

TEST 5 - QUESTION: 1/50 SELECT MULTIPLE

You use the Azure Machine Learning SDK to run a training experiment that trains a classification model and calculates its accuracy metric. The model will be retrained each month as new data is available. You must register the model for use in a batch inference pipeline. You need to register the model and ensure that the models created by subsequent retraining experiments are registered only if their accuracy is higher than the currently registered model. What are two possible ways to achieve this goal? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

☐ A

Specify the model framework version when registering the model, and only register subsequent models if this value is higher.

☐ B

Specify a property named accuracy with the accuracy metric as a value when registering the model, and only register subsequent models if their accuracy is higher than the accuracy property value of the currently registered model.

☐ C

Specify a different name for the model each time you register it.

☐ D

Register the model with the same name each time regardless of accuracy, and always use the latest version of the model in the batch inferencing pipeline.

☐ E

Specify a tag named accuracy with the accuracy metric as a value when registering the model, and only register subsequent models if their accuracy is higher than the accuracy tag value of the currently registered model.

CORRECT ANSWERS: A,E

KEEP OPEN

EXPLANATION:

Explanation: E: Using tags, you can track useful information such as the name and version of the machine learning library used to train the model. Note that tags must be alphanumeric. Reference:

<https://notebooks.azure.com/xavierheriat/projects/azureml-getting-started/html/how-to-use-azureml/deployment/register-model-create-image-deploy-service/register-model-create-image-deploy-service.ipynb>

TEST 5 - QUESTION: 2/50 SELECT MULTIPLE

You use the Azure Machine Learning Python SDK to define a pipeline that consists of multiple steps. When you run the pipeline, you observe that some steps do not run. The cached output from a previous run is used instead. You need to ensure that every step in the pipeline is run, even if the parameters and contents of the source directory have not changed since the previous run. What are two possible ways to achieve this goal? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

☐ A

Restart the compute cluster where the pipeline experiment is configured to run.

☐ B

Use a PipelineData object that references a datastore other than the default datastore.

☐ C

Set the allow_reuse property of each step in the pipeline to False .

☐ D

Set the outputs property of each step in the pipeline to True .

☐ E

Set the regenerate_outputs property of the pipeline to True

CORRECT ANSWERS: C,E

KEEP OPEN

EXPLANATION:

Explanation: B: If regenerate_outputs is set to True, a new submit will always force generation of all step outputs, and disallow data reuse for any step of this run. Once this run is complete, however, subsequent runs may reuse the results of this run. C: Keep the following in mind when working with pipeline steps, input/output data, and step reuse. If data used in a step is in a datastore and allow_reuse is True, then changes to the data change won't be detected. If the data is uploaded as part of the snapshot (under the step's source_directory), though this is not recommended, then the hash will change and will trigger a rerun. Reference: <https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.pipelinesstep>
<https://github.com/Azure/MachineLearningNotebooks/blob/master/how-to-use-azureml/machine-learning-pipelines/intro-to-pipelines/aml-pipelines-getting-started.ipynb>

TEST 5 - QUESTION: 3/50

You are planning to register a trained model in an Azure Machine Learning workspace. You must store additional metadata about the model in a key-value format. You must be able to add new metadata and modify or delete metadata after creation. You need to register the model. Which parameter should you use?

☐ A
description

☐ B
model_framework

☐ C
properties

☐ D
tags

CORRECT ANSWER: C

KEEP OPEN

EXPLANATION:

Explanation: `azureml.core.Model.properties`: Dictionary of key value properties for the Model. These properties cannot be changed after registration, however new key value pairs can be added. Reference: <https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.model.model>

TEST 5 - QUESTION: 4/50

HOTSPOT You need to set up the Permutation Feature Importance module according to the model training requirements. Which properties should you select? To answer, select the appropriate options in the answer area. **NOTE:** Each correct selection is worth one point.

SCENARIO

Case study This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided. To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study. At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section. To start the case study To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

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. You add both files to Azure Machine Learning Studio as separate datasets to the starting point for an experiment. Both datasets contain 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 a 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

An initial investigation shows 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. Data issues Missing values 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 datasets also contain 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. 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-parametric statistics to measure relationships. You must use a feature selection algorithm to analyze the relationship between the MedianValue and AvgRoomsInHouse columns. Model training Permutation Feature Importance Given a trained model and a test dataset, you must compute the Permutation Feature Importance scores of feature variables. You must be determined the absolute fit for the model. Hyperparameters You must configure

hyperparameters in the model learning process to speed the learning phase . 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, must 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. Cross-validation 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. You must complete this task before the data goes through the sampling process. Linear regression module When you train a Linear Regression module, you must 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. The distribution of features across multiple training models must be 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.



CHECK BELOW THE RIGHT ANSWER

Answer Area

▲ Tune Model Hyperparameters

Specify parameter sweeping mode

Random sweep ▼

Maximum number of runs on random sweep

5

Random seed

0

Label column

Selected columns:
Column names: MedianValue

Launch column selector

Metric for measuring performance for classification

▼
F-score
Precision
Recall
Accuracy

Metric for measuring performance for regression

▼
Root of mean squared error
R-squared
Mean zero one error
Mean absolute error

CORRECT ANSWER:

KEEP OPEN

EXPLANATION:

Explanation: Box 1: Accuracy Scenario: 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. Box 2: R-Squared

TEST 5 - QUESTION: 5/50

You are creating a binary classification by using a two-class logistic regression model. You need to evaluate the model results for imbalance. Which evaluation metric should you use?

☐ A

AUC Curve

☐ B

Mean Absolute Error

☐ C

Accuracy

☐ D

Relative Absolute Error

☐ E

Root Mean Square Error

☐ F

Relative Squared Error

CORRECT ANSWER: A

KEEP OPEN

EXPLANATION:

Explanation: One can inspect the true positive rate vs. the false positive rate in the Receiver Operating Characteristic (ROC) curve and the corresponding Area Under the Curve (AUC) value. The closer this curve is to the upper left corner; the better the classifier's performance is (that is maximizing the true positive rate while minimizing the false positive rate). Curves that are close to the diagonal of the plot, result from classifiers that tend to make predictions that are close to random guessing. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/studio/evaluate-model-performance#evaluating-a-binary-classification-model>

TEST 5 - QUESTION: 6/50

HOTSPOT You use Azure Machine Learning to train and register a model. You must deploy the model into production as a real-time web service to an inference cluster named service-compute that the IT department has created in the Azure Machine Learning workspace. Client applications consuming the deployed web service must be authenticated based on their Azure Active Directory service principal. You need to write a script that uses the Azure Machine Learning SDK to deploy the model. The necessary modules have been imported. How should you complete the code? To answer, select the appropriate options in the answer area. **NOTE:** Each correct selection is worth one point.

