[2024 Pass Microsoft DP-100 Premium Files Test Engine pdf - Free Dumps Collection [Q131-Q149

[2024] Pass Microsoft DP-100 Premium Files Test Engine pdf - Free Dumps Collection
New 2024 Realistic DP-100 Dumps Test Engine Exam Questions in here

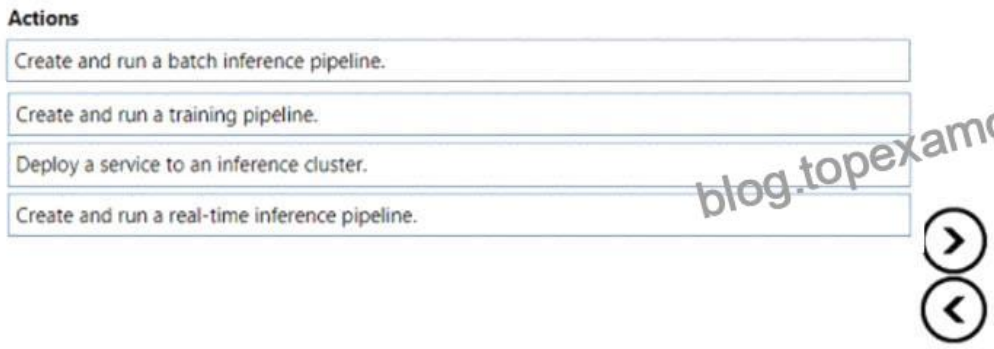
Microsoft DP-100 certification exam is a valuable credential for data scientists and machine learning engineers who want to demonstrate their proficiency in designing and implementing data science solutions on Azure. DP-100 exam covers a wide range of topics related to data science and machine learning and requires candidates to have a deep understanding of Azure data services. To prepare for the exam, candidates can take advantage of various resources provided by Microsoft, including online training courses, study guides, and practice exams.

Microsoft DP-100 certification exam is a comprehensive assessment of the candidate's knowledge and expertise in the field of data science. DP-100 exam covers a wide range of topics, including data exploration and preparation, modeling, feature engineering, training and tuning models, and deploying and managing models in Microsoft Azure. DP-100 exam is designed to test the candidate's ability to design and implement data science solutions using Microsoft Azure data services, including Azure Machine Learning, Azure Databricks, and Azure HDInsight.

NO.131 You are developing a machine learning solution by using the Azure Machine Learning designer.

You need to create a web service that applications can use to submit data feature values and retrieve a predicted label.

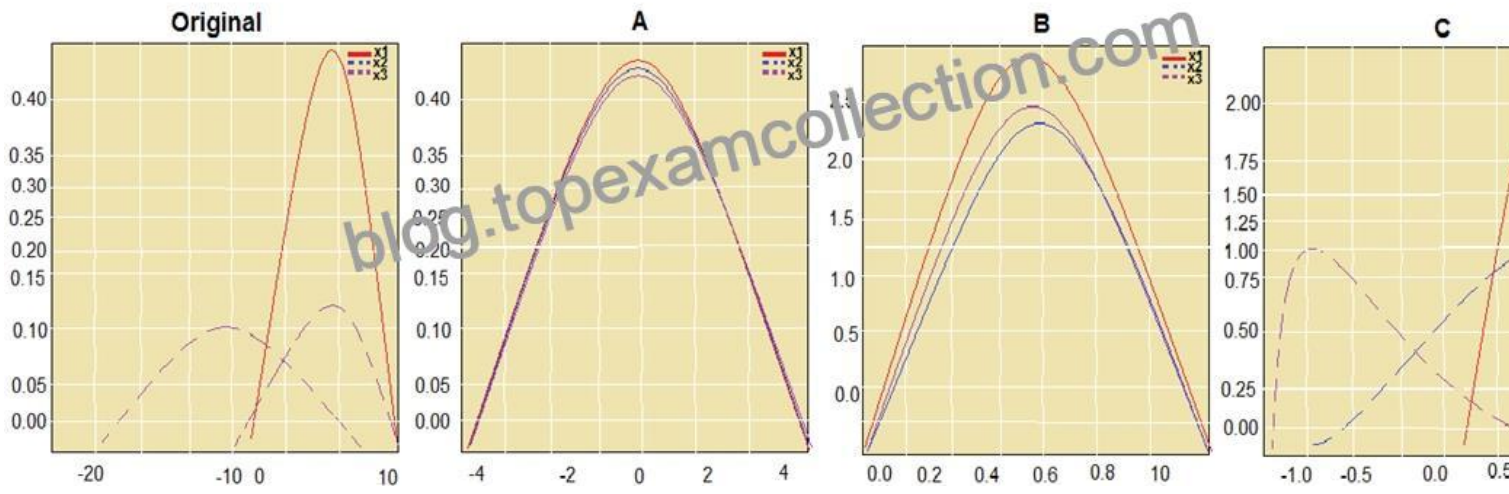
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



Explanation

NO.132 You are performing feature scaling by using the scikit-learn Python library for x.1 x2, and x3 features.

Original and scaled data is shown in the following image.



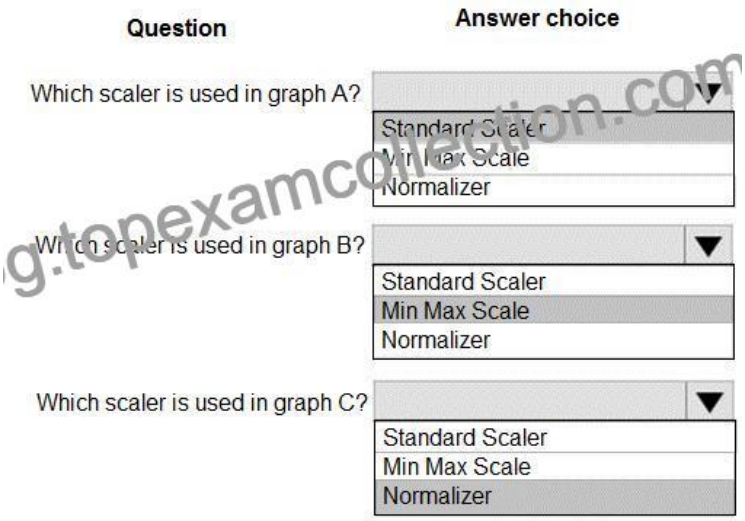
Use the drop-down menus to select the answer choice that answers each question based on the information presented in the graphic.

NOTE: Each correct selection is worth one point.

Question	Answer choice
Which scaler is used in graph A?	Standard Scaler / Min Max Scale / Normalizer
Which scaler is used in graph B?	Standard Scaler / Min Max Scale / Normalizer
Which scaler is used in graph C?	Standard Scaler / Min Max Scale / Normalizer

Question	Answer choice
Which scaler is used in graph A?	Standard Scaler / Min Max Scale / Normalizer
Which scaler is used in graph B?	Standard Scaler / Min Max Scale / Normalizer
Which scaler is used in graph C?	Standard Scaler / Min Max Scale / Normalizer

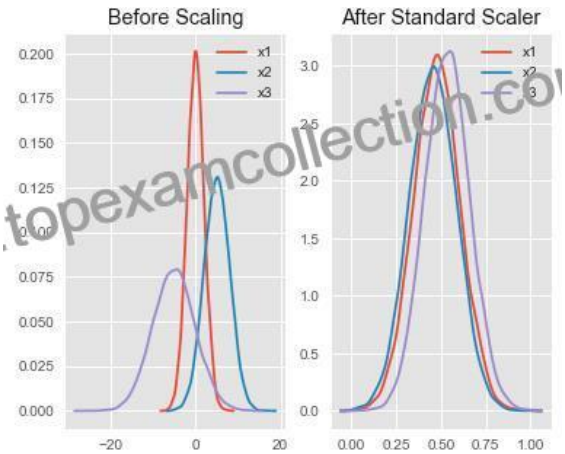
Explanation



Box 1: StandardScaler

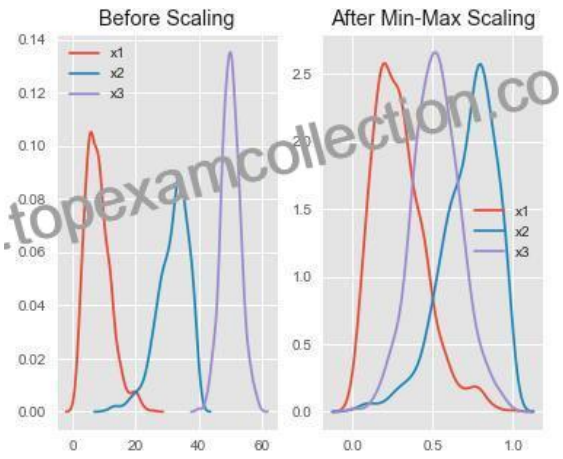
The StandardScaler assumes your data is normally distributed within each feature and will scale them such that the distribution is now centred around 0, with a standard deviation of 1.

