Amazon
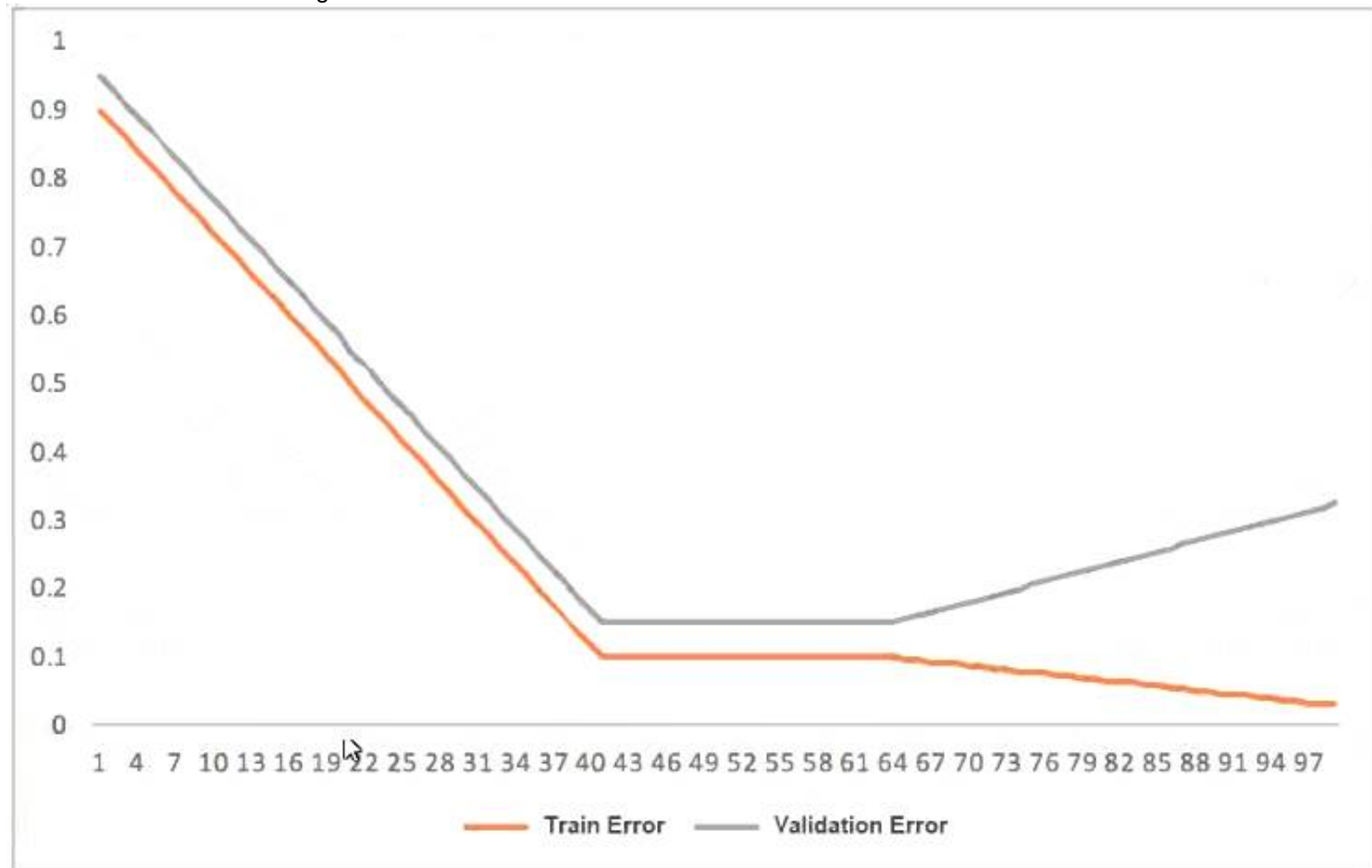
Exam Questions AWS-Certified-Machine-Learning-Specialty

AWS Certified Machine Learning - Specialty

NEW QUESTION 1
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 2
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 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 cluste
B. Execute the pipeline by using SQ
C. Store the results in Amazon S3.
D. Load the data into Amazon DynamoD
E. Convert the scripts to an AWS Lambda functio
F. Execute the pipeline by triggering Lambda execution
G. Store the results in Amazon S3.
H. Create an AWS Glue jo
I. Convert the scripts to PySpar
J. Execute the pipelin
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
An e commerce company wants to launch a new cloud-based product recommendation feature for its web application. Due to data localization regulations, any sensitive data must not leave its on-premises data center, and the product recommendation model must be trained and tested using nonsensitive data only. Data transfer to the cloud must use IPsec. The web application is hosted on premises with a PostgreSQL database that contains all the data. The company wants the data to be uploaded securely to Amazon S3 each day for model retraining.
How should a machine learning specialist meet these requirements?