☐ CHECK BELOW THE RIGHT ANSWER

Answer Area

```
# Assume the necessary modules have been imported
deploy_target =  .(ws, "service-compute")
               
               AksCompute
               AmlCompute
               RemoteCompute
               BatchCompute

deployment_config =  .deploy_configuration(cpu_cores=1, memory_gb=1,
               
               AksWebservice
               AciWebservice
               LocalWebService

 )

token_auth_enabled=True
token_auth_enabled=False
auth_enabled=True
auth_enabled=False

service = Model.deploy(ws, "ml-service",
    [model], inference_config, deployment_config, deploy_target)
service.wait_for_deployment(show_output = True)
```

CORRECT ANSWER:

KEEP OPEN

EXPLANATION:

Explanation: Box 1: AksCompute Example: aks_target = AksCompute(ws,"myaks") # If deploying to a cluster configured for dev/test, ensure that it was created with enough # cores and memory to handle this deployment configuration. Note that memory is also used by # things such as dependencies and AML components. deployment_config = AksWebservice.deploy_configuration(cpu_cores = 1, memory_gb = 1) service = Model.deploy(ws, "myservice", [model], inference_config, deployment_config, aks_target) Box 2: AksWebservice Box 3: token_auth_enabled=Yes Whether or not token auth is enabled for the Webservice. Note: A Service principal defined

in Azure Active Directory (Azure AD) can act as a principal on which authentication and authorization policies can be enforced in Azure Databricks. The Azure Active Directory Authentication Library (ADAL) can be used to programmatically get an Azure AD access token for a user. Incorrect Answers: auth_enabled (bool): Whether or not to enable key auth for this Webservice. Defaults to True. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-azure-kubernetes-service> <https://docs.microsoft.com/en-us/azure/databricks/dev-tools/api/latest/aad/service-prin-aad-token>

TEST 5 - QUESTION: 7/50

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. You are creating a model to predict the price of a student's artwork depending on the following variables: the student's length of education, degree type, and art form. You start by creating a linear regression model. You need to evaluate the linear regression model. Solution: Use the following metrics: Accuracy, Precision, Recall, F1 score, and AUC. Does the solution meet the goal?

☐ A

No

☐ B

Yes

CORRECT ANSWER: A

KEEP OPEN

EXPLANATION:

Explanation: Those are metrics for evaluating classification models, instead use: Mean Absolute Error, Root Mean Absolute Error, Relative Absolute Error, Relative Squared Error, and the Coefficient of Determination. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model>

TEST 5 - QUESTION: 8/50

You register a model that you plan to use in a batch inference pipeline. The batch inference pipeline must use a `ParallelRunStep` step to process files in a file dataset. The script has the `ParallelRunStep` step runs must process six input files each time the inferencing function is called. You need to configure the pipeline. Which configuration setting should you specify in the `ParallelRunConfig` object for the `ParallelRunStep` step?

☐ A
`process_count_per_node= "6"`

☐ B
`node_count= "6"`

☐ C
`mini_batch_size= "6"`

☐ D
`error_threshold= "6"`

CORRECT ANSWER: B

KEEP OPEN

EXPLANATION:

Explanation: `node_count` is the number of nodes in the compute target used for running the `ParallelRunStep`. Incorrect Answers: A: `process_count_per_node` Number of processes executed on each node. (optional, default value is number of cores on node.) C: `mini_batch_size` For `FileDataset` input, this field is the number of files user script can process in one `run()` call. For `TabularDataset` input, this field is the approximate size of data the user script can process in one `run()` call. Example values are 1024, 1024KB, 10MB, and 1GB. D: `error_threshold` The number of record failures for `TabularDataset` and file failures for `FileDataset` that should be ignored during processing. If the error count goes above this value, then the job will be aborted. Reference: <https://docs.microsoft.com/en-us/python/api/azureml-contrib-pipeline-steps/azureml.contrib.pipeline.steps.parallelrunconfig?view=azure-ml-py>

TEST 5 - QUESTION: 9/50

You need to select a feature extraction method. Which method should you use?

SCENARIO

Case study This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided. To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study. At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section. To start the case study To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

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. You add both files to Azure Machine Learning Studio as separate datasets to the starting point for an experiment. Both datasets contain 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 a 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

An initial investigation shows 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. Data issues Missing values 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 datasets also contain 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. 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-parametric statistics to measure relationships. You must use a feature selection algorithm to analyze the relationship between the MedianValue and AvgRoomsInHouse columns. Model training Permutation Feature Importance Given a trained model and a test dataset, you must compute the Permutation Feature Importance scores of feature variables. You must be determined the absolute fit for the model. Hyperparameters You must configure

hyperparameters in the model learning process to speed the learning phase . 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, must 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. Cross-validation 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. You must complete this task before the data goes through the sampling process. Linear regression module When you train a Linear Regression module, you must 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. The distribution of features across multiple training models must be 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.

☐

A

Mood's median test

☐

B

Mutual information

☐

C

Permutation Feature Importance

☐

D

Kendall correlation

CORRECT ANSWER: D

KEEP OPEN

EXPLANATION:

Explanation: In statistics, the Kendall rank correlation coefficient, commonly referred to as Kendall's tau coefficient (after the Greek letter γ), is a statistic used to measure the ordinal association between two measured quantities. It is a supported method of the Azure Machine Learning Feature selection. Note: Both Spearman's and Kendall's can be formulated as special cases of a more general correlation coefficient, and they are both appropriate in this scenario. Scenario: 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. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/feature-selection-modules>

TEST 5 - QUESTION: 10/50

You are a data scientist working for a hotel booking website company. You use the Azure Machine Learning service to train a model that identifies fraudulent transactions. You must deploy the model as an Azure Machine Learning real-time web service using the `Model.deploy` method in the Azure Machine Learning SDK. The deployed web service must return real-time predictions of fraud based on transaction data input. You need to create the script that is specified as the `entry_script` parameter for the `InferenceConfig` class used to deploy the model. What should the entry script do?

☐ A

Start a node on the inference cluster where the web service is deployed.

☐ B

Load the model and use it to predict labels from input data.

☐ C

Register the model with appropriate tags and properties.

☐ D

Create a Conda environment for the web service compute and install the necessary Python packages.

☐ E

Specify the number of cores and the amount of memory required for the inference compute.

CORRECT ANSWER: B

KEEP OPEN

EXPLANATION:

Explanation: The entry script receives data submitted to a deployed web service and passes it to the model. It then takes the response returned by the model and returns that to the client. The script is specific to your model. It must understand the data that the model expects and returns. The two things you need to accomplish in your entry script are: Loading your model (using a function called `init()`) Running your model on input data (using a function called `run()`) Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-and-where>

TEST 5 - QUESTION: 11/50

You use the Azure Machine Learning designer to create and run a training pipeline. You then create a real-time inference pipeline. You must deploy the real-time inference pipeline as a web service. What must you do before you deploy the real-time inference pipeline?

☐ A

Create a batch inference pipeline.

☐ B

Create an Azure Machine Learning compute cluster.

☐ C

Run the real-time inference pipeline.

☐ D

Clone the training pipeline.

