Exam Questions MLS-C01

AWS Certified Machine Learning - Specialty

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

NEW QUESTION 1
A Data Science team within a large company uses Amazon SageMaker notebooks to access data stored in Amazon S3 buckets. The IT Security team is concerned that internet-enabled notebook instances create a security vulnerability where malicious code running on the instances could compromise data privacy. The company mandates that all instances stay within a secured VPC with no internet access, and data communication traffic must stay within the AWS network.
How should the Data Science team configure the notebook instance placement to meet these requirements?

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

Answer: D

NEW QUESTION 2
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 3
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: A

NEW QUESTION 4
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: BD

NEW QUESTION 5
A Machine Learning Specialist is required to build a supervised image-recognition model to identify a cat. The ML Specialist performs some tests and records the following results for a neural network-based image classifier:
Total number of images available = 1,000 Test set images = 100 (constant test set)
The ML Specialist notices that, in over 75% of the misclassified images, the cats were held upside down by their owners.
Which techniques can be used by the ML Specialist to improve this specific test error?

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

Answer: B

NEW QUESTION 6
Example Corp has an annual sale event from October to December. The company has sequential sales data from the past 15 years and wants to use Amazon ML to predict the sales for this year's upcoming event. Which method should Example Corp use to split the data into a training dataset and evaluation dataset?

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

Answer: C

NEW QUESTION 7
A city wants to monitor its air quality to address the consequences of air pollution A Machine Learning Specialist needs to forecast the air quality in parts per million of contaminates for the next 2 days in the city As this is a prototype, only daily data from the last year is available
Which model is MOST likely to provide the best results in Amazon SageMaker?

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

Answer: C

NEW QUESTION 8
A Machine Learning team uses Amazon SageMaker to train an Apache MXNet handwritten digit classifier model using a research dataset. The team wants to receive a notification when the model is overfitting. Auditors want to view the Amazon SageMaker log activity report to ensure there are no unauthorized API calls.
What should the Machine Learning team do to address the requirements with the least amount of code and fewest steps?

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

Answer: C

NEW QUESTION 9
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: BD

NEW QUESTION 10
A Machine Learning Specialist was given a dataset consisting of unlabeled data The Specialist must create a model that can help the team classify the data into different buckets What model should be used to complete this work?

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

Answer: A

NEW QUESTION 10
A Machine Learning Specialist is 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 13
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 14
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 17
A company wants to classify user behavior as either fraudulent or normal. Based on internal research, a Machine Learning Specialist would like to build a binary classifier based on two features: age of account and transaction month. The class distribution for these features is illustrated in the figure provided.
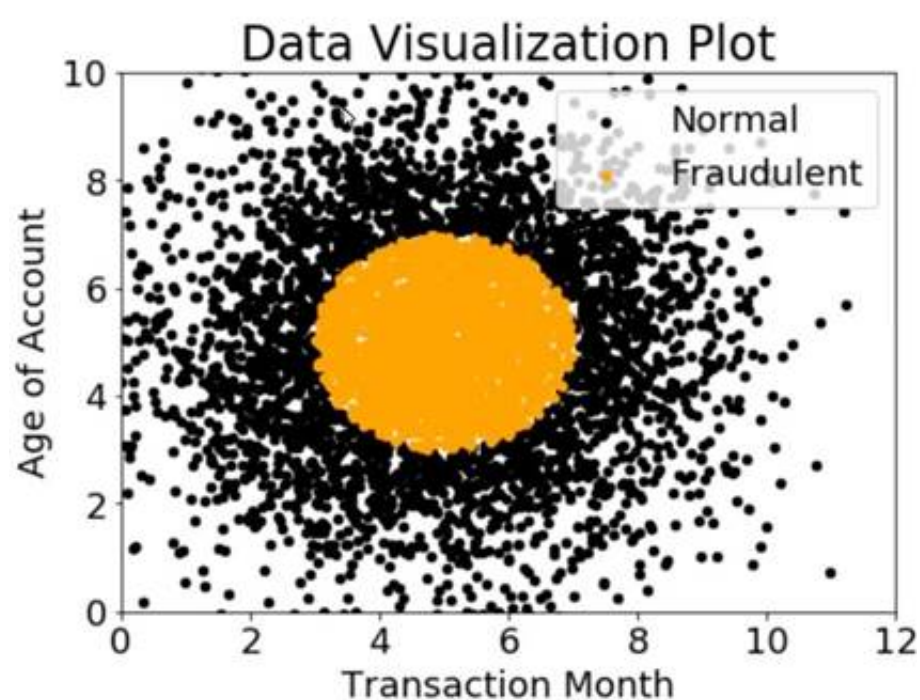
Based on this information, which model would have the HIGHEST recall with respect to the fraudulent class?