Example:



All features are now on the same scale relative to one another.

Box 2: Min Max Scaler



Notice that the skewness of the distribution is maintained but the 3 distributions are brought into the same scale so that they overlap.

Box 3: Normalizer

References:

http://benalexkeen.com/feature-scaling-with-scikit-learn/

NO.133 You are implementing hyperparameter tuning for a model training from a notebook. The notebook is in an Azure Machine Learning workspace. You add code that imports all relevant Python libraries.

You must configure Bayesian sampling over the search space for the num_hidden_layers and batch_size hyperparameters.

You need to complete the following Python code to configure Bayesian sampling.

Which code segments should you use? To answer, select the appropriate options in the answer area NOTE: Each correct selection is worth one point.

Answer Area


```
param_sampling = BayesianParameterSampling( {
    "learning_rate": uniform(0.05, 0.1),
    "batch_size": loguniform ( range (16, 128, 16))
  }
)
```

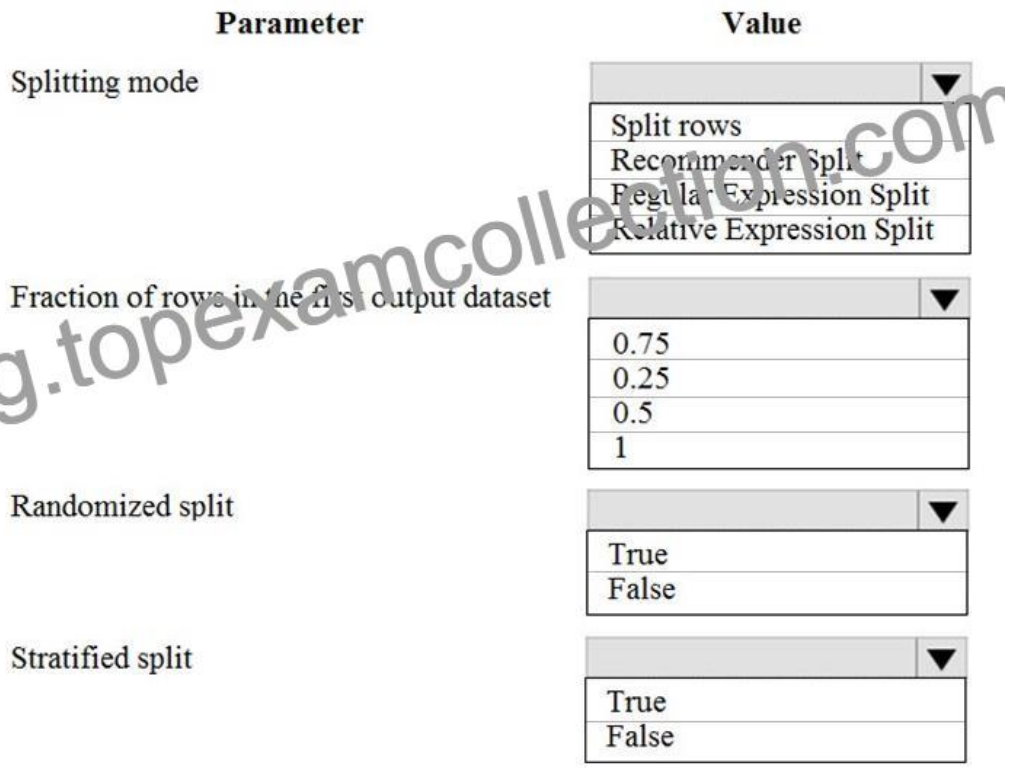
NO.134 You are performing a classification task in Azure Machine Learning Studio.

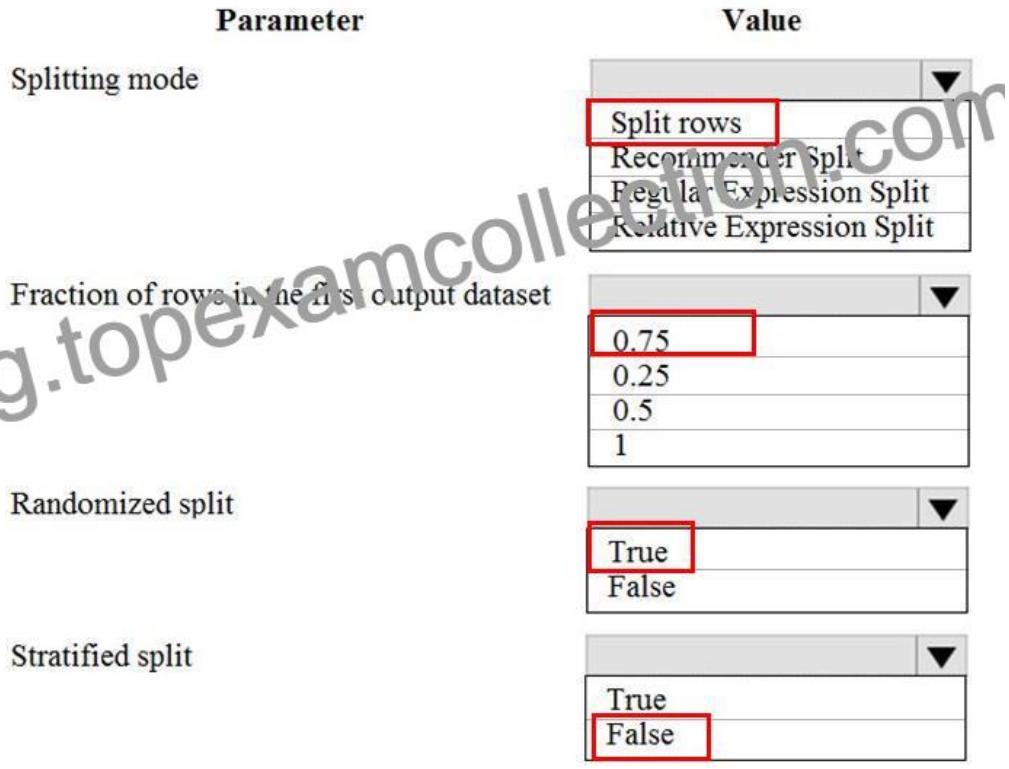
You must prepare balanced testing and training samples based on a provided data set.

