Exam Questions MLS-C01

AWS Certified Machine Learning - Specialty

https://www.2passeasy.com/dumps/MLS-C01/

NEW QUESTION 1
A Machine Learning Specialist is implementing a full Bayesian network on a dataset that describes public transit in New York City. One of the random variables is discrete, and represents the number of minutes New Yorkers wait for a bus given that the buses cycle every 10 minutes, with a mean of 3 minutes.
Which prior probability distribution should the ML Specialist use for this variable?

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

Answer: D

NEW QUESTION 2
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: D

NEW QUESTION 3
A Machine Learning Specialist was given a dataset consisting of unlabeled data The Specialist must create a model that can help the team classify the data into different buckets What model should be used to complete this work?

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

Answer: A

NEW QUESTION 4
A company wants to classify user behavior as either fraudulent or normal. Based on internal research, a Machine Learning Specialist would like to build a binary classifier based on two features: age of account and transaction month. The class distribution for these features is illustrated in the figure provided.
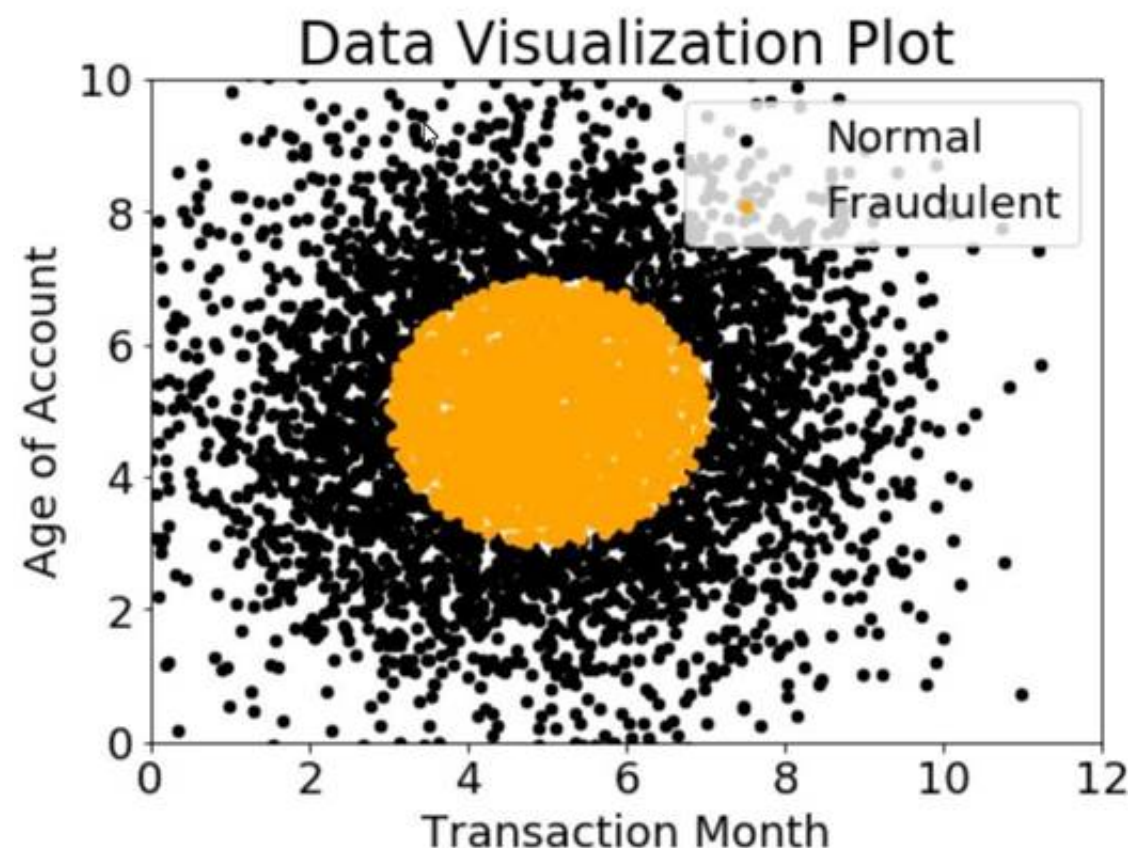
Based on this information which model would have the HIGHEST accuracy?