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

Answer: C

NEW QUESTION 5
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_tex
B. Use the Amazon SageMaker Object2Vec algorithm and the .csv file to train a mode
C. Use SageMaker to deploy the model to an inference endpoin
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_tex
F. Use the Amazon SageMaker Latent Dirichlet Allocation (LDA) algorithm and the .csv file to train a mode
G. Use the LDA algorithm to detect labels automaticall
H. Use SageMaker to deploy the model to an inference endpoin
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_tex
K. Use Amazon Comprehend custom classification and the extracted information to train the custom classifie
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_tex
N. Use Amazon Comprehend custom classification and the .csv file to train the custom classifie
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 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 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 7
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 8
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 9
A web-based company wants to improve its conversion rate on its landing page Using a large historical dataset of customer visits, the company has repeatedly trained a multi-class deep learning network algorithm on Amazon SageMaker However there is an overfitting problem training data shows 90% accuracy in predictions, while test data shows 70% accuracy only
The company needs to boost the generalization of its model before deploying it into production to maximize conversions of visits to purchases
Which action is recommended to provide the HIGHEST accuracy model for the company's test and validation data?

A. Increase the randomization of training data in the mini-batches used in training.
B. Allocate a higher proportion of the overall data to the training dataset
C. Apply L1 or L2 regularization and dropouts to the training.
D. Reduce the number of layers and units (or neurons) from the deep learning network.

Answer: C

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

NEW QUESTION 10
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?
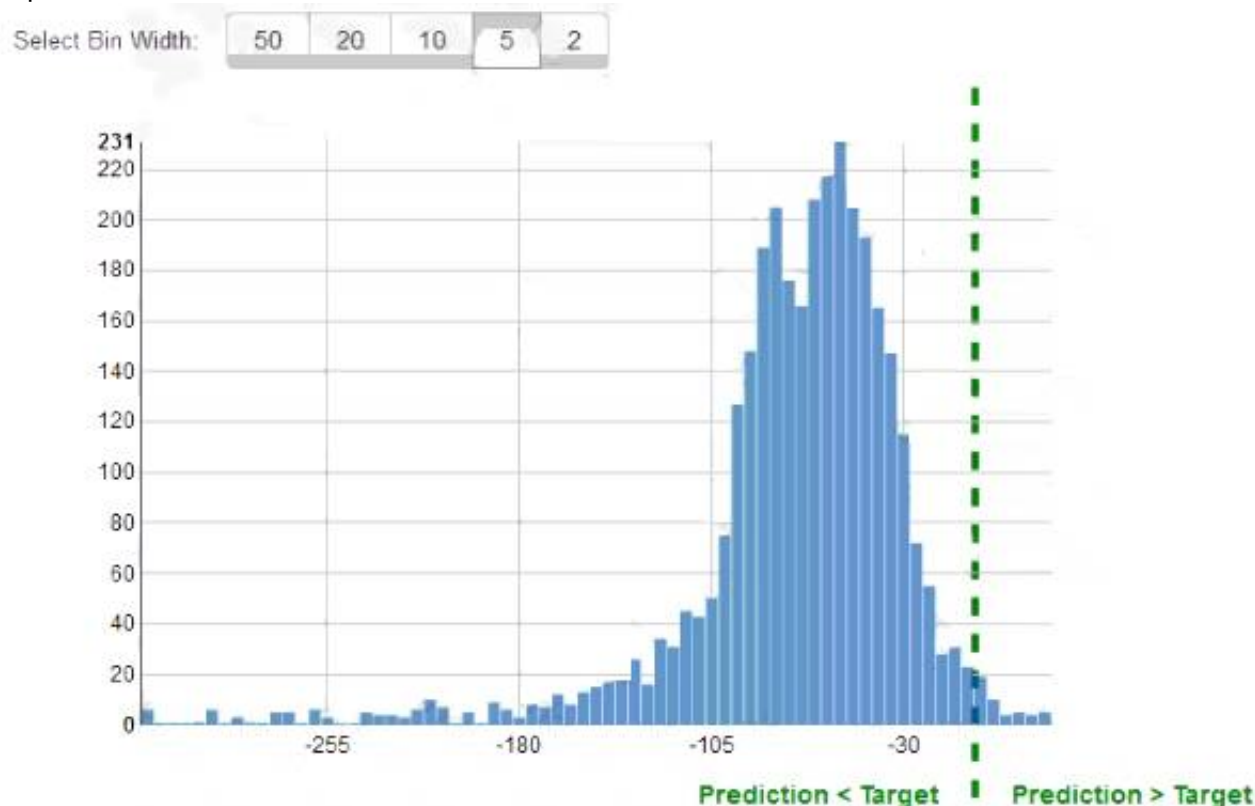
A. The model might have prediction errors over a range of target values.
B. The dataset cannot be accurately represented using the regression model

C. There are too many variables in the model
D. The model is predicting its target values perfectly.

Answer: D

NEW QUESTION 10
A machine learning (ML) specialist is using Amazon SageMaker hyperparameter optimization (HPO) to improve a model's accuracy. The learning rate parameter is specified in the following HPO configuration:

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

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

A. Modify the HPO configuration as follows: C:\Users\Admin\Desktop\Data\Odt data\Untitled.jpgSelect the most accurate hyperparameter configuration form this HPO job.