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

Answer: C

NEW QUESTION 20
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 24
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: ABD

NEW QUESTION 25
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 26
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 27
A large mobile network operating company is building a machine learning model to predict customers who are likely to unsubscribe from the service. The company plans to offer an incentive for these customers as the cost of churn is far greater than the cost of the incentive.
The model produces the following confusion matrix after evaluating on a test dataset of 100 customers: Based on the model evaluation results, why is this a viable model for production?

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

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

Answer: B

NEW QUESTION 32
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 36

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 37
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 do1?

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 Create TrainingJob request body

Answer: A

NEW QUESTION 39
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 44
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 48
A Mobile Network Operator is building an analytics platform to analyze and optimize a company's operations using Amazon Athena and Amazon S3
The source systems send data in CSV format in real lime The Data Engineering team wants to transform the data to the Apache Parquet format before storing it on Amazon S3
Which solution takes the LEAST effort to implement?

A. Ingest .CSV data using Apache Kafka Streams on Amazon EC2 instances and use Kafka Connect S3 to serialize data as Parquet
B. Ingest .CSV data from Amazon Kinesis Data Streams and use Amazon Glue to convert data into Parquet.
C. Ingest .CSV data using Apache Spark Structured Streaming in an Amazon EMR cluster and use Apache Spark to convert data into Parquet.
D. Ingest .CSV data from Amazon Kinesis Data Streams and use Amazon Kinesis Data Firehose to convert data into Parquet.

Answer: C

NEW QUESTION 51
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 55
A Machine Learning Specialist has completed a proof of concept for a company using a small data sample and now the Specialist is ready to implement an end-to-end solution in AWS using Amazon SageMaker The historical training data is stored in Amazon RDS
Which approach should the Specialist use for training a model using that data?

A. Write a direct connection to the SQL database within the notebook and pull data in

B. Push the data from Microsoft SQL Server to Amazon S3 using an AWS Data Pipeline and provide the S3 location within the notebook.
C. Move the data to Amazon DynamoDB and set up a connection to DynamoDB within the notebook to pull data in
D. Move the data to Amazon ElastiCache using AWS DMS and set up a connection within the notebook to pull data in for fast access.

Answer: B

NEW QUESTION 56
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: C

NEW QUESTION 59
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 63
A Machine Learning Specialist is developing a daily ETL workflow containing multiple ETL jobs The workflow consists of the following processes
* Start the workflow as soon as data is uploaded to Amazon S3
* When all the datasets are available in Amazon S3, start an ETL job to join the uploaded datasets with multiple terabyte-sized datasets already stored in Amazon S3
* Store the results of joining datasets in Amazon S3
* If one of the jobs fails, send a notification to the Administrator Which configuration will meet these requirements?

A. Use AWS Lambda to trigger an AWS Step Functions workflow to wait for dataset uploads to complete in Amazon S3. Use AWS Glue to join the datasets Use an Amazon CloudWatch alarm to send an SNS notification to the Administrator in the case of a failure
B. Develop the ETL workflow using AWS Lambda to start an Amazon SageMaker notebook instance Use a lifecycle configuration script to join the datasets and persist the results in Amazon S3 Use an Amazon CloudWatch alarm to send an SNS notification to the Administrator in the case of a failure
C. Develop the ETL workflow using AWS Batch to trigger the start of ETL jobs when data is uploaded to Amazon S3 Use AWS Glue to join the datasets in Amazon S3 Use an Amazon CloudWatch alarm to send an SNS notification to the Administrator in the case of a failure
D. Use AWS Lambda to chain other Lambda functions to read and join the datasets in Amazon S3 as soon as the data is uploaded to Amazon S3 Use an Amazon CloudWatch alarm to send an SNS notification to the Administrator in the case of a failure