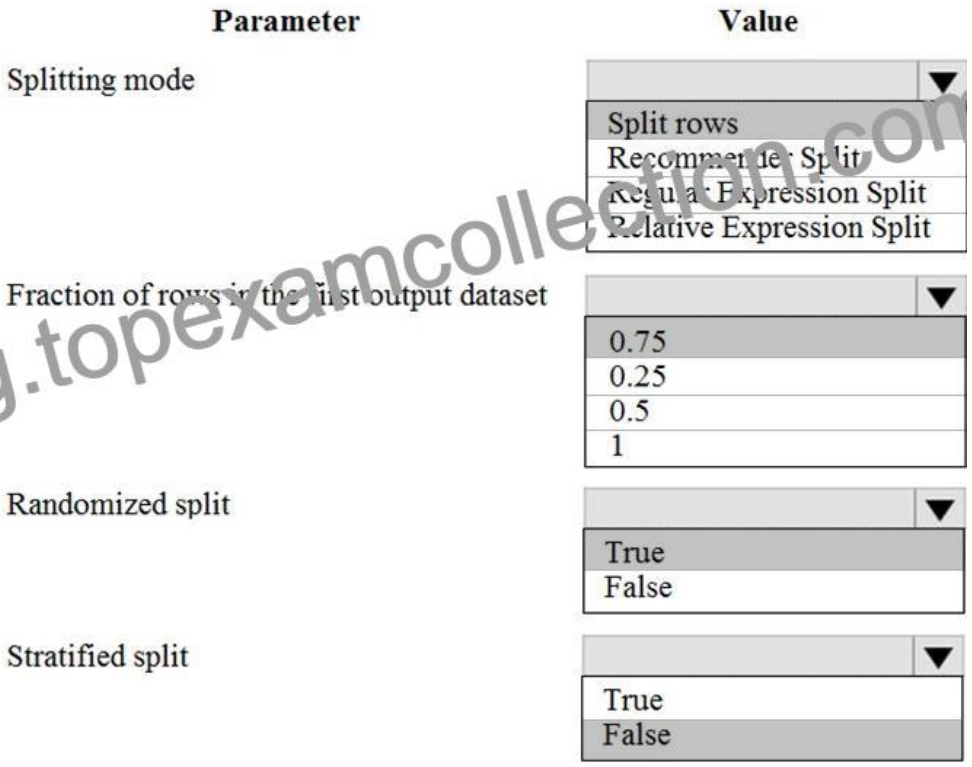
You need to split the data with a 0.75:0.25 ratio.

Which value should you use for each parameter? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.





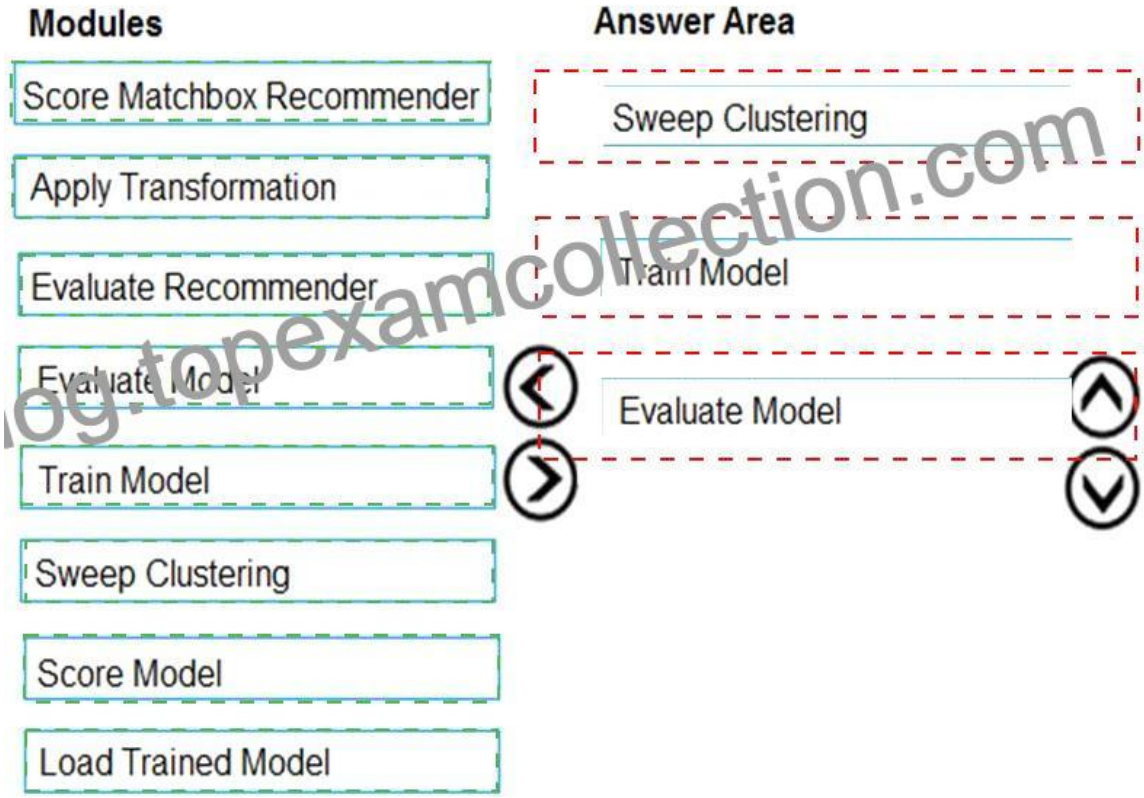


Reference:

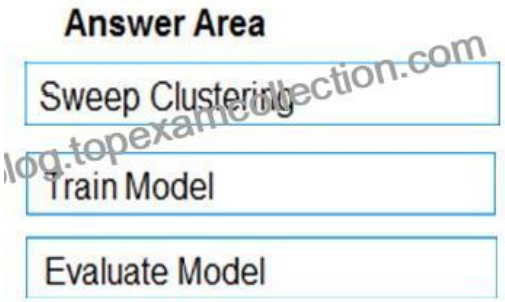
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/split-data

NO.135 You need to produce a visualization for the diagnostic test evaluation according to the data visualization requirements.

Which three modules should you recommend be used in sequence? To answer, move the appropriate modules from the list of modules to the answer area and arrange them in the correct order.



Explanation



Step 1: Sweep Clustering

Start by using the “Tune Model Hyperparameters” module to select the best sets of parameters for each of the models we’re considering.

One of the interesting things about the “Tune Model Hyperparameters” module is that it not only outputs the results from the Tuning, it also outputs the Trained Model.

Step 2: Train Model

Step 3: Evaluate Model

Scenario: You need to provide the test results to the Fabrikam Residences team. You create data visualizations to aid in presenting the results.

You must produce a Receiver Operating Characteristic (ROC) curve to conduct a diagnostic test evaluation of the model. You need to select appropriate methods for producing the ROC curve in Azure Machine Learning Studio to compare the Two-Class Decision Forest and the Two-Class Decision Jungle modules with one another.

References:

http://breaking-bi.blogspot.com/2017/01/azure-machine-learning-model-evaluation.html

NO.136 You need to implement a feature engineering strategy for the crowd sentiment local models.

What should you do?
* Apply an analysis of variance (ANOVA).
* Apply a Pearson correlation coefficient.
* Apply a Spearman correlation coefficient.
* Apply a linear discriminant analysis.

Explanation/Reference:

Explanation:

The linear discriminant analysis method works only on continuous variables, not categorical or ordinal variables.

Linear discriminant analysis is similar to analysis of variance (ANOVA) in that it works by comparing the means of the variables.

Scenario:

Data scientists must build notebooks in a local environment using automatic feature engineering and model building in machine learning pipelines.

Experiments for local crowd sentiment models must combine local penalty detection data.

All shared features for local models are continuous variables.

Incorrect Answers:

B: The Pearson correlation coefficient, sometimes called Pearson’s R test, is a statistical value that measures the linear relationship between two variables. By examining the coefficient values, you can infer something about the strength of the relationship between the two variables, and whether they are positively correlated or negatively correlated.

C: Spearman’s correlation coefficient is designed for use with non-parametric and non-normally distributed data. Spearman’s coefficient is a nonparametric measure of statistical dependence between two variables, and is sometimes denoted by the Greek letter rho. The Spearman’s coefficient expresses the degree to which two variables are monotonically related. It is also called Spearman rank correlation, because it can be used with ordinal variables.

References:

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/fisher-linear- discriminant-analysis

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/compute-linear- correlation Testlet 2 Case study Overview You are a data scientist for Fabrikam Residences, a company specializing in quality private and commercial property in the United States. Fabrikam Residences is considering expanding into Europe and has asked you to investigate prices for private residences in major European cities. You use Azure Machine Learning Studio to measure the median value of properties. You produce a regression model to predict property prices by using the Linear Regression and Bayesian Linear Regression modules.