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

B. Run three different HPO jobs that use different learning rates form the following intervals for MinValue and MaxValue while using the same number of training jobs for each HPO job:[0.01, 0.1][0.001, 0.01][0.0001, 0.001]Select the most accurate hyperparameter configuration form these three HPO jobs.
C. Modify the HPO configuration as follows: C:\Users\Admin\Desktop\Data\Odt data\Untitled.jpg

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

Select the most accurate hyperparameter configuration form this training job.
D. Run three different HPO jobs that use different learning rates form the following intervals for MinValue and MaxValu
E. Divide the number of training jobs for each HPO job by three:[0.01, 0.1][0.001, 0.01][0.0001, 0.001]Select the most accurate hyperparameter configuration form these three HPO jobs.

Answer: C

NEW QUESTION 14
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 18
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 23
A company will use Amazon SageMaker to train and host a machine learning (ML) model for a marketing campaign. The majority of data is sensitive customer data. The data must be encrypted at rest. The company wants AWS to maintain the root of trust for the master keys and wants encryption key usage to be logged. Which implementation will meet these requirements?

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

Answer: C

NEW QUESTION 28
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 31
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 35
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 36
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 37
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 38
A power company wants to forecast future energy consumption for its customers in residential properties and commercial business properties. Historical power consumption data for the last 10 years is available. A team of data scientists who performed the initial data analysis and feature selection will include the historical power consumption data and data such as weather, number of individuals on the property, and public holidays.
The data scientists are using Amazon Forecast to generate the forecasts.
Which algorithm in Forecast should the data scientists use to meet these requirements?

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

Answer: B

NEW QUESTION 43
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 45
A data science team is planning to build a natural language processing (NLP) application. The application's text preprocessing stage will include part-of-speech tagging and key phase extraction. The preprocessed text will be input to a custom classification algorithm that the data science team has already written and trained using Apache MXNet.
Which solution can the team build MOST quickly to meet these requirements?

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

Answer: B

NEW QUESTION 46
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 51
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 54
A company wants to use automatic speech recognition (ASR) to transcribe messages that are less than 60 seconds long from a voicemail-style application. The company requires the correct identification of 200 unique product names, some of which have unique spellings or pronunciations.
The company has 4,000 words of Amazon SageMaker Ground Truth voicemail transcripts it can use to customize the chosen ASR model. The company needs to ensure that everyone can update their customizations multiple times each hour.
Which approach will maximize transcription accuracy during the development phase?

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

Answer: A

NEW QUESTION 58
A Machine Learning Specialist needs to move and transform data in preparation for training Some of the data needs to be processed in near-real time and other data can be moved hourly There are existing Amazon EMR MapReduce jobs to clean and feature engineering to perform on the data
Which of the following services can feed data to the MapReduce jobs? (Select TWO)

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

Answer: BC

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

NEW QUESTION 61
A Machine Learning Specialist is configuring automatic model tuning in Amazon SageMaker
When using the hyperparameter optimization feature, which of the following guidelines should be followed to improve optimization?
Choose the maximum number of hyperparameters supported by

A. Amazon SageMaker to search the largest number of combinations possible
B. Specify a very large hyperparameter range to allow Amazon SageMaker to cover every possible value.
C. Use log-scaled hyperparameters to allow the hyperparameter space to be searched as quickly as possible
D. Execute only one hyperparameter tuning job at a time and improve tuning through successive rounds of experiments

Answer: C

NEW QUESTION 63
A Data Scientist received a set of insurance records, each consisting of a record ID, the final outcome among 200 categories, and the date of the final outcome. Some partial information on claim contents is also provided, but only for a few of the 200 categories. For each outcome category, there are hundreds of records distributed over the past 3 years. The Data Scientist wants to predict how many claims to expect in each category from month to month, a few months in advance.
What type of machine learning model should be used?

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

Answer: C

NEW QUESTION 66
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 70
A Data Scientist is working on an application that performs sentiment analysis. The validation accuracy is poor and the Data Scientist thinks that the cause may be a rich vocabulary and a low average frequency of words in the dataset
Which tool should be used to improve the validation accuracy?

A. Amazon Comprehend syntax analysts and entity detection
B. Amazon SageMaker BlazingText allow mode
C. Natural Language Toolkit (NLTK) stemming and stop word removal
D. Scikit-learn term frequency-inverse document frequency (TF-IDF) vectorizers

Answer: A

NEW QUESTION 72
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 75
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 77
A real-estate company is launching a new product that predicts the prices of new houses. The historical data for the properties and prices is stored in .csv format in an Amazon S3 bucket. The data has a header, some categorical fields, and some missing values. The company's data scientists have used Python with a common open-source library to fill the missing values with zeros. The data scientists have dropped all of the categorical fields and have trained a model by using the open-source linear regression algorithm with the default parameters.
The accuracy of the predictions with the current model is below 50%. The company wants to improve the model performance and launch the new product as soon as possible.
Which solution will meet these requirements with the LEAST operational overhead?

