. Ingest the data into Apache Spark Streaming using Amazon EM
- C. and use Spark MLlib with k-means to perform anomaly detection Then store the results in an Apache Hadoop Distributed File System (HDFS) using Amazon EMR with a replication factor of three as the data lake
- D. Ingest the data and store it in Amazon S3 Use AWS Batch along with the AWS Deep Learning AMIs to train a k-means model using TensorFlow on the data in Amazon S3.
- E. Ingest the data and store it in Amazon S3. Have an AWS Glue job that is triggered on demand transform the new data Then use the built-in Random Cut Forest (RCF) model within Amazon SageMaker to detect anomalies in the data

Answer: A

NEW QUESTION 113

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 114

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 115

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 118

A company's Machine Learning Specialist needs to improve the training speed of a time-series forecasting model using TensorFlow. The training is currently implemented on a single-GPU machine and takes approximately 23 hours to complete. The training needs to be run daily. The model accuracy js acceptable, but the company anticipates a continuous increase in the size of the training data and a need to update the model on an hourly, rather than a daily, basis. The company also wants to minimize coding effort and infrastructure changes What should the Machine Learning Specialist do to the training solution to allow it to scale for future demand?

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

Answer: B

NEW QUESTION 122

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

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

How should the data scientist transform the data?

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

Answer: A

Explanation:

<https://docs.aws.amazon.com/forecast/latest/dg/dataset-import-guidelines-troubleshooting.html>

NEW QUESTION 126

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

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

Which solution will meet these requirements?

- A. Move the training data to an Amazon S3 bucket
- B. Train and evaluate the model by using Amazon SageMaker
- C. Optimize the model by using SageMaker Ne
- D. Deploy the model on a SageMaker hosting services endpoint.
- E. Train and evaluate the model on premise
- F. Upload the model to an Amazon S3 bucket
- G. Deploy the model on an Amazon SageMaker hosting services endpoint.
- H. Move the training data to an Amazon S3 bucket
- I. Train and evaluate the model by using Amazon SageMaker
- J. Optimize the model by using SageMaker Ne
- K. Set up an edge device in the manufacturing facilities with AWS IoT Greengrass
- L. Deploy the model on the edge device.
- M. Train the model on premise
- N. Upload the model to an Amazon S3 bucket
- O. Set up an edge device in the manufacturing facilities with AWS IoT Greengrass
- P. Deploy the model on the edge device.

Answer: A

NEW QUESTION 127

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 use?

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

Answer: D

NEW QUESTION 132

A logistics company needs a forecast model to predict next month's inventory requirements for a single item in 10 warehouses. A machine learning specialist uses

Amazon Forecast to develop a forecast model from 3 years of monthly data. There is no missing data. The specialist selects the DeepAR+ algorithm to train a predictor. The predictor means absolute percentage error (MAPE) is much larger than the MAPE produced by the current human forecasters. Which changes to the CreatePredictor API call could improve the MAPE? (Choose two.)

- A. Set PerformAutoML to true.
- B. Set ForecastHorizon to 4.
- C. Set ForecastFrequency to W for weekly.
- D. Set PerformHPO to true.
- E. Set FeaturizationMethodName to filling.

Answer: CD

NEW QUESTION 136

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 139

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 143

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 146

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 147

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

Total number of images available = 1,000 Test set images = 100 (constant test set)

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

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

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

Answer: A

NEW QUESTION 152

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 153

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 156

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 contaminants for the next 2 days in the city. As this is a prototype, only daily data from the last year is available.

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

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

Answer: C

NEW QUESTION 160

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 163

.....

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questions and Answers in PDF Format

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

100% Actual & Verified — Instant Download, Please Click
[Order The AWS-Certified-Machine-Learning-Specialty Practice Test Here](#)