CORRECT ANSWER: B

KEEP OPEN

EXPLANATION:

Explanation: You need to create an inferencing cluster. Deploy the real-time endpoint After your AKS service has finished provisioning, return to the real-time inferencing pipeline to complete deployment. Select Deploy above the canvas. Select Deploy new real-time endpoint. Select the AKS cluster you created. Select Deploy. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-designer-automobile-price-deploy>

TEST 5 - QUESTION: 12/50

You create a deep learning model for image recognition on Azure Machine Learning service using GPU-based training. You must deploy the model to a context that allows for real-time GPU-based inferencing. You need to configure compute resources for model inferencing. Which compute type should you use?

- ☐ A
Azure Container Instance
- ☐ B
Machine Learning Compute
- ☐ C
Field Programmable Gate Array
- ☐ D
Azure Kubernetes Service

CORRECT ANSWER: D

KEEP OPEN

EXPLANATION:

Explanation: You can use Azure Machine Learning to deploy a GPU-enabled model as a web service. Deploying a model on Azure Kubernetes Service (AKS) is one option. The AKS cluster provides a GPU resource that is used by the model for inference. Inference, or model scoring, is the phase where the deployed model is used to make predictions. Using GPUs instead of CPUs offers performance advantages on highly parallelizable computation. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-inferencing-gpus>

TEST 5 - QUESTION: 13/50

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. You train and register a machine learning model. You plan to deploy the model as a real-time web service. Applications must use key-based authentication to use the model. You need to deploy the web service. Solution: Create an AciWebService instance. Set the value of the auth_enabled property to False . Set the value of the token_auth_enabled property to True . Deploy the model to the service. Does the solution meet the goal?

☐ A

Yes

☐ B

No

CORRECT ANSWER: B

KEEP OPEN

EXPLANATION:

Explanation: Instead use only auth_enabled = TRUE Note: Key-based authentication. Web services deployed on AKS have key-based auth enabled by default. ACI-deployed services have key-based auth disabled by default, but you can enable it by setting auth_enabled = TRUE when creating the ACI web service. The following is an example of creating an ACI deployment configuration with key-based auth enabled. deployment_config <- aci_web_service_deployment_config(cpu_cores = 1, memory_gb = 1, auth_enabled = TRUE) Reference: <https://azure.github.io/azureml-sdk-for-r/articles/deploying-models.html>

TEST 5 - QUESTION: 14/50

You are a data scientist working for a bank and have used Azure ML to train and register a machine learning model that predicts whether a customer is likely to repay a loan. You want to understand how your model is making selections and must be sure that the model does not violate government regulations such as denying loans based on where an applicant lives. You need to determine the extent to which each feature in the customer data is influencing predictions. What should you do?

☐ A

Add tags to the model registration indicating the names of the features in the training dataset.

☐ B

Use the Hyperdrive library to test the model with multiple hyperparameter values.

☐ C

Enable data drift monitoring for the model and its training dataset.

☐ D

Use the interpretability package to generate an explainer for the model.

☐ E

Score the model against some test data with known label values and use the results to calculate a confusion matrix.

CORRECT ANSWER: D

KEEP OPEN

EXPLANATION:

Explanation: When you compute model explanations and visualize them, you're not limited to an existing model explanation for an automated ML model. You can also get an explanation for your model with different test data. The steps in this section show you how to compute and visualize engineered feature importance based on your test data. Incorrect Answers: A: In the context of machine learning, data drift is the change in model input data that leads to model performance degradation. It is one of the top reasons where model accuracy degrades over time, thus monitoring data drift helps detect model performance issues. B: A confusion matrix is used to describe the performance of a classification model. Each row displays the instances of the true, or actual class in your dataset, and each column represents the instances of the class that was predicted by the model. C: Hyperparameters are adjustable parameters you choose for model training that guide the training process. The HyperDrive package helps you automate choosing these parameters. Reference:

TEST 5 - QUESTION: 15/50

HOTSPOT You are a lead data scientist for a project that tracks the health and migration of birds. You create a multi-image classification deep learning model that uses a set of labeled bird photos collected by experts. You plan to use the model to develop a cross-platform mobile app that predicts the species of bird captured by app users. You must test and deploy the trained model as a web service. The deployed model must meet the following requirements: An authenticated connection must not be required for testing. The deployed model must perform with low latency during inferencing. The REST endpoints must be scalable and should have a capacity to handle large number of requests when multiple end users are using the mobile application. You need to verify that the web service returns predictions in the expected JSON format when a valid REST request is submitted. Which compute resources should you use? To answer, select the appropriate options in the answer area. **NOTE :** Each correct selection is worth one point.

☐ CHECK BELOW THE RIGHT ANSWER

Answer Area

Context	Resource
Test	<div><div></div><div>ds-workstation notebook VM</div><div>aks-compute cluster</div><div>cpu-compute cluster</div><div>gpu-compute cluster</div></div>
Production	<div><div></div><div>ds-workstation notebook VM</div><div>aks-compute cluster</div><div>cpu-compute cluster</div><div>gpu-compute cluster</div></div>

CORRECT ANSWER:

KEEP OPEN

EXPLANATION:

Explanation: Box 1: ds-workstation notebook VM An authenticated connection must not be required for testing. On a Microsoft Azure virtual machine (VM), including a Data Science Virtual Machine (DSVM), you create local user accounts while provisioning the VM. Users then authenticate to the VM by using these credentials. Box 2: gpu-compute cluster Image classification is well suited for GPU compute clusters Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/dsvm-common-identity> <https://docs.microsoft.com/en-us/azure/architecture/reference-architectures/ai/training-deep-learning>

TEST 5 - QUESTION: 16/50

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. You train a classification model by using a logistic regression algorithm. You must be able to explain the model's predictions by calculating the importance of each feature, both as an overall global relative importance value and as a measure of local importance for a specific set of predictions. You need to create an explainer that you can use to retrieve the required global and local feature importance values. Solution: Create a TabularExplainer. Does the solution meet the goal?