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

Answer: B

NEW QUESTION 5
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: A

NEW QUESTION 6
The displayed graph is from a foresting model for testing a time series.
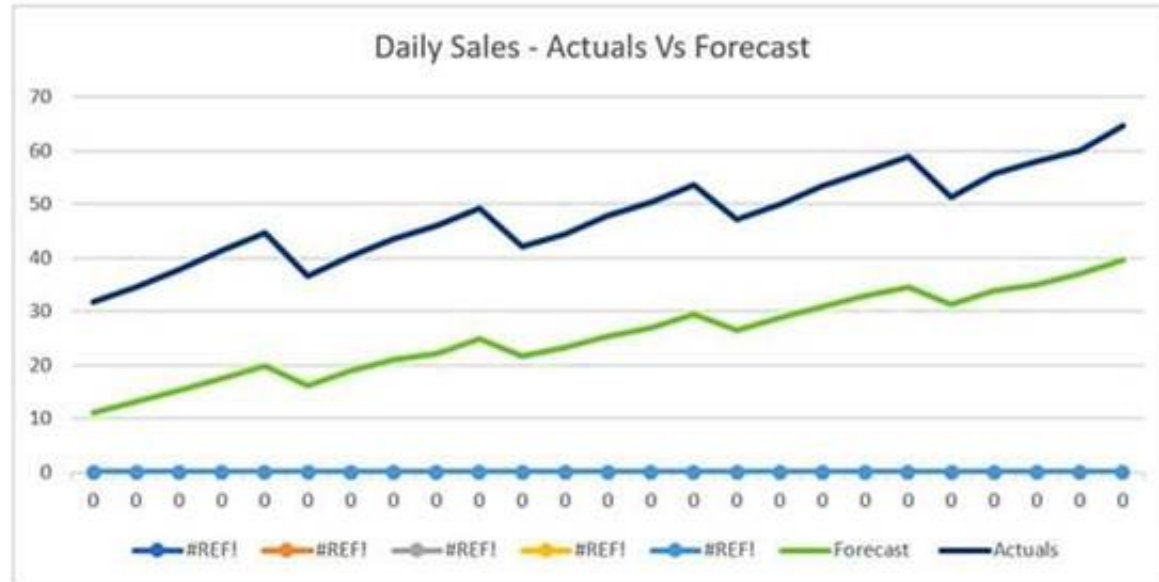
Considering the graph only, which conclusion should a Machine Learning Specialist make about the behavior of the model?

A. The model predicts both the trend and the seasonality well.
B. The model predicts the trend well, but not the seasonality.
C. The model predicts the seasonality well, but not the trend.
D. The model does not predict the trend or the seasonality well.

Answer: D

NEW QUESTION 7
While reviewing the histogram for residuals on regression evaluation data a Machine Learning Specialist notices that the residuals do not form a zero-centered bell shape as shown What does this mean?
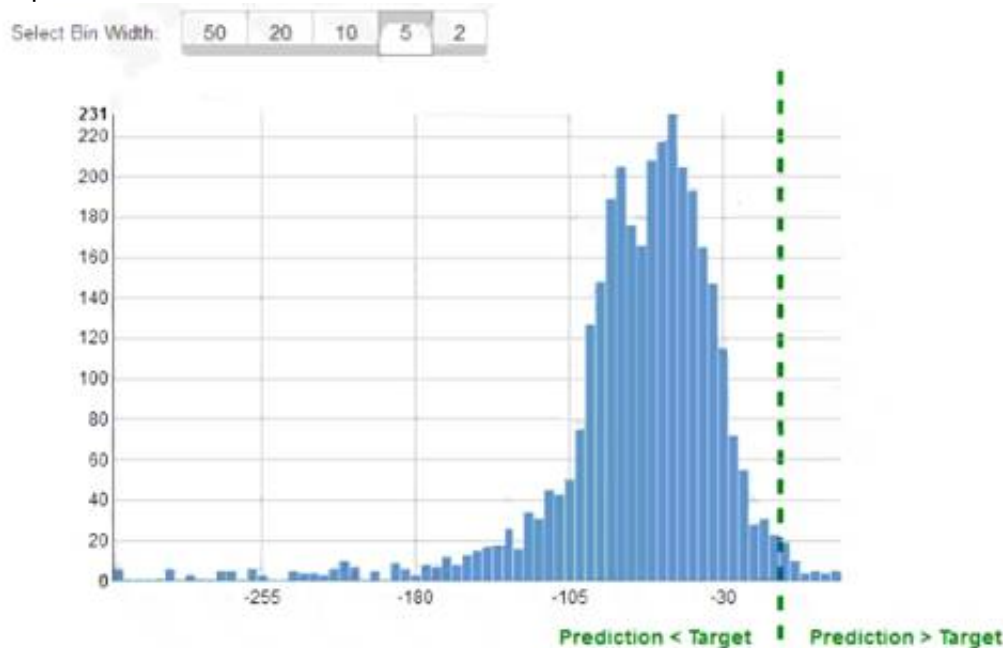
A. The model might have prediction errors over a range of target values.
B. The dataset cannot be accurately represented using the regression model
C. There are too many variables in the model
D. The model is predicting its target values perfectly.