A. Create a service-linked role for Amazon Elastic Container Service (Amazon ECS) with access to the S3 bucke
B. Create an ECS cluster that is based on an AWS Deep Learning Containers imag
C. Write the code to perform the feature engineerin
D. Train a logistic regression model for predicting the price, pointing to the bucket with the datase
E. Wait for the training job to complet
F. Perform the inferences.
G. Create an Amazon SageMaker notebook with a new IAM role that is associated with the noteboo
H. Pull the dataset from the S3 bucke
I. Explore different combinations of feature engineering transformations,regression algorithms, and hyperparameter
J. Compare all the results in the notebook, and deploy the most accurate configuration in an endpoint for predictions.
K. Create an IAM role with access to Amazon S3, Amazon SageMaker, and AWS Lambd
L. Create a training job with the SageMaker built-in XGBoost model pointing to the bucket with the datase

M. Specify the price as the target featur
N. Wait for the job to complet
O. Load the model artifact to a Lambda function for inference on prices of new houses.
P. Create an IAM role for Amazon SageMaker with access to the S3 bucke
Q. Create a SageMaker AutoML job with SageMaker Autopilot pointing to the bucket with the datase
R. Specify the price as the target attribut
S. Wait for the job to complet
T. Deploy the best model for predictions.

Answer: A

NEW QUESTION 82
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 85
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 89
A Machine Learning Specialist is deciding between building a naive Bayesian model or a full Bayesian network for a classification problem. The Specialist computes the Pearson correlation coefficients between each feature and finds that their absolute values range between 0.1 to 0.95.
Which model describes the underlying data in this situation?

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

Answer: C

NEW QUESTION 94
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 99

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 101
A manufacturing company has structured and unstructured data stored in an Amazon S3 bucket A Machine Learning Specialist wants to use SQL to run queries on this data. Which solution requires the LEAST effort to be able to query this data?

A. Use AWS Data Pipeline to transform the data and Amazon RDS to run queries.
B. Use AWS Glue to catalogue the data and Amazon Athena to run queries
C. Use AWS Batch to run ETL on the data and Amazon Aurora to run the quenes
D. Use AWS Lambda to transform the data and Amazon Kinesis Data Analytics to run queries

Answer: D

NEW QUESTION 105
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 107
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 111
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 116
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 118
A company uses a long short-term memory (LSTM) model to evaluate the risk factors of a particular energy sector. The model reviews multi-page text documents to analyze each sentence of the text and categorize it as either a potential risk or no risk. The model is not performing well, even though the Data Scientist has experimented with many different network structures and tuned the corresponding hyperparameters.
Which approach will provide the MAXIMUM performance boost?

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

Answer: C

NEW QUESTION 119
A Data Scientist wants to gain real-time insights into a data stream of GZIP files. Which solution would allow the use of SQL to query the stream with the LEAST latency?

A. Amazon Kinesis Data Analytics with an AWS Lambda function to transform the data.
B. AWS Glue with a custom ETL script to transform the data.
C. An Amazon Kinesis Client Library to transform the data and save it to an Amazon ES cluster.
D. Amazon Kinesis Data Firehose to transform the data and put it into an Amazon S3 bucket.

Answer: A

NEW QUESTION 121
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 126
A Machine Learning Specialist prepared the following graph displaying the results of k-means for k = [1:10]

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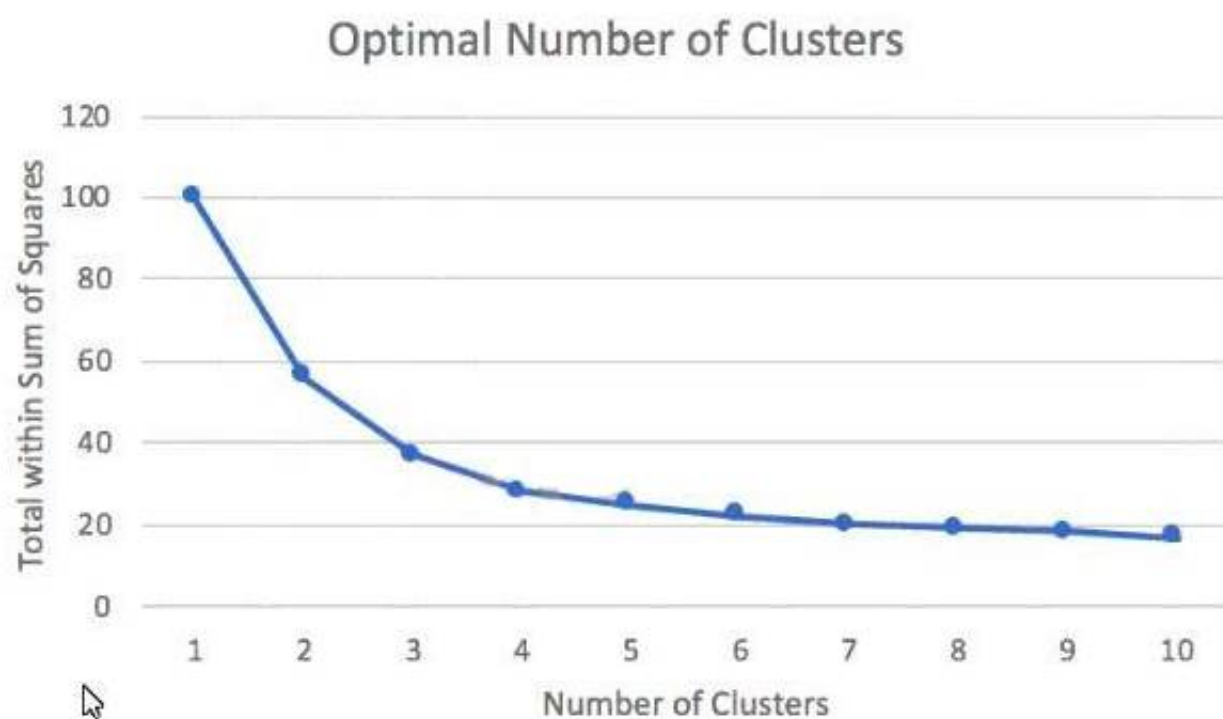
Considering the graph, what is a reasonable selection for the optimal choice of k?