☐ A

No

☐ B

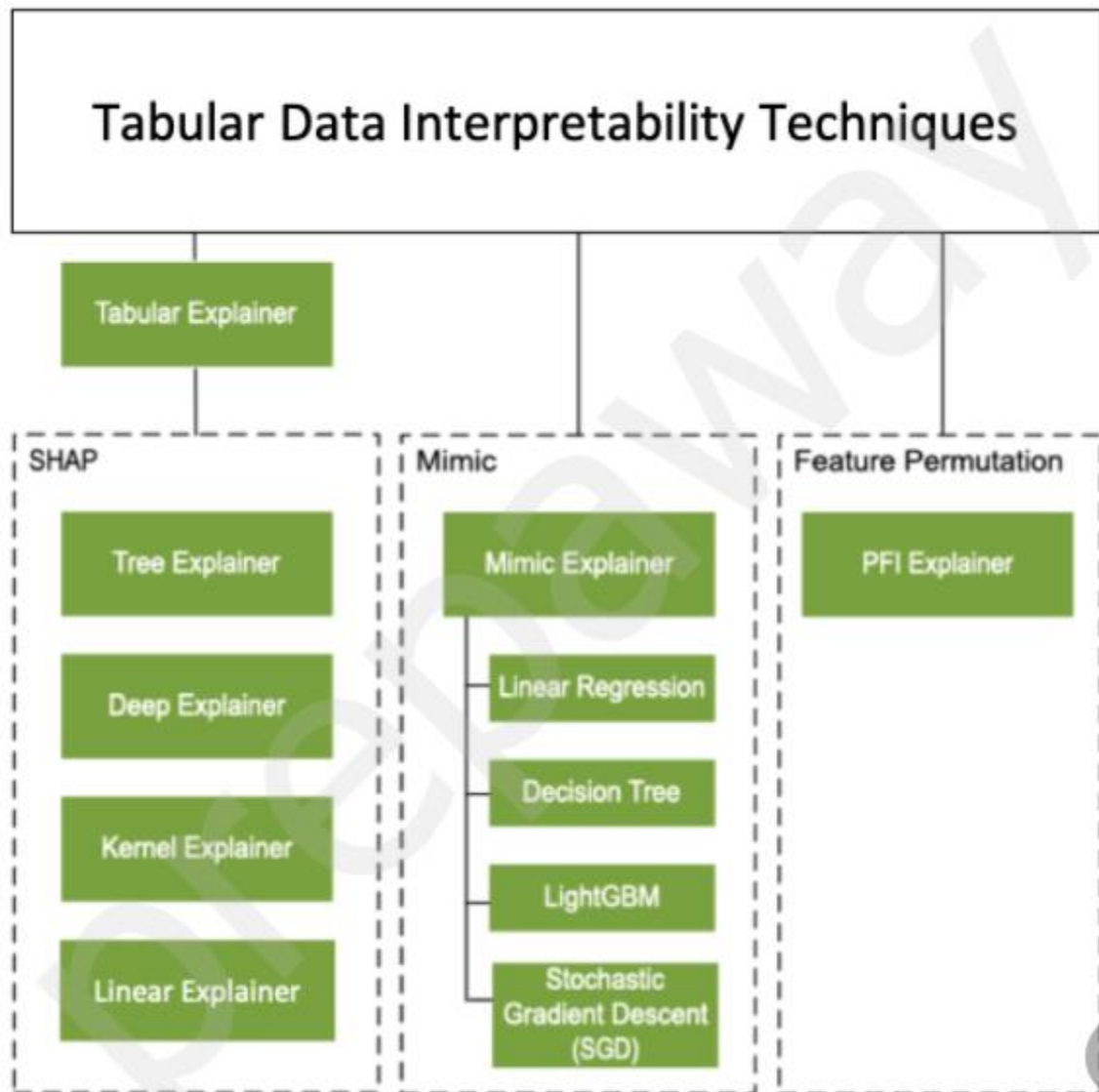
Yes

CORRECT ANSWER: A

KEEP OPEN

EXPLANATION:

Explanation: Instead use Permutation Feature Importance Explainer (PFI). Note 1:



Note 2: Permutation Feature Importance Explainer (PFI): Permutation Feature Importance is a technique used to explain classification and regression models. At a high level, the way it works is by randomly shuffling data one feature at a time for the entire dataset and calculating how much the performance metric of interest changes. The larger the change, the more important that feature is. PFI can explain the overall behavior of any underlying model but does not explain individual predictions. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability>

TEST 5 - QUESTION: 17/50

HOTSPOT You create an Azure Machine Learning workspace. You need to detect data drift between a baseline dataset and a subsequent target dataset by using the DataDriftDetector class. 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.

☐ CHECK BELOW THE RIGHT ANSWER

Answer Area

```
from azureml.core import Workspace, Dataset
from datetime import datetime
```

```
ws = Workspace.from_config()
dset = Dataset.get_by_name(ws, 'target')
baseline = target.time_before(datetime(2021, 2, 1))
features = ['windAngle', 'windSpeed', 'temperature', 'stationName']
```

```
monitor = DataDriftDetector. (ws, 'drift-monitor', baseline,
                             backfill
                             create_from_datasets
                             create_from_model
                             target, compute_target='cpu-cluster', frequency='Week', feature_list=None,
                             drift_threshold=.6, latency=24)
```

```
monitor = DataDriftDetector.get_by_name(ws, 'drift-monitor')
monitor = monitor.update(feature_list=features)
complete = monitor. (datetime(2021, 1, 1), datetime.today())
```

backfill
list
update

CORRECT ANSWER:

KEEP OPEN

EXPLANATION:

Explanation: Box 1: create_from_datasets The create_from_datasets method creates a new DataDriftDetector object from a baseline tabular dataset and a target time series dataset. Box 2: backfill The backfill method runs a backfill job over a given specified start and end date. Syntax: backfill(start_date, end_date, compute_target=None, create_compute_target=False) Incorrect Answers: List and update do not have datetime parameters. Reference: [https://docs.microsoft.com/en-us/python/api/azureml-datadrift/azureml.datadrift.datadriftdetector\(class\)](https://docs.microsoft.com/en-us/python/api/azureml-datadrift/azureml.datadrift.datadriftdetector(class))

TEST 5 - QUESTION: 18/50

You train and register a machine learning model. You create a batch inference pipeline that uses the model to generate predictions from multiple data files. You must publish the batch inference pipeline as a service that can be scheduled to run every night. You need to select an appropriate compute target for the inference service. Which compute target should you use?

☐ A

Azure Machine Learning compute instance

☐ B

Azure Machine Learning compute cluster

☐ C

Azure Container Instance (ACI) compute target

☐ D

Azure Kubernetes Service (AKS)-based inference cluster

CORRECT ANSWER: B

KEEP OPEN

EXPLANATION:

Explanation: Azure Machine Learning compute clusters is used for Batch inference. Run batch scoring on serverless compute. Supports normal and low-priority VMs. No support for real-time inference. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target>

TEST 5 - QUESTION: 19/50

HOTSPOT You need to configure the Permutation Feature Importance module for the model training requirements. What should you do? To answer, select the appropriate options in the dialog box in the answer area. **NOTE:** Each correct selection is worth one point.

SCENARIO

Case study This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided. To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study. At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section. To start the case study To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

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. You add both files to Azure Machine Learning Studio as separate datasets to the starting point for an experiment. Both datasets contain 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 a 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

An initial investigation shows 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. Data issues Missing values 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 datasets also contain 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. 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-parametric statistics to measure relationships. You must use a feature selection algorithm to analyze the relationship between the MedianValue and AvgRoomsInHouse columns. Model training Permutation Feature Importance Given a trained model and a test dataset, you must compute the Permutation Feature Importance scores of feature variables. You must be determined the absolute fit for the model. Hyperparameters You must configure

hyperparameters in the model learning process to speed the learning phase . 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, must 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. Cross-validation 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. You must complete this task before the data goes through the sampling process. Linear regression module When you train a Linear Regression module, you must 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. The distribution of features across multiple training models must be 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.