Datasets

There are two datasets in CSV format that contain property details for two cities, London and Paris, with the following columns:

Column heading	Description
CapitaCrimeRate	per capita crime rate by town
Zoned	proportion of residential land zoned for lots over 25.000 square feet
NonRetailAcres	proportion of retail business acres per town
NextToRiver	proximity of the property to the river
NitrogenOxideConcentration	nitric oxides concentration (parts per 10 million)
AvgRoomsPerHouse	average number of rooms per dwelling
Age	proportion of owner-occupied units built prior to 1940
DistanceToEmploymentCenter	weighted distances to employment centers
AccessibilityToHighway	index of accessibility to radial highways to a value of two decimal places
Tax	full value property tax rate per $10,000
PupilTeacherRatio	pupil to teacher ratio by town
ProfessionalClass	professional class percentage
LowerStatus	percentage lower status of the population
MedianValue	median value of owner-occupied homes in $1000s

The two datasets have been added to Azure Machine Learning Studio as separate datasets and included as the starting point of the experiment.

Dataset issues

The AccessibilityToHighway column in both datasets contains missing values. The missing data must be replaced with new data so that it is modeled conditionally using the other variables in the data before filling in the missing values.

Columns in each dataset contain missing and null values. The dataset also contains many outliers. The Age column has a high proportion of outliers. You need to remove the rows that have outliers in the Age column. The MedianValue and AvgRoomsinHouse columns both hold data in numeric format. You need to select a feature selection algorithm to analyze the relationship between the two columns in more detail.

Model fit

The model shows signs of overfitting. You need to produce a more refined regression model that reduces the overfitting.

Experiment requirements

You must set up the experiment to cross-validate the Linear Regression and Bayesian Linear Regression modules to evaluate performance.

In each case, the predictor of the dataset is the column named MedianValue. An initial investigation showed that the datasets are identical in structure apart from the MedianValue column. The smaller Paris dataset contains the MedianValue in text format, whereas the larger London dataset contains the MedianValue in numerical format. You must ensure that the datatype of the MedianValue column of the Paris dataset matches the structure of the London dataset.

You must prioritize the columns of data for predicting the outcome. You must use non-parameters statistics to measure the relationships.

You must use a feature selection algorithm to analyze the relationship between the MedianValue and AvgRoomsinHouse columns.

Model training

Given a trained model and a test dataset, you need to compute the permutation feature importance scores of feature variables. You need to set up the Permutation Feature Importance module to select the correct metric to investigate the model’s accuracy and replicate the findings.

You want to configure hyperparameters in the model learning process to speed the learning phase by using hyperparameters. In addition, this configuration should cancel the lowest performing runs at each evaluation interval, thereby directing effort and resources towards models that are more likely to be successful.

You are concerned that the model might not efficiently use compute resources in hyperparameter tuning.

You also are concerned that the model might prevent an increase in the overall tuning time. Therefore, you need to implement an early stopping criterion on models that provides savings without terminating promising jobs.

Testing

You must produce multiple partitions of a dataset based on sampling using the Partition and Sample module in Azure Machine

Learning Studio. You must create three equal partitions for cross-validation. You must also configure the cross-validation process so that the rows in the test and training datasets are divided evenly by properties that are near each city’s main river. The data that identifies that a property is near a river is held in the column named NextToRiver. You want to complete this task before the data goes through the sampling process.

When you train a Linear Regression module using a property dataset that shows data for property prices for a large city, you need to determine the best features to use in a model. You can choose standard metrics provided to measure performance before and after the feature importance process completes. You must ensure that the distribution of the features across multiple training models is consistent.

Data visualization

You need to provide the test results to the Fabrikam Residences team. You create data visualizations to aid in presenting the results.

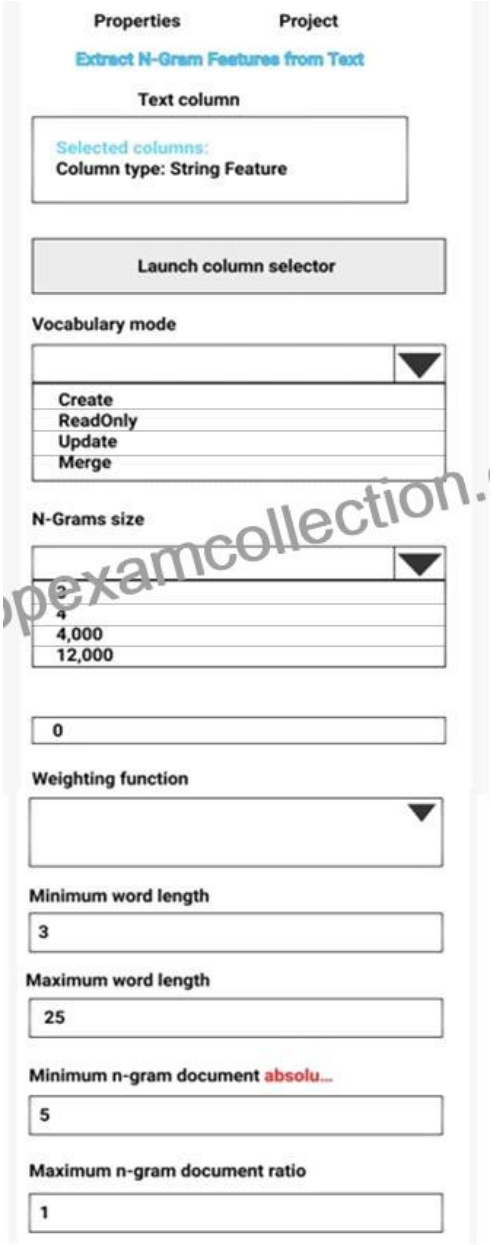
You must produce a Receiver Operating Characteristic (ROC) curve to conduct a diagnostic test evaluation of the model. You need to select appropriate methods for producing the ROC curve in Azure Machine Learning Studio to compare the Two-Class Decision Forest and the Two-Class Decision Jungle modules with one another.

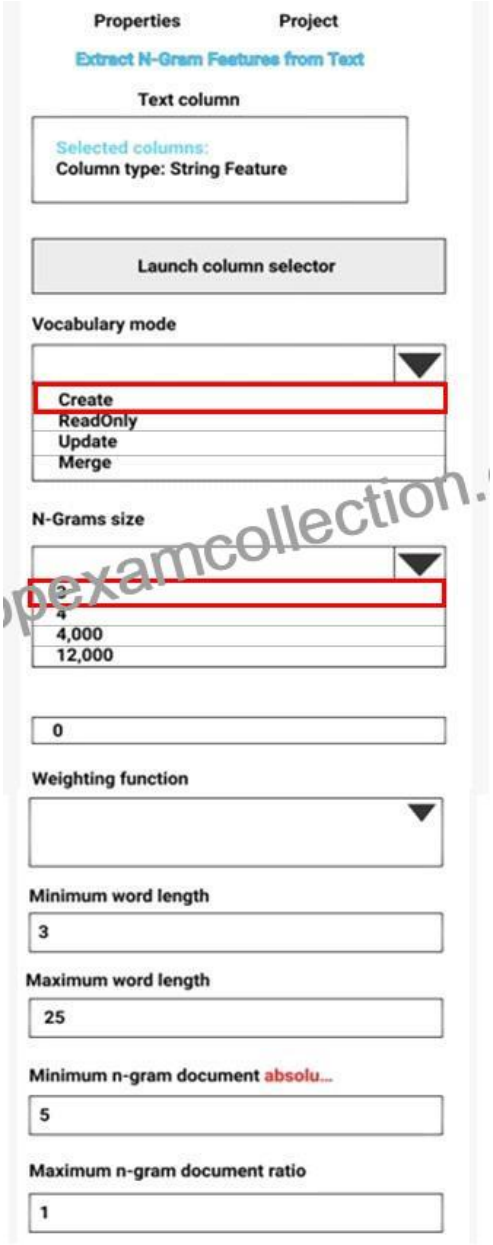
NO.137 You are performing sentiment analysis using a CSV file that includes 12,000 customer reviews written in a short sentence format. You add the CSV file to Azure Machine Learning Studio and configure it as the starting point dataset of an experiment. You add the Extract N-Gram Features from Text module to the experiment to extract key phrases from the customer review column in the dataset.