Answer: A

NEW QUESTION 68
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 69
A Data Scientist is developing a machine learning model to classify whether a financial transaction is fraudulent. The labeled data available for training consists of 100,000 non-fraudulent observations and 1,000 fraudulent observations.
The Data Scientist applies the XGBoost algorithm to the data, resulting in the following confusion matrix when the trained model is applied to a previously unseen validation dataset. The accuracy of the model is 99.1%, but the Data Scientist has been asked to reduce the number of false negatives.

```
Predicted       0      1
Actual     0 99,966 | 34
           1    877 | 123
```

Which combination of steps should the Data Scientist take to reduce the number of false positive predictions by the model? (Select TWO.)

A. Change the XGBoost eval_metric parameter to optimize based on rmse instead of error.
B. Increase the XGBoost scale_pos_weight parameter to adjust the balance of positive and negative weights.
C. Increase the XGBoost max_depth parameter because the model is currently underfitting the data.

D. Change the XGBoost evaljnetric parameter to optimize based on AUC instead of error.
E. Decrease the XGBoost max_depth parameter because the model is currently overfitting the data.

Answer: DE

NEW QUESTION 74
When submitting Amazon SageMaker training jobs using one of the built-in algorithms, which common parameters MUST be specified? (Select THREE.)

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

Answer: AEF

NEW QUESTION 78
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 81
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 85
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 89
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 92
An online reseller has a large, multi-column dataset with one column missing 30% of its data A Machine Learning Specialist believes that certain columns in the dataset could be used to reconstruct the missing data
Which reconstruction approach should the Specialist use to preserve the integrity of the dataset?
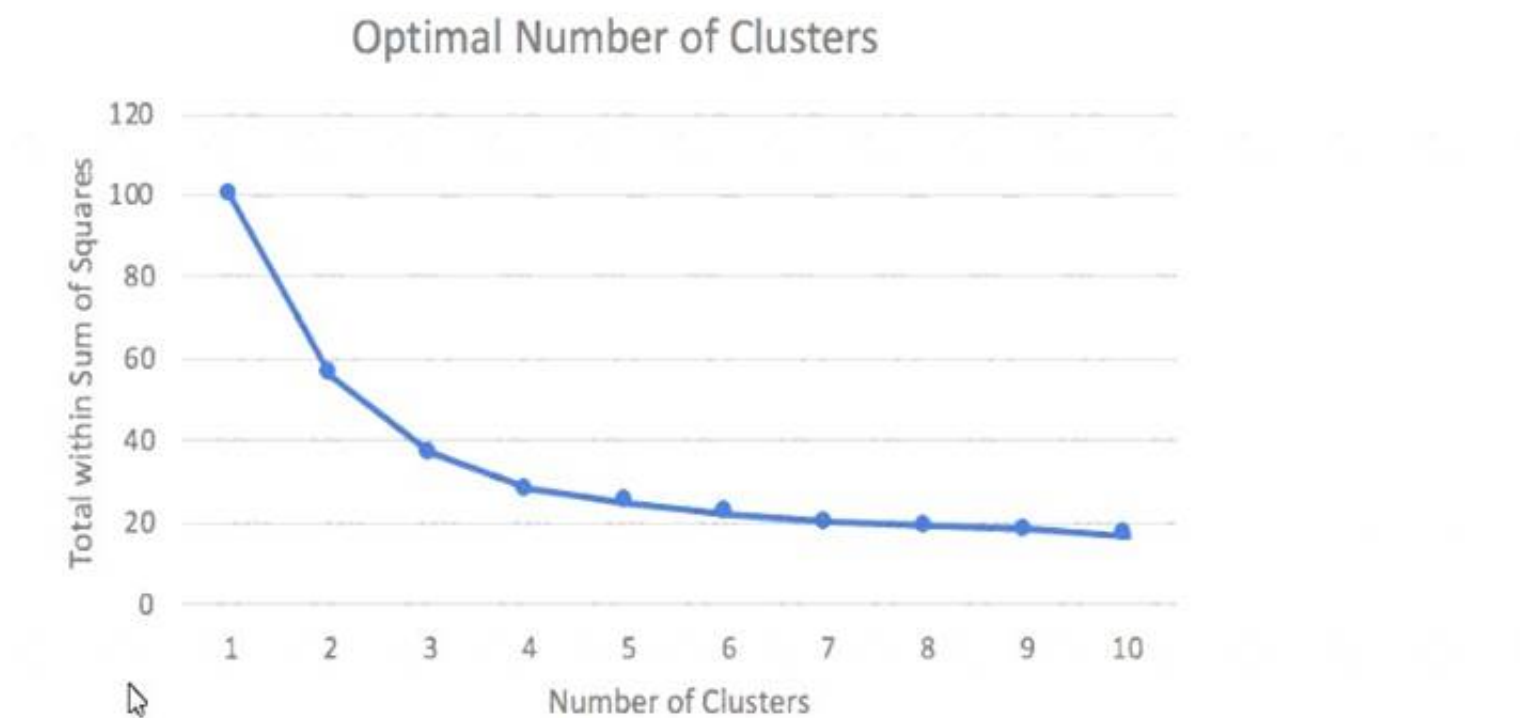
A. Listwise deletion