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

Answer: C

NEW QUESTION 128
A technology startup is using complex deep neural networks and GPU compute to recommend the company's products to its existing customers based upon each customer's habits and interactions. The solution currently pulls each dataset from an Amazon S3 bucket before loading the data into a TensorFlow model pulled from the company's Git repository that runs locally. This job then runs for several hours while continually outputting its progress to the same S3 bucket. The job can be paused, restarted, and continued at any time in the event of a failure, and is run from a central queue.
Senior managers are concerned about the complexity of the solution's resource management and the costs involved in repeating the process regularly. They ask for the workload to be automated so it runs once a week, starting Monday and completing by the close of business Friday.
Which architecture should be used to scale the solution at the lowest cost?

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

Answer: C

NEW QUESTION 129
A library is developing an automatic book-borrowing system that uses Amazon Rekognition. Images of library members' faces are stored in an Amazon S3 bucket. When members borrow books, the Amazon Rekognition CompareFaces API operation compares real faces against the stored faces in Amazon S3.
The library needs to improve security by making sure that images are encrypted at rest. Also, when the images are used with Amazon Rekognition. they need to be encrypted in transit. The library also must ensure that the images are not used to improve Amazon Rekognition as a service.
How should a machine learning specialist architect the solution to satisfy these requirements?

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

Answer: B

NEW QUESTION 134
A Data Scientist is developing a binary classifier to predict whether a patient has a particular disease on a series of test results. The Data Scientist has data on 400 patients randomly selected from the population. The disease is seen in 3% of the population.
Which cross-validation strategy should the Data Scientist adopt?

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

Answer: B

NEW QUESTION 136
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 138
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 139
A Machine Learning Specialist uploads a dataset to an Amazon S3 bucket protected with server-side encryption using AWS KMS.
How should the ML Specialist define the Amazon SageMaker notebook instance so it can read the same dataset from Amazon S3?

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

Answer: D

NEW QUESTION 144
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 145
A data scientist is working on a public sector project for an urban traffic system. While studying the traffic patterns, it is clear to the data scientist that the traffic behavior at each light is correlated, subject to a small stochastic error term. The data scientist must model the traffic behavior to analyze the traffic patterns and reduce congestion.
How will the data scientist MOST effectively model the problem?

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

Answer: D

NEW QUESTION 149
A Machine Learning Specialist has created a deep learning neural network model that performs well on the training data but performs poorly on the test data.
Which of the following methods should the Specialist consider using to correct this? (Select THREE.)

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

Answer: BCD

NEW QUESTION 151
A Data Scientist needs to analyze employment data. The dataset contains approximately 10 million observations on people across 10 different features. During the preliminary analysis, the Data Scientist notices that income and age distributions are not normal. While income levels shows a right skew as expected, with fewer individuals having a higher income, the age distribution also show a right skew, with fewer older individuals participating in the workforce.
Which feature transformations can the Data Scientist apply to fix the incorrectly skewed data? (Choose two.)

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

Answer: AB

NEW QUESTION 152
A manufacturing company asks its Machine Learning Specialist to develop a model that classifies defective parts into one of eight defect types. The company has provided roughly 100000 images per defect type for training During the injial training of the image classification model the Specialist notices that the validation accuracy is 80%, while the training accuracy is 90% It is known that human-level performance for this type of image classification is around 90%
What should the Specialist consider to fix this issue1?

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

Answer: D

NEW QUESTION 153
A company sells thousands of products on a public website and wants to automatically identify products with potential durability problems. The company has 1.000 reviews with date, star rating, review text, review summary, and customer email fields, but many reviews are incomplete and have empty fields. Each review has already been labeled with the correct durability result.
A machine learning specialist must train a model to identify reviews expressing concerns over product durability. The first model needs to be trained and ready to review in 2 days.

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

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

Answer: B

NEW QUESTION 156
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 157
A Machine Learning Specialist previously trained a logistic regression model using scikit-learn on a local machine, and the Specialist now wants to deploy it to production for inference only.
What steps should be taken to ensure Amazon SageMaker can host a model that was trained locally?