CHECK BELOW THE RIGHT ANSWER

Answer Area

Permutation Feature importance

Random seed

	▼
0 500	
	▼
Regression – Root Mean Square Error Regression – R-squared Regression – Mean Zero One Error Regression – Mean Absolute Error	

CORRECT ANSWER:

KEEP OPEN

EXPLANATION:

Explanation: Box 1: 500 For Random seed, type a value to use as seed for randomization. If you specify 0 (the default), a number is generated based on the system clock. A seed value is optional, but you should provide a value if you want reproducibility across runs of the same experiment. Here we must replicate the findings. Box 2: Mean Absolute Error Scenario: Given a trained model and a test dataset, you must 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. Regression. Choose one of the following: Precision, Recall, Mean Absolute Error, Root Mean Squared Error, Relative Absolute Error, Relative Squared Error, Coefficient of Determination Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/permutation-feature-importance>

TEST 5 - QUESTION: 20/50

You deploy a real-time inference service for a trained model. The deployed model supports a business-critical application, and it is important to be able to monitor the data submitted to the web service and the predictions the data generates. You need to implement a monitoring solution for the deployed model using minimal administrative effort. What should you do?

☐ A

View the log files generated by the experiment used to train the model.

☐ B

Create an ML Flow tracking URI that references the endpoint, and view the data logged by ML Flow.

☐ C

View the explanations for the registered model in Azure ML studio.

☐ D

Enable Azure Application Insights for the service endpoint and view logged data in the Azure portal.

CORRECT ANSWER: D

KEEP OPEN

EXPLANATION:

Explanation: Configure logging with Azure Machine Learning studio You can also enable Azure Application Insights from Azure Machine Learning studio. When you're ready to deploy your model as a web service, use the following steps to enable Application Insights: Sign in to the studio at <https://ml.azure.com>. Go to Models and select the model you want to deploy. Select +Deploy. Populate the Deploy model form. Expand the Advanced menu. Select Enable Application Insights diagnostics and data collection.

▼ Advanced

Enable Application Insights diagnostics and data collection
☒ Enable Application Insights diagnostics and data collection

Enable SSL
☐ Enable SSL

Max concurrent requests per container

CPU reserve capacity ⓘ

Memory reserve capacity ⓘ

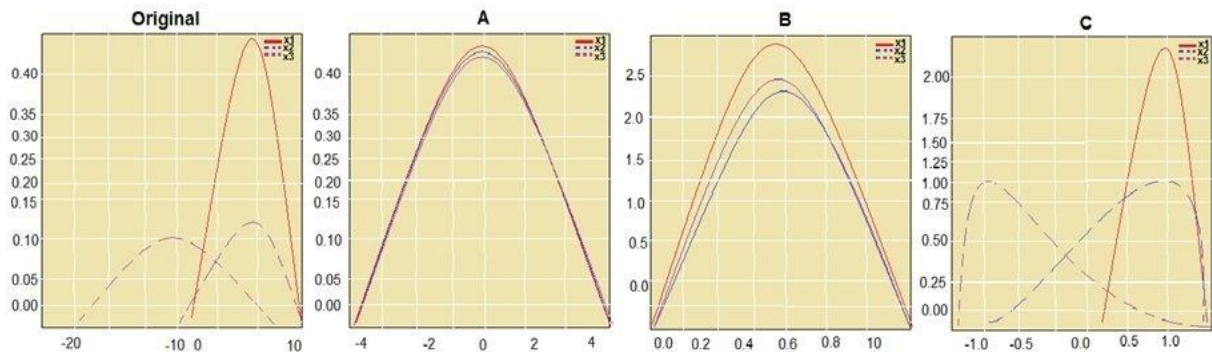
Deploy Cancel

praw343167

Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-enable-app-insights>

TEST 5 - QUESTION: 21/50

HOTSPOT You are performing feature scaling by using the scikit-learn Python library for x_1 , x_2 , and x_3 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.

☐ CHECK BELOW THE RIGHT ANSWER

Answer Area

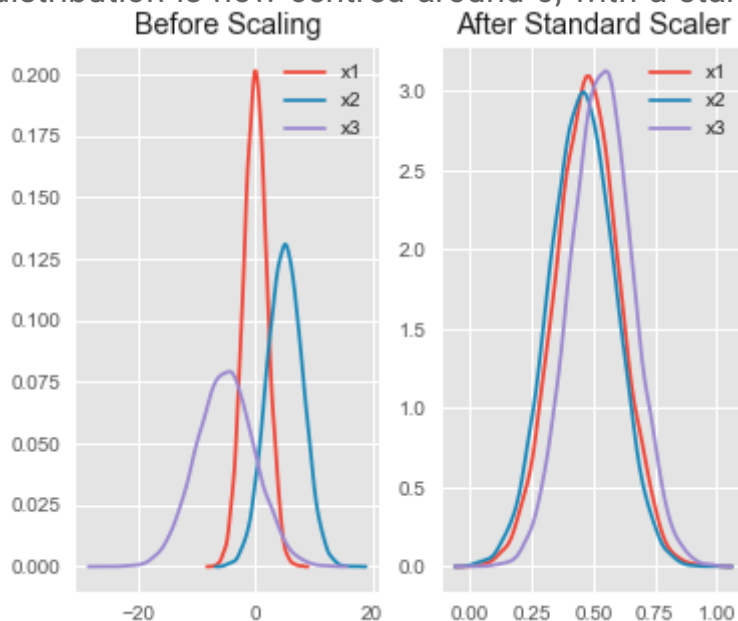
Question	Answer choice
Which scaler is used in graph A?	<div>Standard Scaler</div> <div>Min Max Scale</div> <div>Normalizer</div>
Which scaler is used in graph B?	<div>Standard Scaler</div> <div>Min Max Scale</div> <div>Normalizer</div>
Which scaler is used in graph C?	<div>Standard Scaler</div> <div>Min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

CORRECT ANSWER:

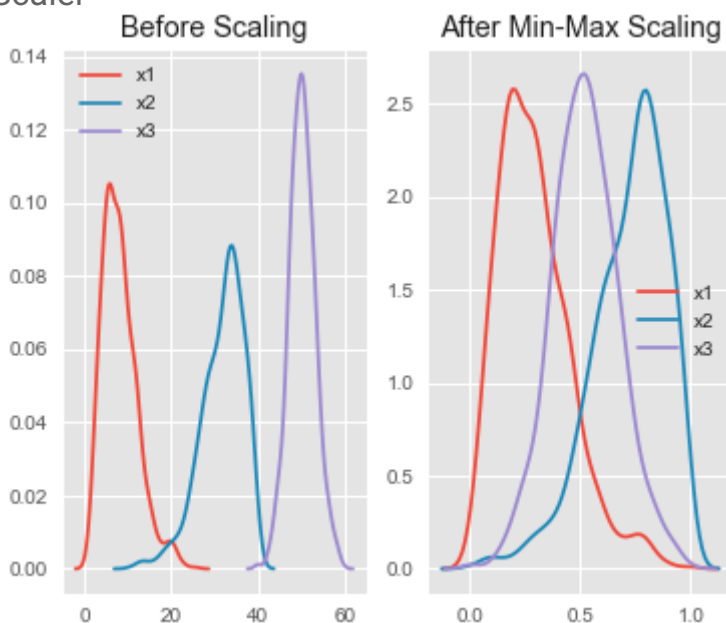
KEEP OPEN

EXPLANATION:

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

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

TEST 5 - QUESTION: 22/50

You are determining if two sets of data are significantly different from one another by using Azure Machine Learning Studio. Estimated values in one set of data may be more than or less than reference values in the other set of data. You must produce a distribution that has a constant Type I error as a function of the correlation. You need to produce the distribution. Which type of distribution should you produce?



A

Unpaired t-test with a two-tail option



B

Unpaired t-test with a one-tail option



C

Paired t-test with a one-tail option



D

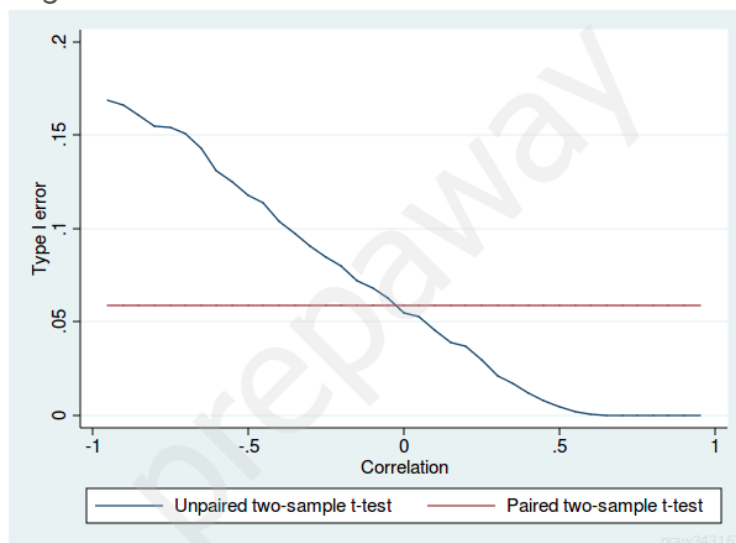
Paired t-test with a two-tail option

CORRECT ANSWER: D

KEEP OPEN

EXPLANATION:

Explanation: Choose a one-tail or two-tail test. The default is a two-tailed test. This is the most common type of test, in which the expected distribution is symmetric around zero. Example: Type I error of unpaired and paired two-sample t-tests as a function of the correlation. The simulated random numbers originate from a bivariate normal distribution with a variance of 1.



Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/test-hypothesis-using-t-test>
https://en.wikipedia.org/wiki/Student%27s_t-test

TEST 5 - QUESTION: 23/50

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. You train and register a machine learning model. You plan to deploy the model as a real-time web service. Applications must use key-based authentication to use the model. You need to deploy the web service. Solution: Create an AciWebservice instance. Set the value of the ssl_enabled property to True . Deploy the model to the service. Does the solution meet the goal?

☐ A

Yes

☐ B

No

CORRECT ANSWER: B

KEEP OPEN

EXPLANATION:

Explanation: Instead use only `auth_enabled = TRUE` Note: Key-based authentication. Web services deployed on AKS have key-based auth enabled by default. ACI-deployed services have key-based auth disabled by default, but you can enable it by setting `auth_enabled = TRUE` when creating the ACI web service. The following is an example of creating an ACI deployment configuration with key-based auth enabled.

```
deployment_config <-  
aci_webservice_deployment_config(cpu_cores = 1, memory_gb = 1,  
auth_enabled = TRUE)
```

Reference: <https://azure.github.io/azureml-sdk-for-r/articles/deploying-models.html>

TEST 5 - QUESTION: 24/50 SELECT MULTIPLE

You train and register a model in your Azure Machine Learning workspace. You must publish a pipeline that enables client applications to use the model for batch inferencing. You must use a pipeline with a single ParallelRunStep step that runs a Python inferencing script to get predictions from the input data. You need to create the inferencing script for the ParallelRunStep pipeline step. Which two functions should you include? Each correct answer presents part of the solution. NOTE : Each correct selection is worth one point.

☐ A`init()`☐ B`batch()`☐ C`score(mini_batch)`☐ D`run(mini_batch)`☐ E`main()`**CORRECT ANSWERS: A,D**

KEEP OPEN

EXPLANATION:

Reference:

<https://github.com/Azure/MachineLearningNotebooks/tree/master/how-to-use-azureml/machine-learning-pipelines/parallel-run>

TEST 5 - QUESTION: 25/50

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. You train a classification model by using a logistic regression algorithm. You must be able to explain the model's predictions by calculating the importance of each feature, both as an overall global relative importance value and as a measure of local importance for a specific set of predictions. You need to create an explainer that you can use to retrieve the required global and local feature importance values. Solution: Create a PFIE explainer. Does the solution meet the goal?

☐ A

Yes

☐ B

No

CORRECT ANSWER: A

KEEP OPEN

EXPLANATION:

Explanation: Permutation Feature Importance Explainer (PFI): Permutation Feature Importance is a technique used to explain classification and regression models. At a high level, the way it works is by randomly shuffling data one feature at a time for the entire dataset and calculating how much the performance metric of interest changes. The larger the change, the more important that feature is. PFI can explain the overall behavior of any underlying model but does not explain individual predictions. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability>

TEST 5 - QUESTION: 26/50

You create a multi-class image classification deep learning model. You train the model by using PyTorch version 1.2. You need to ensure that the correct version of PyTorch can be identified for the inferencing environment when the model is deployed. What should you do?

☐ A

Deploy the model on computer that is configured to use the default Azure Machine Learning conda environment.

☐ B

Save the model locally as a .pt file, and deploy the model as a local web service.

☐ C

Register the model with a .pt file extension and the default version property.

☐ D

Register the model, specifying the model_framework and model_framework_version properties.

CORRECT ANSWER: D

KEEP OPEN

EXPLANATION:

Explanation: framework_version: The PyTorch version to be used for executing training code. Reference: <https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.dnn.pytorch?view=azure-ml-py>

TEST 5 - QUESTION: 27/50

You train a machine learning model. You must deploy the model as a real-time inference service for testing. The service requires low CPU utilization and less than 48 MB of RAM. The compute target for the deployed service must initialize automatically while minimizing cost and administrative overhead. Which compute target should you use?