B. Last observation carried forward
C. Multiple imputation
D. Mean substitution

Answer: C

NEW QUESTION 96
A Machine Learning Specialist prepared the following graph displaying the results of k-means for k = [1:10]

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

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

Answer: C

NEW QUESTION 98
A Machine Learning Specialist must build out a process to query a dataset on Amazon S3 using Amazon Athena The dataset contains more than 800.000 records stored as plaintext CSV files Each record contains 200 columns and is approximately 1 5 MB in size Most queries will span 5 to 10 columns only
How should the Machine Learning Specialist transform the dataset to minimize query runtime?

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

Answer: A

NEW QUESTION 102
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 107
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 111
A Machine Learning Specialist kicks off a hyperparameter tuning job for a tree-based ensemble model using Amazon SageMaker with Area Under the ROC Curve (AUC) as the objective metric This workflow will eventually be deployed in a pipeline that retrains and tunes hyperparameters each night to model click-through on

data that goes stale every 24 hours
With the goal of decreasing the amount of time it takes to train these models, and ultimately to decrease costs, the Specialist wants to reconfigure the input hyperparameter range(s)
Which visualization will accomplish this?

A. A histogram showing whether the most important input feature is Gaussian.
B. A scatter plot with points colored by target variable that uses (-Distributed Stochastic Neighbor Embedding (I-SNE) to visualize the large number of input variables in an easier-to-read dimension.
C. A scatter plot showing (he performance of the objective metric over each training iteration
D. A scatter plot showing the correlation between maximum tree depth and the objective metric.

Answer: B

NEW QUESTION 112
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 117
A Marketing Manager at a pet insurance company plans to launch a targeted marketing campaign on social media to acquire new customers Currently, the company has the following data in Amazon Aurora
• Profiles for all past and existing customers
• Profiles for all past and existing insured pets
• Policy-level information
• Premiums received
• Claims paid
What steps should be taken to implement a machine learning model to identify potential new customers on social media?