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

Answer: D

NEW QUESTION 158
An ecommerce company sends a weekly email newsletter to all of its customers. Management has hired a team of writers to create additional targeted content. A data scientist needs to identify five customer segments based on age, income, and location. The customers' current segmentation is unknown. The data scientist previously built an XGBoost model to predict the likelihood of a customer responding to an email based on age, income, and location.
Why does the XGBoost model NOT meet the current requirements, and how can this be fixed?

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

Answer: C

NEW QUESTION 160
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 163
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 165
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.

Data Visualization Plot
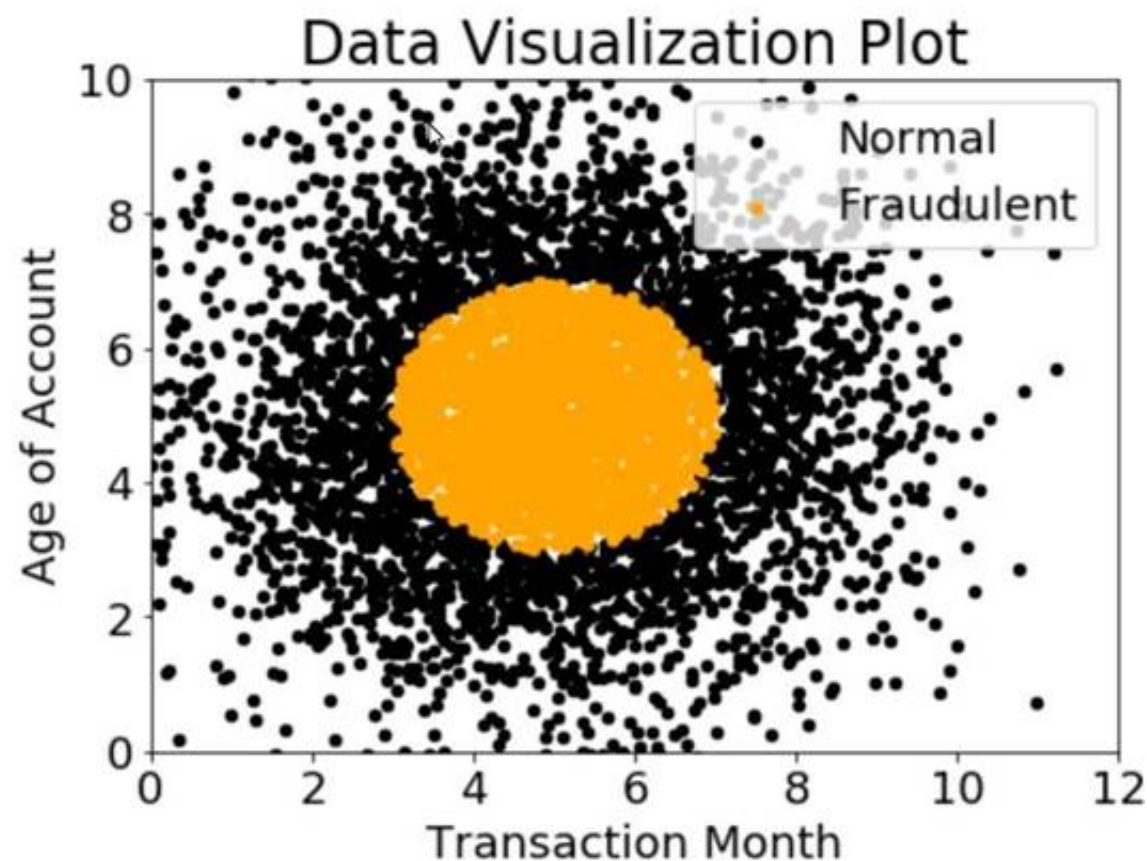
Based on this information, which model would have the HIGHEST recall with respect to the fraudulent class?

A. Decision tree
B. Linear support vector machine (SVM)
C. Naive Bayesian classifier
D. Single Perceptron with sigmoidal activation function

Answer: C

NEW QUESTION 169
A company is running an Amazon SageMaker training job that will access data stored in its Amazon S3 bucket A compliance policy requires that the data never be transmitted across the internet How should the company set up the job?

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

Answer: D

NEW QUESTION 173
A Machine Learning Specialist is building a logistic regression model that will predict whether or not a person will order a pizza. The Specialist is trying to build the optimal model with an ideal classification threshold.
What model evaluation technique should the Specialist use to understand how different classification thresholds will impact the model's performance?

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

Answer: A

NEW QUESTION 176
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 lof 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 180
A Machine Learning Specialist is working with a large company to leverage machine learning within its products. The company wants to group its customers into categories based on which customers will and will not churn within the next 6 months. The company has labeled the data available to the Specialist.
Which machine learning model type should the Specialist use to accomplish this task?

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

Answer: B

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

≫ Will this customer churn or not?

≫ Will a customer renew their subscription?

≫ Will a user downgrade a pricing plan?

≫ Are there any signs of unusual customer behavior?

NEW QUESTION 184
A Machine Learning Specialist is building a prediction model for a large number of features using linear models, such as linear regression and logistic regression
During exploratory data analysis the Specialist observes that many features are highly correlated with each other This may make the model unstable
What should be done to reduce the impact of having such a large number of features?