☐ A

Azure Kubernetes Service (AKS) inference cluster

☐ B

Azure Machine Learning compute cluster

☐ C

attached Azure Databricks cluster

☐ D

Azure Container Instance (ACI)

CORRECT ANSWER: D

KEEP OPEN

EXPLANATION:

Explanation: Azure Container Instances (ACI) are suitable only for small models less than 1 GB in size. Use it for low-scale CPU-based workloads that require less than 48 GB of RAM. Note: Microsoft recommends using single-node Azure Kubernetes Service (AKS) clusters for dev-test of larger models. Reference: <https://docs.microsoft.com/id-id/azure/machine-learning/how-to-deploy-and-where>

TEST 5 - QUESTION: 28/50

You deploy a model as an Azure Machine Learning real-time web service using the following code.

```
# ws, model, inference_config, and deployment_config defined previously
service = Model.deploy(ws, 'classification-service', [model], inference_config, deployment_config)
service.wait_for_deployment(True)
```

The deployment fails. You need to troubleshoot the deployment failure by determining the actions that were performed during deployment and identifying the specific action that failed. Which code segment should you run?

☐ A
service.update_deployment_state()

☐ B
service.get_logs()

☐ C
service.serialize()

☐ D
service.state

CORRECT ANSWER: B

KEEP OPEN

EXPLANATION:

Explanation: You can print out detailed Docker engine log messages from the service object. You can view the log for ACI, AKS, and Local deployments. The following example demonstrates how to print the logs. # if you already have the service object handy print(service.get_logs()) # if you only know the name of the service (note there might be multiple services with the same name but different version number) print(ws.webservices['mysvc'].get_logs()) Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-troubleshoot-deployment>

TEST 5 - QUESTION: 29/50

An organization creates and deploys a multi-class image classification deep learning model that uses a set of labeled photographs. The software engineering team reports there is a heavy inferencing load for the prediction web services during the summer. The production web service for the model fails to meet demand despite having a fully-utilized compute cluster where the web service is deployed. You need to improve performance of the image classification web service with minimal downtime and minimal administrative effort. What should you advise the IT Operations team to do?

☐ A

Increase the VM size of nodes in the compute cluster where the web service is deployed.

☐ B

Increase the node count of the compute cluster where the web service is deployed.

☐ C

Increase the minimum node count of the compute cluster where the web service is deployed.

☐ D

Create a new compute cluster by using larger VM sizes for the nodes, redeploy the web service to that cluster, and update the DNS registration for the service endpoint to point to the new cluster.

CORRECT ANSWER: B

KEEP OPEN

EXPLANATION:

Explanation: The Azure Machine Learning SDK does not provide support scaling an AKS cluster. To scale the nodes in the cluster, use the UI for your AKS cluster in the Azure Machine Learning studio. You can only change the node count, not the VM size of the cluster. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-kubernetes>

TEST 5 - QUESTION: 30/50

You retrain an existing model. You need to register the new version of a model while keeping the current version of the model in the registry. What should you do?

☐ A

Save the new model in the default datastore with the same name as the existing model. Do not register the new model.

☐ B

Delete the existing model and register the new one with the same name.

☐ C

Register the model with the same name as the existing model.

☐ D

Register a model with a different name from the existing model and a custom property named version with the value 2

CORRECT ANSWER: C

KEEP OPEN

EXPLANATION:

Explanation: Model version: A version of a registered model. When a new model is added to the Model Registry, it is added as Version 1. Each model registered to the same model name increments the version number. Reference: <https://docs.microsoft.com/en-us/azure/databricks/applications/mlflow/model-registry>

TEST 5 - QUESTION: 31/50

HOTSPOT You deploy a model in Azure Container Instance. You must use the Azure Machine Learning SDK to call the model API. You need to invoke the deployed model using native SDK classes and methods. How should you complete the command? To answer, select the appropriate options in the answer areas. NOTE : Each correct selection is worth one point.

☐ CHECK BELOW THE RIGHT ANSWER

Answer Area

```
from azureml.core import Workspace

from azureml.core.webservice import requests
from azureml.core.webservice import Webservice
from azureml.core.webservice import LocalWebservice

import json
ws = Workspace.from_config()
service_name = "mlmodel1-service"
service = Webservice(name=service_name, workspace=ws)
x_new = [[2,101.5,1,24,21], [1,89.7,4,41,21]]
input_json = json.dumps({"data": x_new})

predictions = service.run(input_json)
predictions = requests.post(service.scoring_uri, input_json)
predictions = service.deserialize(ws, input_json)
```

CORRECT ANSWER:

KEEP OPEN

EXPLANATION:

Explanation: Box 1: from azureml.core.webservice import Webservice The following code shows how to use the SDK to update the model, environment, and entry script for a web service to Azure Container Instances: from azureml.core import Environment from azureml.core.webservice import Webservice from azureml.core.model import Model, InferenceConfig Box 2: predictions = service.run(input_json) Example: The following code demonstrates sending data to the service: import json test_sample = json.dumps({'data': [[1, 2, 3, 4, 5, 6, 7, 8, 9, 10], [10, 9, 8, 7, 6, 5, 4, 3, 2, 1]]}) test_sample = bytes(test_sample, encoding='utf8') prediction = service.run(input_data=test_sample) print(prediction) Reference: <https://docs.microsoft.com/bs-latn-ba/azure/machine-learning/how-to-deploy-azure-container-instance> <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-troubleshoot-deployment>

TEST 5 - QUESTION: 32/50

HOTSPOT You need to configure the Feature Based Feature Selection module based on the experiment requirements and datasets. How should you configure the module properties? To answer, select the appropriate options in the dialog box in the answer area. **NOTE:** Each correct selection is worth one point.

SCENARIO

Case study This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided. To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study. At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section. To start the case study To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

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. You add both files to Azure Machine Learning Studio as separate datasets to the starting point for an experiment. Both datasets contain 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 a 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

An initial investigation shows 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. Data issues Missing values 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 datasets also contain 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. 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-parametric statistics to measure relationships. You must use a feature selection algorithm to analyze the relationship between the MedianValue and AvgRoomsInHouse columns. Model training Permutation Feature Importance Given a trained model and a test dataset, you must compute the Permutation Feature Importance scores of feature variables. You must be determined the absolute fit for the model. Hyperparameters You must configure

hyperparameters in the model learning process to speed the learning phase . 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, must 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. Cross-validation 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. You must complete this task before the data goes through the sampling process. Linear regression module When you train a Linear Regression module, you must 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. The distribution of features across multiple training models must be 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.



CHECK BELOW THE RIGHT ANSWER

Answer Area

Filter Based Feature Selection

Feature scoring method

▼

Fisher Score
Chi-squared
Mutual information
Counts

☒ Operate on feature columns only

Target column

▼

MedianValue
AvgRoomsInHouse

Launch column selector

Number of desired features

1

CORRECT ANSWER:

KEEP OPEN

EXPLANATION:

Explanation: Box 1: Mutual Information. The mutual information score is particularly useful in feature selection because it maximizes the mutual information between the joint distribution and target variables in datasets with many dimensions. Box 2: MedianValue MedianValue is the feature column, , it is the predictor of the dataset. Scenario: 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. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/filter-based-feature-selection>

TEST 5 - QUESTION: 33/50

You use the Azure Machine Learning designer to create and run a training pipeline. The pipeline must be run every night to inference predictions from a large volume of files. The folder where the files will be stored is defined as a dataset. You need to publish the pipeline as a REST service that can be used for the nightly inferencing run. What should you do?