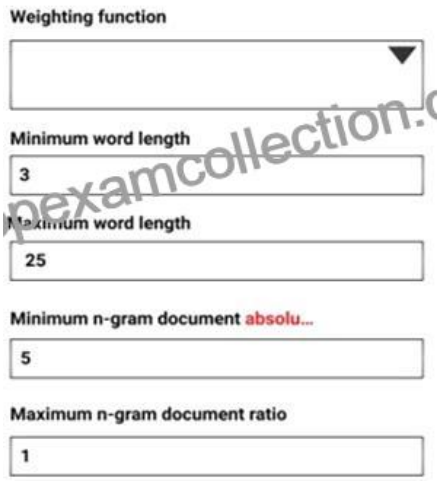
You must create a new n-gram dictionary from the customer review text and set the maximum n-gram size to trigrams.

What should you select? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.




Properties
Project
Extract N-Gram Features from Text
Text column
Selected columns:
Column type: String Feature
Launch column selector
Vocabulary mode
Create
ReadOnly
Update
Merge
N-Grams size
4
4,000
12,000
0
Weighting function
Minimum word length
3
Maximum word length
25
Minimum n-gram document absolu...
5
Maximum n-gram document ratio
1


Weighting function
Minimum word length
3
Maximum word length
25
Minimum n-gram document absolu...
5
Maximum n-gram document ratio
1

References:

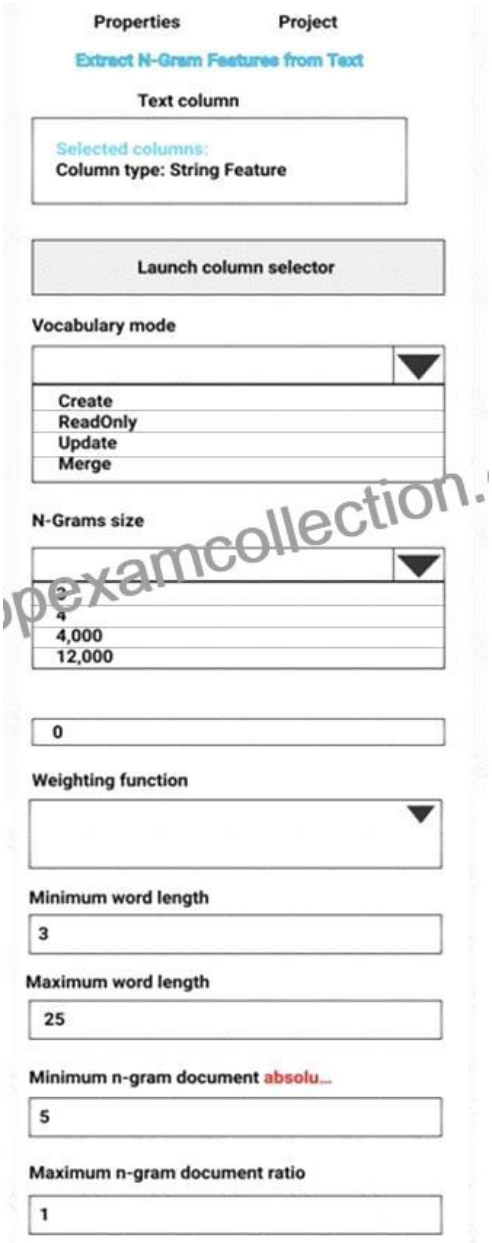
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/extract-n-gram-features-from-text

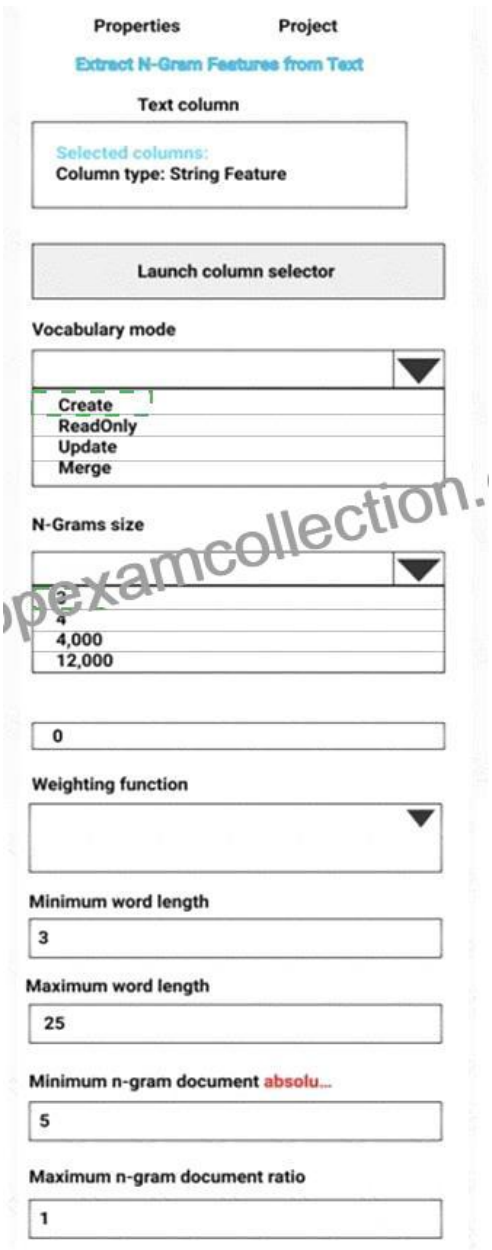
NO.138 You are performing sentiment analysis using a CSV file that includes 12,000 customer reviews written in a short sentence format. You add the CSV file to Azure Machine Learning Studio and configure it as the starting point dataset of an experiment. You add the Extract N-Gram Features from Text module to the experiment to extract key phrases from the customer review column in the dataset.

You must create a new n-gram dictionary from the customer review text and set the maximum n-gram size to trigrams.

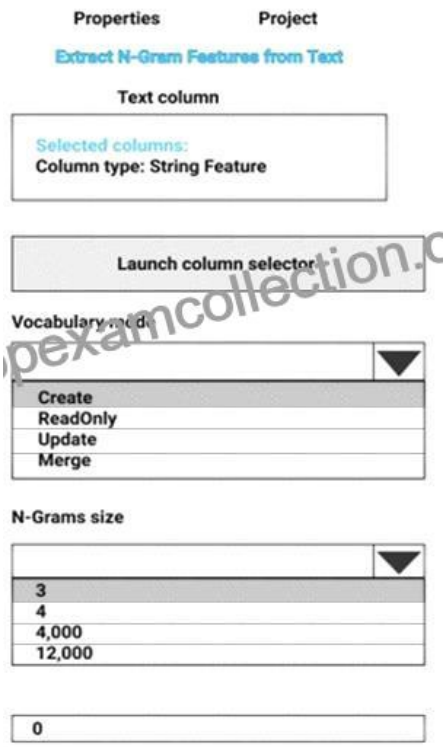
What should you select? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.





Explanation



Vocabulary mode: Create

For Vocabulary mode, select Create to indicate that you are creating a new list of n-gram features.

N-Grams size: 3

For N-Grams size, type a number that indicates the maximum size of the n-grams to extract and store. For example, if you type 3, unigrams, bigrams, and trigrams will be created.