A. Perform one-hot encoding on highly correlated features
B. Use matrix multiplication on highly correlated features.
C. Create a new feature space using principal component analysis (PCA)
D. Apply the Pearson correlation coefficient

Answer: B

NEW QUESTION 189
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 this instance as a Jupyter notebook in Amazon SageMaker.

Answer: AD

NEW QUESTION 190
A company uses camera images of the tops of items displayed on store shelves to determine which items were removed and which ones still remain. After several hours of data labeling, the company has a total of 1,000 hand-labeled images covering 10 distinct items. The training results were poor.
Which machine learning approach fulfills the company's long-term needs?

A. Convert the images to grayscale and retrain the model
B. Reduce the number of distinct items from 10 to 2, build the model, and iterate
C. Attach different colored labels to each item, take the images again, and build the model

D. Augment training data for each item using image variants like inversions and translations, build the model, and iterate.

Answer: A

NEW QUESTION 192
A Machine Learning Specialist working for an online fashion company wants to build a data ingestion solution for the company's Amazon S3-based data lake.
The Specialist wants to create a set of ingestion mechanisms that will enable future capabilities comprised of:
• Real-time analytics
• Interactive analytics of historical data
• Clickstream analytics
• Product recommendations
Which services should the Specialist use?

A. AWS Glue as the data dialog; Amazon Kinesis Data Streams and Amazon Kinesis Data Analytics for real-time data insights; Amazon Kinesis Data Firehose for delivery to Amazon ES for clickstream analytics; Amazon EMR to generate personalized product recommendations
B. Amazon Athena as the data catalog; Amazon Kinesis Data Streams and Amazon Kinesis Data Analytics for near-realtime data insights; Amazon Kinesis Data Firehose for clickstream analytics; AWS Glue to generate personalized product recommendations
C. AWS Glue as the data catalog; Amazon Kinesis Data Streams and Amazon Kinesis Data Analytics for historical data insights; Amazon Kinesis Data Firehose for delivery to Amazon ES for clickstream analytics; Amazon EMR to generate personalized product recommendations
D. Amazon Athena as the data catalog; Amazon Kinesis Data Streams and Amazon Kinesis Data Analytics for historical data insights; Amazon DynamoDB streams for clickstream analytics; AWS Glue to generate personalized product recommendations

Answer: A

NEW QUESTION 193
A Machine Learning Specialist needs to be able to ingest streaming data and store it in Apache Parquet files for exploration and analysis. Which of the following services would both ingest and store this data in the correct format?

A. AWSDMS
B. Amazon Kinesis Data Streams
C. Amazon Kinesis Data Firehose
D. Amazon Kinesis Data Analytics

Answer: C

NEW QUESTION 198
A Machine Learning Specialist trained a regression model, but the first iteration needs optimizing. The Specialist needs to understand whether the model is more frequently overestimating or underestimating the target.
What option can the Specialist use to determine whether it is overestimating or underestimating the target value?

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

Answer: B

NEW QUESTION 203
An agricultural company is interested in using machine learning to detect specific types of weeds in a 100-acre grassland field. Currently, the company uses tractor-mounted cameras to capture multiple images of the field as 10 × 10 grids. The company also has a large training dataset that consists of annotated images of popular weed classes like broadleaf and non-broadleaf docks.
The company wants to build a weed detection model that will detect specific types of weeds and the location of each type within the field. Once the model is ready, it will be hosted on Amazon SageMaker endpoints. The model will perform real-time inferencing using the images captured by the cameras. Which approach should a Machine Learning Specialist take to obtain accurate predictions?

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

Answer: C

NEW QUESTION 206
A company has set up and deployed its machine learning (ML) model into production with an endpoint using Amazon SageMaker hosting services. The ML team has configured automatic scaling for its SageMaker instances to support workload changes. During testing, the team notices that additional instances are being launched before the new instances are ready. This behavior needs to change as soon as possible.
How can the ML team solve this issue?

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

Answer: A

NEW QUESTION 207
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 211
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 212
A retail company uses a machine learning (ML) model for daily sales forecasting. The company's brand manager reports that the model has provided inaccurate results for the past 3 weeks.
At the end of each day, an AWS Glue job consolidates the input data that is used for the forecasting with the actual daily sales data and the predictions of the model. The AWS Glue job stores the data in Amazon S3. The company's ML team is using an Amazon SageMaker Studio notebook to gain an understanding about the source of the model's inaccuracies.
What should the ML team do on the SageMaker Studio notebook to visualize the model's degradation MOST accurately?

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

Answer: C

NEW QUESTION 213
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 215
A Machine Learning Specialist is assigned a TensorFlow project using Amazon SageMaker for training, and needs to continue working for an extended period with no Wi-Fi access.
Which approach should the Specialist use to continue working?

A. Install Python 3 and boto3 on their laptop and continue the code development using that environment.
B. Download the TensorFlow Docker container used in Amazon SageMaker from GitHub to their local environment, and use the Amazon SageMaker Python SDK to test the code.
C. Download TensorFlow from tensorflow.org to emulate the TensorFlow kernel in the SageMaker environment.
D. Download the SageMaker notebook to their local environment then install Jupyter Notebooks on their laptop and continue the development in a local notebook.