Answer: D

NEW QUESTION 8
A 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 9
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 10
A large consumer goods manufacturer has the following products on sale
• 34 different toothpaste variants
• 48 different toothbrush variants
• 43 different mouthwash variants
The entire sales history of all these products is available in Amazon S3 Currently, the company is using custom-built autoregressive integrated moving average (ARIMA) models to forecast demand for these products The company wants to predict the demand for a new product that will soon be launched
Which solution should a Machine Learning Specialist apply?

A. Train a custom ARIMA model to forecast demand for the new product.
B. Train an Amazon SageMaker DeepAR algorithm to forecast demand for the new product
C. Train an Amazon SageMaker k-means clustering algorithm to forecast demand for the new product.
D. Train a custom XGBoost model to forecast demand for the new product

Answer: B

Explanation:
The Amazon SageMaker DeepAR forecasting algorithm is a supervised learning algorithm for forecasting scalar (one-dimensional) time series using recurrent neural networks (RNN). Classical forecasting methods, such as autoregressive integrated moving average (ARIMA) or exponential smoothing (ETS), fit a single model to each individual time series. They then use that model to extrapolate the time series into the future.

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

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

Answer: BD

NEW QUESTION 13
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: A

NEW QUESTION 17
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 21
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 25
An office security agency conducted a successful pilot using 100 cameras installed at key locations within the main office. Images from the cameras were uploaded to Amazon S3 and tagged using Amazon Rekognition, and the results were stored in Amazon ES. The agency is now looking to expand the pilot into a full production system using thousands of video cameras in its office locations globally. The goal is to identify activities performed by non-employees in real time. Which solution should the agency consider?

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

Answer: D

NEW QUESTION 26
A Machine Learning Specialist is developing a custom video recommendation model for an application The dataset used to train this model is very large with millions of data points and is hosted in an Amazon S3 bucket The Specialist wants to avoid loading all of this data onto an Amazon SageMaker notebook instance because it would take hours to move and will exceed the attached 5 GB Amazon EBS volume on the notebook instance.
Which approach allows the Specialist to use all the data to train the model?

A. Load a smaller subset of the data into the SageMaker notebook and train locall
B. Confirm that the training code is executing and the model parameters seem reasonabl
C. Initiate a SageMaker training job using the full dataset from the S3 bucket using Pipe input mode.
D. Launch an Amazon EC2 instance with an AWS Deep Learning AMI and attach the S3 bucket to theinstanc
E. Train on a small amount of the data to verify the training code and hyperparameter
F. Go back toAmazon SageMaker and train using the full dataset
G. Use AWS Glue to train a model using a small subset of the data to confirm that the data will be compatiblewith Amazon SageMake
H. Initiate a SageMaker training job using the full dataset from the S3 bucket usingPipe input mode.
I. Load a smaller subset of the data into the SageMaker notebook and train locall
J. Confirm that the training code is executing and the model parameters seem reasonabl
K. Launch an Amazon EC2 instance with an AWS Deep Learning AMI and attach the S3 bucket to train the full dataset.

Answer: A

NEW QUESTION 28
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 30
A manufacturing company has structured and unstructured data stored in an Amazon S3 bucket. A Machine Learning Specialist wants to use SQL to run queries

on this data.
Which solution requires the LEAST effort to be able to query this data?

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

Answer: B

NEW QUESTION 33
A company has raw user and transaction data stored in AmazonS3 a MySQL database, and Amazon RedShift A Data Scientist needs to perform an analysis by joining the three datasets from Amazon S3, MySQL, and Amazon RedShift, and then calculating the average-of a few selected columns from the joined data Which AWS service should the Data Scientist use?

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

Answer: A

NEW QUESTION 36
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: D

NEW QUESTION 41
A financial services company is building a robust serverless data lake on Amazon S3. The data lake should be flexible and meet the following requirements:
* Support querying old and new data on Amazon S3 through Amazon Athena and Amazon Redshift Spectrum.
* Support event-driven ETL pipelines.
* Provide a quick and easy way to understand metadata. Which approach meets trfese requirements?

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

Answer: B

NEW QUESTION 46
A Machine Learning Specialist is using Amazon SageMaker to host a model for a highly available customer-facing application .
The Specialist has trained a new version of the model, validated it with historical data, and now wants to deploy it to production To limit any risk of a negative customer experience, the Specialist wants to be able to monitor the model and roll it back, if needed
What is the SIMPLEST approach with the LEAST risk to deploy the model and roll it back, if needed?