Weighting function: Leave blank

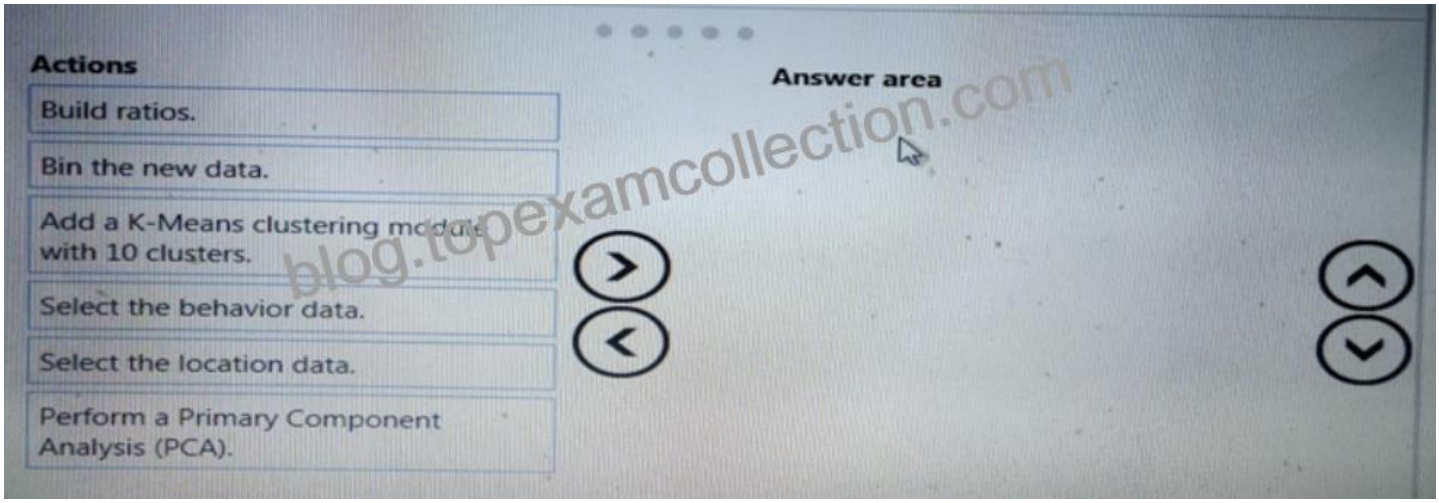
The option, Weighting function, is required only if you merge or update vocabularies. It specifies how terms in the two vocabularies and their scores should be weighted against each other.

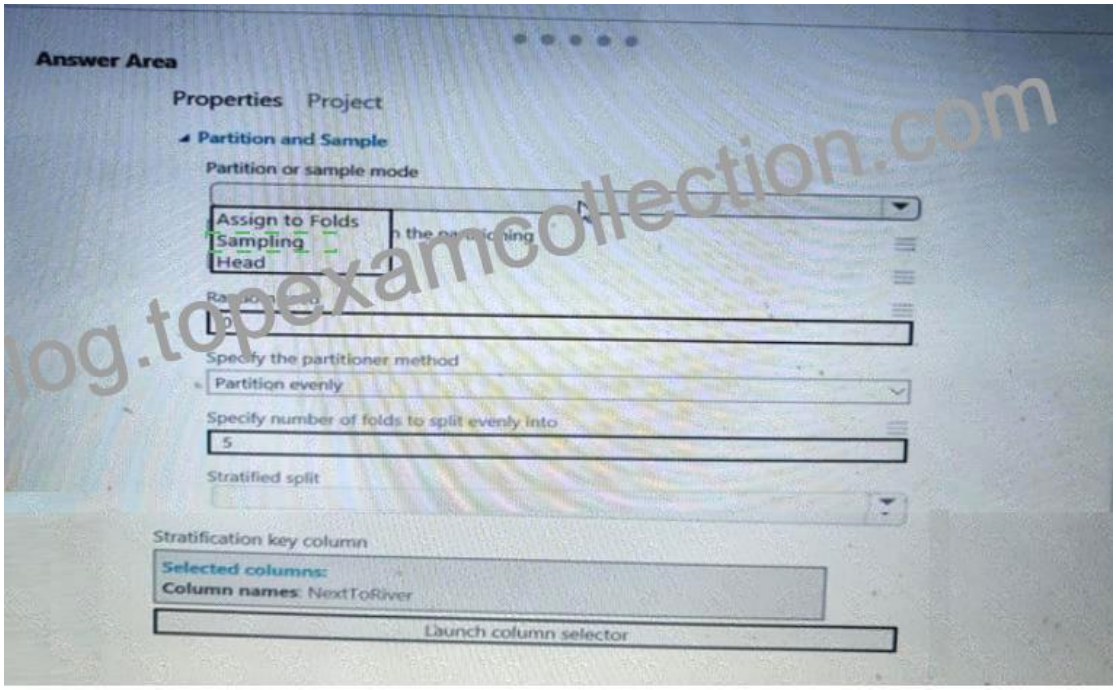
References:

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/extract-n-gram-features-from-

NO.139 You need to modify the inputs for the global penalty event model to address the bias and variance issue.

Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.





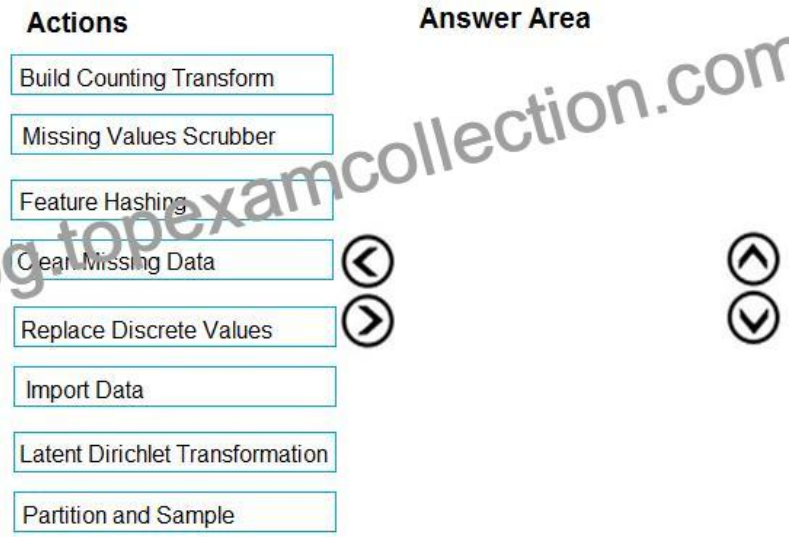
NO.140 You are creating an experiment by using Azure Machine Learning Studio.

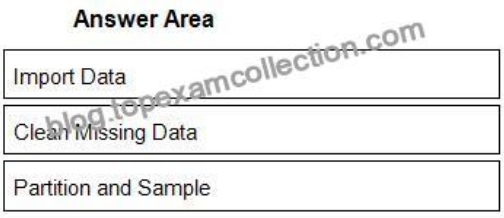
You must divide the data into four subsets for evaluation. There is a high degree of missing values in the dat a. You must prepare the data for analysis.

You need to select appropriate methods for producing the experiment.

Which three modules should you run in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.





1 – Import Data

2 – Clean Missing Data

3 – Partition and Sample

Reference:

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

NO.141 You are evaluating a completed binary classification machine learning model.

You need to use the precision as the valuation metric.

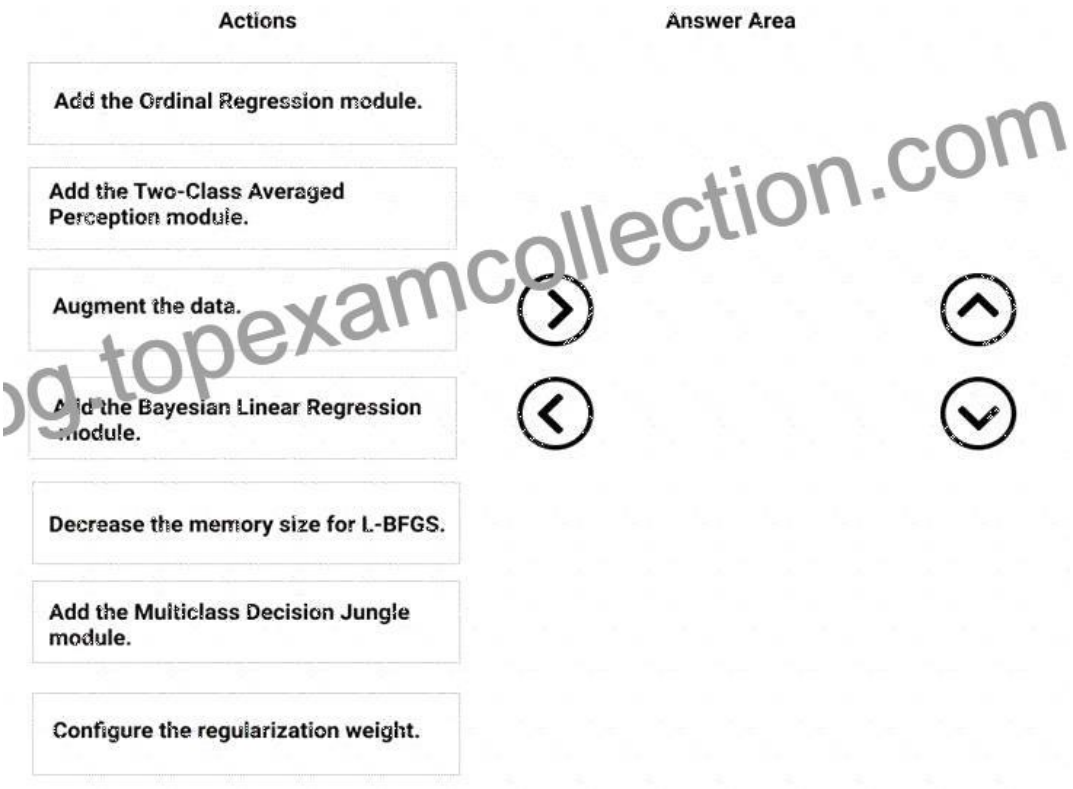
Which visualization should you use?
* Binary classification confusion matrix
* box plot
* Gradient descent
* coefficient of determination