Answer: D

NEW QUESTION 218
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 222
A Machine Learning Specialist is designing a system for improving sales for a company. The objective is to use the large amount of information the company has on users' behavior and product preferences to predict which products users would like based on the users' similarity to other users.
What should the Specialist do to meet this objective?

A. Build a content-based filtering recommendation engine with Apache Spark ML on Amazon EMR.
B. Build a collaborative filtering recommendation engine with Apache Spark ML on Amazon EMR.
C. Build a model-based filtering recommendation engine with Apache Spark ML on Amazon EMR.
D. Build a combinative filtering recommendation engine with Apache Spark ML on Amazon EMR.

Answer: B

Explanation:
Many developers want to implement the famous Amazon model that was used to power the "People who bought this also bought these items" feature on Amazon.com. This model is based on a method called Collaborative Filtering. It takes items such as movies, books, and products that were rated highly by a set of users and recommending them to other users who also gave them high ratings. This method works well in domains where explicit ratings or implicit user actions can be gathered and analyzed.

NEW QUESTION 224
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 cod
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 SageMake
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 228
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 bucke
B. Train and evaluate the model by using Amazon SageMake
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 bucke
G. Deploy the model on an Amazon SageMaker hosting services endpoint.
H. Move the training data to an Amazon S3 bucke
I. Train and evaluate the model by using Amazon SageMake
J. Optimize the model by using SageMaker Ne
K. Set up an edge device in the manufacturing facilities with AWS IoT Greengras
L. Deploy the model on the edge device.
M. Train the model on premise
N. Upload the model to an Amazon S3 bucke
O. Set up an edge device in the manufacturing facilities with AWS IoT Greengras
P. Deploy the model on the edge device.

Answer: A

NEW QUESTION 231
A Machine Learning Specialist is working for an online retailer that wants to run analytics on every customer visit, processed through a machine learning pipeline.
The data needs to be ingested by Amazon Kinesis Data Streams at up to 100 transactions per second, and the JSON data blob is 100 KB in size.
What is the MINIMUM number of shards in Kinesis Data Streams the Specialist should use to successfully ingest this data?

A. 1 shards
B. 10 shards
C. 100 shards

D. 1,000 shards

Answer: B

NEW QUESTION 234

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

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

Answer: D

NEW QUESTION 237

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

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

Answer: D

NEW QUESTION 238

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

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

Answer: C

NEW QUESTION 243

A Machine Learning Specialist wants to determine the appropriate SageMakerVariant Invocations Per Instance setting for an endpoint automatic scaling configuration. The Specialist has performed a load test on a single instance and determined that peak requests per second (RPS) without service degradation is about 20 RPS As this is the first deployment, the Specialist intends to set the invocation safety factor to 0 5
Based on the stated parameters and given that the invocations per instance setting is measured on a per-minute basis, what should the Specialist set as the sageMakervariantinvocationsPerinstance setting?

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

Answer: C

NEW QUESTION 244

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 245

A data scientist uses an Amazon SageMaker notebook instance to conduct data exploration and analysis. This requires certain Python packages that are not natively available on Amazon SageMaker to be installed on the notebook instance.
How can a machine learning specialist ensure that required packages are automatically available on the notebook instance for the data scientist to use?

A. Install AWS Systems Manager Agent on the underlying Amazon EC2 instance and use Systems Manager Automation to execute the package installation

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

Answer: D

Explanation:
https://docs.aws.amazon.com/sagemaker/latest/dg/nbi-add-external.html

NEW QUESTION 246
A Machine Learning Specialist has built a model using Amazon SageMaker built-in algorithms and is not getting expected accurate results The Specialist wants to use hyperparameter optimization to increase the model's accuracy
Which method is the MOST repeatable and requires the LEAST amount of effort to achieve this?

A. Launch multiple training jobs in parallel with different hyperparameters
B. Create an AWS Step Functions workflow that monitors the accuracy in Amazon CloudWatch Logs and relaunches the training job with a defined list of hyperparameters
C. Create a hyperparameter tuning job and set the accuracy as an objective metric.
D. Create a random walk in the parameter space to iterate through a range of values that should be used for each individual hyperparameter

Answer: B

NEW QUESTION 248
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 252
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 255
A Machine Learning Specialist is packaging a custom ResNet model into a Docker container so the company can leverage Amazon SageMaker for training. The Specialist is using Amazon EC2 P3 instances to train the model and needs to properly configure the Docker container to leverage the NVIDIA GPUs.
What does the Specialist need to do?

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

Answer: B

NEW QUESTION 259
A data scientist is using an Amazon SageMaker notebook instance and needs to securely access data stored in a specific Amazon S3 bucket.
How should the data scientist accomplish this?

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

Answer: C

NEW QUESTION 262
For the given confusion matrix, what is the recall and precision of the model?

		Actual	
		Yes	No
Predicted	Yes	12	3
	No	1	9

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

Answer: C

NEW QUESTION 263
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 264
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 267
......

Thank You for Trying Our Product

We offer two products:

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

2nd - Questons 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