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

Answer: A

NEW QUESTION 48
A Machine Learning Specialist works for a credit card processing company and needs to predict which transactions may be fraudulent in near-real time.
Specifically, the Specialist must train a model that returns the probability that a given transaction may be fraudulent
How should the Specialist frame this business problem'?

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

Answer: A

NEW QUESTION 51
An e-commerce company needs a customized training model to classify images of its shirts and pants products The company needs a proof of concept in 2 to 3 days with good accuracy Which compute choice should the Machine Learning Specialist select to train and achieve good accuracy on the model quickly?

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

Answer: C

NEW QUESTION 53
A Machine Learning Specialist is working with a media company to perform classification on popular articles from the company's website. The company is using random forests to classify how popular an article will be before it is published A sample of the data being used is below.
Given the dataset, the Specialist wants to convert the Day-Of_Week column to binary values. What technique should be used to convert this column to binary values.

Article_Title	Author	Top_Keywords	Day_Of_Week	URL_of_Article	Page_Views
Building a Big Data Platform	Jane Doe	Big Data, Spark, Hadoop	Tuesday	http://examplecorp.com/data_platform.html	1300456
Getting Started with Deep Learning	John Doe	Deep Learning, Machine Learning, Spark	Tuesday	http://examplecorp.com/started_deep_learning.html	1230661
MXNet ML Guide	Jane Doe	Machine Learning, MXNet, Logistic Regression	Thursday	http://examplecorp.com/mxnet_guide.html	937291
Intro to NoSQL Databases	Mary Major	NoSQL, Operations, Database	Monday	http://examplecorp.com/nosql_intro_guide.html	407812

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

Answer: B

NEW QUESTION 58
A Machine Learning Specialist is configuring Amazon SageMaker so multiple Data Scientists can access notebooks, train models, and deploy endpoints. To ensure the best operational performance, the Specialist needs to be able to track how often the Scientists are deploying models, GPU and CPU utilization on the deployed SageMaker endpoints, and all errors that are generated when an endpoint is invoked.
Which services are integrated with Amazon SageMaker to track this information? (Select TWO.)

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

Answer: AD

NEW QUESTION 63
During mini-batch training of a neural network for a classification problem, a Data Scientist notices that training accuracy oscillates What is the MOST likely cause of this issue?

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

Answer: D

NEW QUESTION 66
A Machine Learning Specialist needs to create a data repository to hold a large amount of time-based training data for a new model. In the source system, new files are added every hour Throughout a single 24-hour period, the volume of hourly updates will change significantly. The Specialist always wants to train on the last 24 hours of the data
Which type of data repository is the MOST cost-effective solution?

A. An Amazon EBS-backed Amazon EC2 instance with hourly directories
B. An Amazon RDS database with hourly table partitions
C. An Amazon S3 data lake with hourly object prefixes

D. An Amazon EMR cluster with hourly hive partitions on Amazon EBS volumes

Answer: C

NEW QUESTION 67
A manufacturing company has a large set of labeled historical sales data The manufacturer would like to predict how many units of a particular part should be produced each quarter Which machine learning approach should be used to solve this problem?

A. Logistic regression
B. Random Cut Forest (RCF)
C. Principal component analysis (PCA)
D. Linear regression

Answer: B

NEW QUESTION 68
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 73
A Data Scientist is developing a machine learning model to predict future patient outcomes based on information collected about each patient and their treatment plans. The model should output a continuous value as its prediction. The data available includes labeled outcomes for a set of 4,000 patients. The study was conducted on a group of individuals over the age of 65 who have a particular disease that is known to worsen with age.
Initial models have performed poorly. While reviewing the underlying data, the Data Scientist notices that, out of 4,000 patient observations, there are 450 where the patient age has been input as 0. The other features for these observations appear normal compared to the rest of the sample population.
How should the Data Scientist correct this issue?

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

Answer: A

NEW QUESTION 75
......

THANKS FOR TRYING THE DEMO OF OUR PRODUCT

Visit Our Site to Purchase the Full Set of Actual MLS-C01 Exam Questions With Answers.

We Also Provide Practice Exam Software That Simulates Real Exam Environment And Has Many Self-Assessment Features. Order the MLS-C01 Product From:

https://www.2passeasy.com/dumps/MLS-C01/

Money Back Guarantee

MLS-C01 Practice Exam Features:

* MLS-C01 Questions and Answers Updated Frequently

* MLS-C01 Practice Questions Verified by Expert Senior Certified Staff

* MLS-C01 Most Realistic Questions that Guarantee you a Pass on Your FirstTry

* MLS-C01 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year