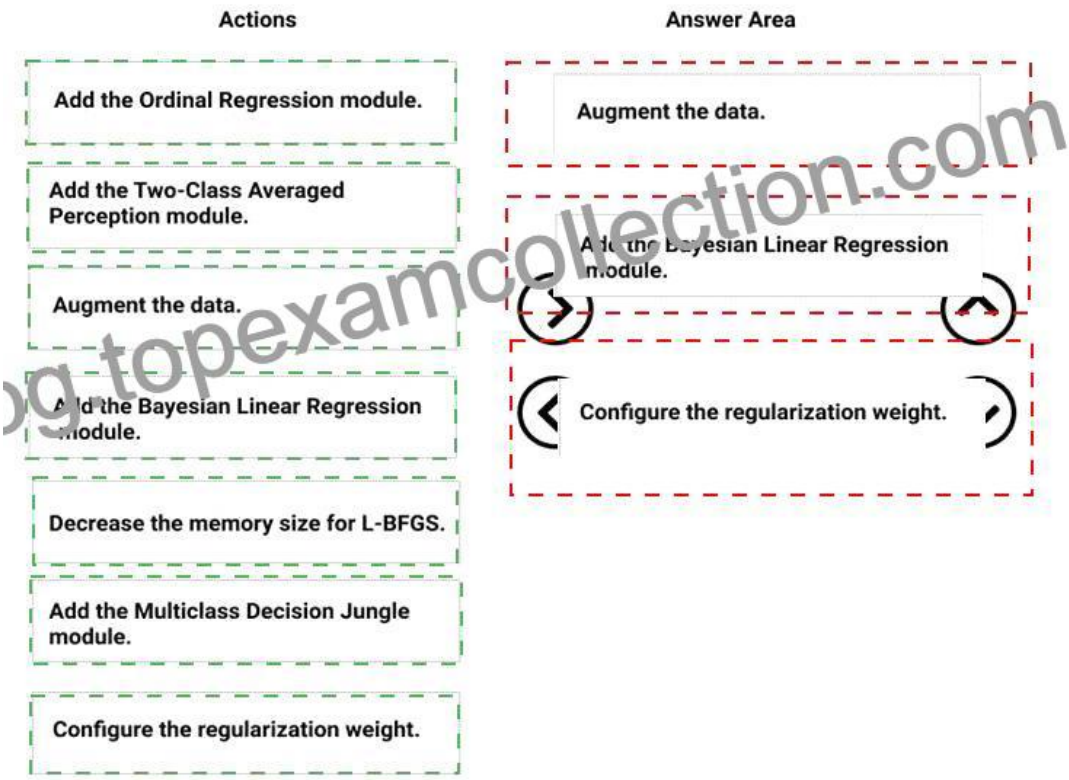
Reference:

https://machinelearningknowledge.ai/confusion-matrix-and-performance-metrics-machine-learning/

NO.142 You need to correct the model fit issue.

Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.





Explanation

Step 1: Augment the data

Scenario: Columns in each dataset contain missing and null values. The datasets also contain many outliers.

Step 2: Add the Bayesian Linear Regression module.

Scenario: You produce a regression model to predict property prices by using the Linear Regression and Bayesian Linear Regression modules.

Step 3: Configure the regularization weight.

Regularization typically is used to avoid overfitting. For example, in L2 regularization weight, type the value to use as the weight for L2 regularization. We recommend that you use a non-zero value to avoid overfitting.

Scenario:

Model fit: The model shows signs of overfitting. You need to produce a more refined regression model that reduces the overfitting.

NO.143 You need to use the Python language to build a sampling strategy for the global penalty detection models.

How should you complete the code segment? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Explanation

Box 1: import pytorch as deeplearninglib

Box 2: ..DistributedSampler(Sampler)..

DistributedSampler(Sampler):

Sampler that restricts data loading to a subset of the dataset.

It is especially useful in conjunction with class:`torch.nn.parallel.DistributedDataParallel`. In such case, each process can pass a DistributedSampler instance as a DataLoader sampler, and load a subset of the original dataset that is exclusive to it.

Scenario: Sampling must guarantee mutual and collective exclusively between local and global segmentation models that share the same features.

Box 3: optimizer = deeplearninglib.train. GradientDescentOptimizer(learning_rate=0.10)

NO.144 You plan to preprocess text from CSV files. You load the Azure Machine Learning Studio default stop words list.

You need to configure the Preprocess Text module to meet the following requirements:

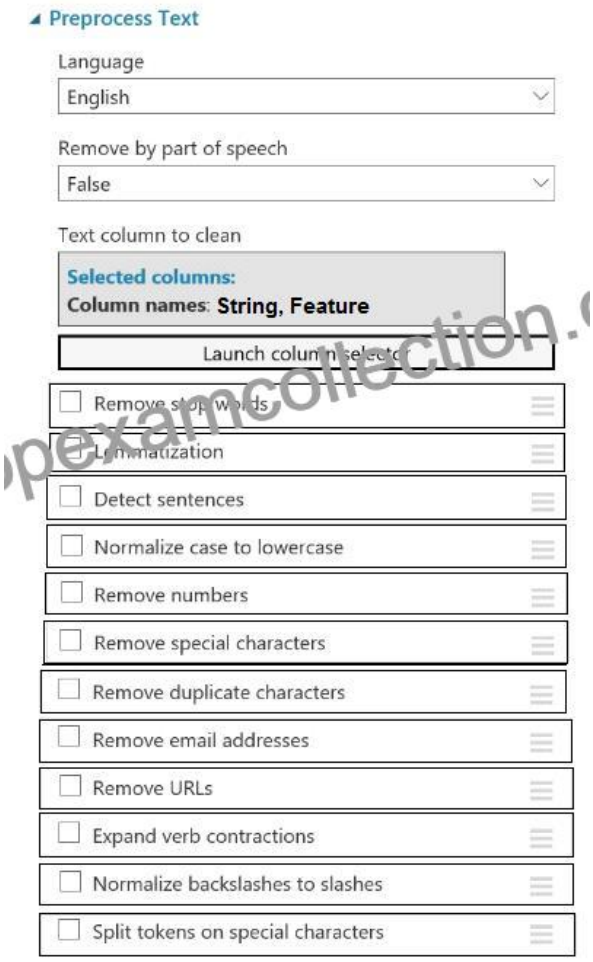
Ensure that multiple related words from a single canonical form.