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

Answer: C

NEW QUESTION 118
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: A

NEW QUESTION 122
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
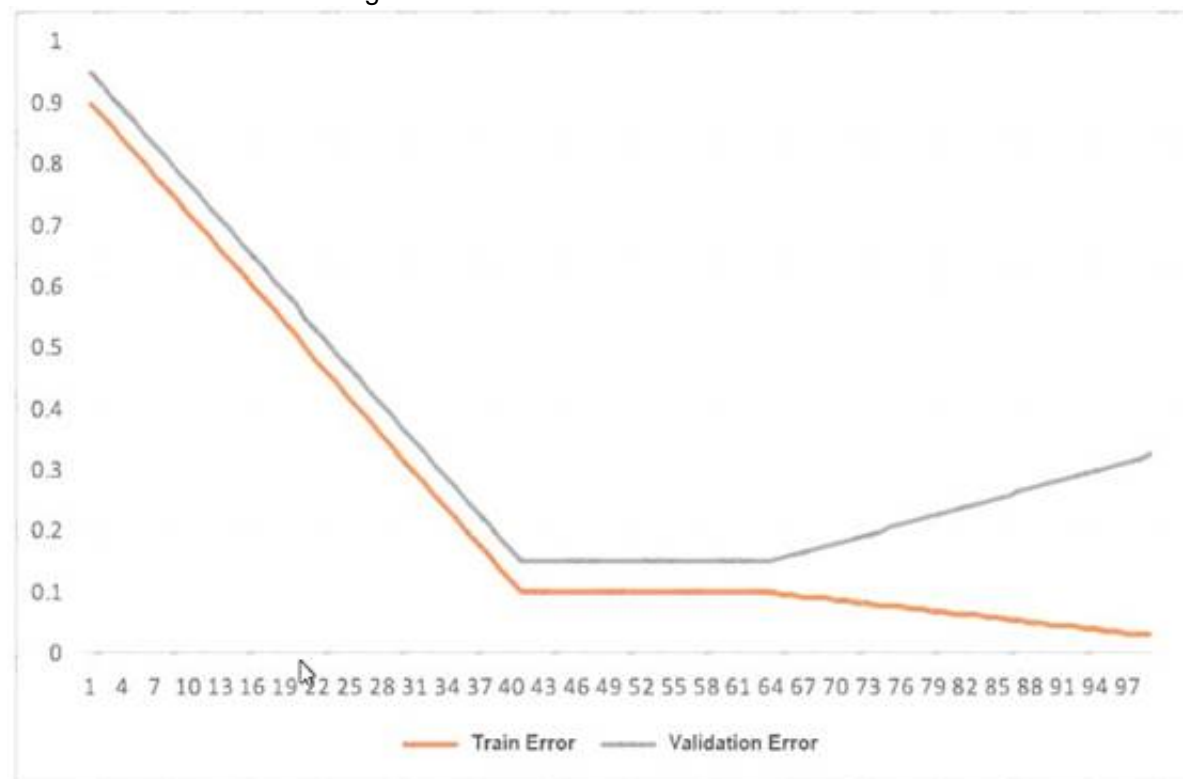
Which technique can be used to improve model performance in terms of accuracy in the validation set?

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 125
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 128
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 service accounts.
D. Amazon SageMaker notebook instances are based on AWS ECS instances running within AWS service accounts.

Answer: C

NEW QUESTION 133
IT leadership wants Jo transition a company's existing machine learning data storage environment to AWS as a temporary ad hoc solution The company currently uses a custom software process that heavily leverages SOL as a query language and exclusively stores generated csv documents for machine learning
The ideal state for the company would be a solution that allows it to continue to use the current workforce of SQL experts The solution must also support the storage of csv and JSON files, and be able to query over semi-structured data The following are high priorities for the company:
• Solution simplicity
• Fast development time
• Low cost
• High flexibility
What technologies meet the company's requirements?

A. Amazon S3 and Amazon Athena
B. Amazon Redshift and AWS Glue
C. Amazon DynamoDB and DynamoDB Accelerator (DAX)
D. Amazon RDS and Amazon ES

Answer: B

NEW QUESTION 138

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 139
An Amazon SageMaker notebook instance is launched into Amazon VPC The SageMaker notebook references data contained in an Amazon S3 bucket in another account The bucket is encrypted using SSE-KMS The instance returns an access denied error when trying to access data in Amazon S3.
Which of the following are required to access the bucket and avoid the access denied error? (Select THREE)

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

Answer: ACF

NEW QUESTION 140
......

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