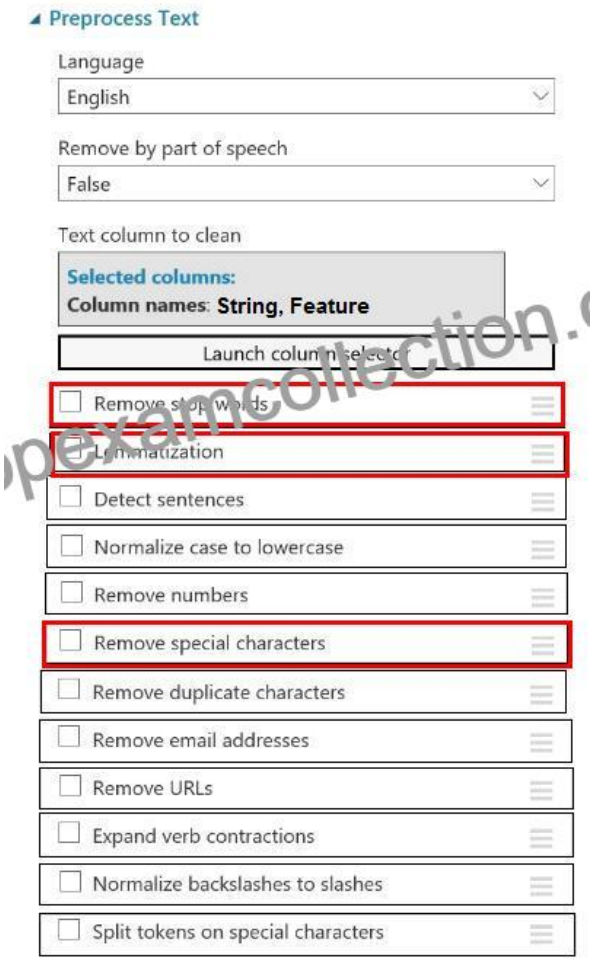
Remove pipe characters from text.

Remove words to optimize information retrieval.

Which three options should you select? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.



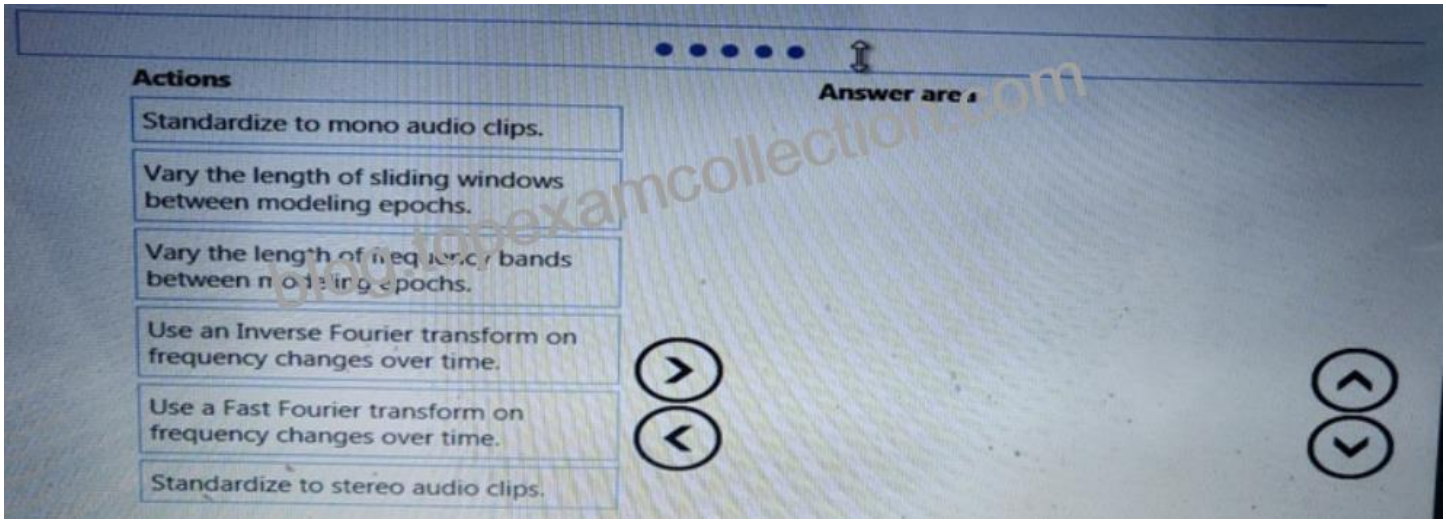


References:

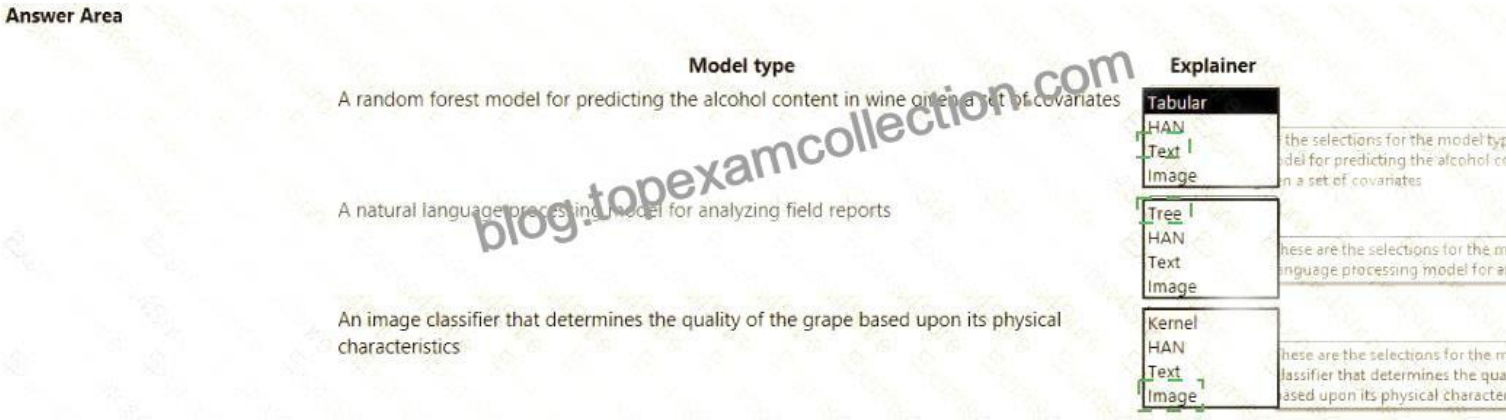
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/preprocess-text

NO.145 You need to define a process for penalty event detection.

Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



Answer Area



NO.146 Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

An IT department creates the following Azure resource groups and resources:

Resource group	Resources
ml_resources	• an Azure Machine Learning workspace named amlworkspace • an Azure Storage account named amlworkspace12345 • an Application Insights instance named amlworkspace54321 • an Azure Key Vault named amlworkspace67890 • an Azure Container Registry named amlworkspace09876
general_compute	A virtual machine named mlvm with the following configuration: • Operating system: Ubuntu Linux • Software installed: Python 3.6 and Jupyter Notebooks • Size: NC6 (6 vCPUs, 1 vGPU, 56 Gb RAM)

The IT department creates an Azure Kubernetes Service (AKS)-based inference compute target named aks-cluster in the Azure Machine Learning workspace. You have a Microsoft Surface Book computer with a GPU. Python 3.6 and Visual Studio Code are installed.

You need to run a script that trains a deep neural network (DNN) model and logs the loss and accuracy metrics.

Solution: Install the Azure ML SDK on the Surface Book. Run Python code to connect to the workspace. Run the training script as an experiment on the aks-cluster compute target.

Does the solution meet the goal?
* Yes
* No
Need to attach the mlvm virtual machine as a compute target in the Azure Machine Learning workspace.

Reference:

https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target

NO.147 You create a training pipeline by using the Azure Machine Learning designer. You need to load data into a machine learning pipeline by using the Import Data component. Which two data sources could you use? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point
* Azure Blob storage container through a registered datastore
* Azure SQL Database
* URL via HTTP
* Azure Data Lake Storage Gen2
* Registered dataset

NO.148 You have an Azure Machine Learning workspace. You are connecting an Azure Data Lake Storage Gen2 account to the workspace as a data store. You need to authorize access from the workspace to the Azure Data Lake Storage Gen2 account.

What should you use?
* Managed identity
* SAS token
* Service principal
* Account key

NO.149 You create an Azure Machine Learning workspace. You are training a classification model with no-code AutoML in Azure Machine Learning studio.

The model must predict if a client of a financial institution will subscribe to a fixed-term deposit. You must preview the data profile in Azure Machine Learning studio once the dataset is created.

You need to train the model.

Which four actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

Explanation

Updated Official licence for DP-100 Certified by DP-100 Dumps PDF:
https://www.topexamcollection.com/DP-100-vce-collection.html]