☐ A

Set the compute target for the pipeline to an inference cluster

☐ B

Clone the pipeline

☐ C

Create a batch inference pipeline

☐ D

Create a real-time inference pipeline

CORRECT ANSWER: C

KEEP OPEN

EXPLANATION:

Explanation: Azure Machine Learning Batch Inference targets large inference jobs that are not time-sensitive. Batch Inference provides cost-effective inference compute scaling, with unparalleled throughput for asynchronous applications. It is optimized for high-throughput, fire-and-forget inference over large collections of data. You can submit a batch inference job by pipeline_run, or through REST calls with a published pipeline. Reference: <https://github.com/Azure/MachineLearningNotebooks/blob/master/how-to-use-azureml/machine-learning-pipelines/parallel-run/README.md>

TEST 5 - QUESTION: 34/50

You use the designer to create a training pipeline for a classification model. The pipeline uses a dataset that includes the features and labels required for model training. You create a real-time inference pipeline from the training pipeline. You observe that the schema for the generated web service input is based on the dataset and includes the label column that the model predicts. Client applications that use the service must not be required to submit this value. You need to modify the inference pipeline to meet the requirement. What should you do?

☐ A

Replace the dataset in the inference pipeline with an Enter Data Manually module that includes data for the feature columns but not the label column.

☐ B

Delete the Web Service Input module from the inference pipeline.

☐ C

Add a Select Columns in Dataset module to the inference pipeline after the dataset and use it to select all columns other than the label.

☐ D

Delete the dataset from the training pipeline and recreate the real-time inference pipeline.

CORRECT ANSWER: C

KEEP OPEN

EXPLANATION:

Explanation: By default, the Web Service Input will expect the same data schema as the module output data which connects to the same downstream port as it. You can remove the target variable column in the inference pipeline using Select Columns in Dataset module. Make sure that the output of Select Columns in Dataset removing target variable column is connected to the same port as the output of the Web Service Input module. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-designer-automobile-price-deploy>

TEST 5 - QUESTION: 35/50

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. You train and register a machine learning model. You plan to deploy the model as a real-time web service. Applications must use key-based authentication to use the model. You need to deploy the web service. Solution: Create an AciWebservice instance. Set the value of the auth_enabled property to True . Deploy the model to the service. Does the solution meet the goal?

☐ A

Yes

☐ B

No

CORRECT ANSWER: A

KEEP OPEN

EXPLANATION:

Explanation: Key-based authentication. Web services deployed on AKS have key-based auth enabled by default. ACI-deployed services have key-based auth disabled by default, but you can enable it by setting auth_enabled = TRUE when creating the ACI web service. The following is an example of creating an ACI deployment configuration with key-based auth enabled. deployment_config <- aci_webservice_deployment_config(cpu_cores = 1, memory_gb = 1, auth_enabled = TRUE) Reference: <https://azure.git> hub.io/azureml-sdk-for-r/articles/deploying-models.html

TEST 5 - QUESTION: 36/50

You are a data scientist creating a linear regression model. You need to determine how closely the data fits the regression line. Which metric should you review?

☐ A

Root Mean Square Error

☐ B

Mean absolute error

☐ C

Coefficient of determination

☐ D

Recall

☐ E

Precision

CORRECT ANSWER: C

KEEP OPEN

EXPLANATION:

Explanation: Coefficient of determination, often referred to as R^2 , represents the predictive power of the model as a value between 0 and 1. Zero means the model is random (explains nothing); 1 means there is a perfect fit. However, caution should be used in interpreting R^2 values, as low values can be entirely normal and high values can be suspect. Incorrect Answers: A: Root mean squared error (RMSE) creates a single value that summarizes the error in the model. By squaring the difference, the metric disregards the difference between over-prediction and under-prediction. C: Recall is the fraction of all correct results returned by the model. D: Precision is the proportion of true results over all positive results. E: Mean absolute error (MAE) measures how close the predictions are to the actual outcomes; thus, a lower score is better. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model>

TEST 5 - QUESTION: 37/50

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. You are creating a model to predict the price of a student's artwork depending on the following variables: the student's length of education, degree type, and art form. You start by creating a linear regression model. You need to evaluate the linear regression model. Solution: Use the following metrics: Mean Absolute Error, Root Mean Absolute Error, Relative Absolute Error, Relative Squared Error, and the Coefficient of Determination. Does the solution meet the goal?

☐ A

Yes

☐ B

No

CORRECT ANSWER: A

KEEP OPEN

EXPLANATION:

Explanation: The following metrics are reported for evaluating regression models. When you compare models, they are ranked by the metric you select for evaluation. Mean absolute error (MAE) measures how close the predictions are to the actual outcomes; thus, a lower score is better. Root mean squared error (RMSE) creates a single value that summarizes the error in the model. By squaring the difference, the metric disregards the difference between over-prediction and under-prediction. Relative absolute error (RAE) is the relative absolute difference between expected and actual values; relative because the mean difference is divided by the arithmetic mean. Relative squared error (RSE) similarly normalizes the total squared error of the predicted values by dividing by the total squared error of the actual values. Mean Zero One Error (MZOE) indicates whether the prediction was correct or not. In other words: $\text{ZeroOneLoss}(x,y) = 1$ when $x \neq y$; otherwise 0. Coefficient of determination, often referred to as R^2 , represents the predictive power of the model as a value between 0 and 1. Zero means the model is random (explains nothing); 1 means there is a perfect fit. However, caution should be used in interpreting R^2 values, as low values can be entirely normal and high values can be suspect. AUC.

TEST 5 - QUESTION: 38/50

You use Azure Machine Learning designer to create a training pipeline for a regression model. You need to prepare the pipeline for deployment as an endpoint that generates predictions asynchronously for a dataset of input data values. What should you do?

☐ A

Clone the training pipeline.

☐ B

Create a batch inference pipeline from the training pipeline.

☐ C

Replace the dataset in the training pipeline with an Enter Data Manually module.

☐ D

Create a real-time inference pipeline from the training pipeline.

CORRECT ANSWER: D

KEEP OPEN

EXPLANATION:

Explanation: You must first convert the training pipeline into a real-time inference pipeline. This process removes training modules and adds web service inputs and outputs to handle requests. Incorrect Answers: A: Use the Enter Data Manually module to create a small dataset by typing values. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-designer-automobile-price-deploy> <https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/enter-data-manually>

TEST 5 - QUESTION: 39/50 SELECT MULTIPLE

You create a batch inference pipeline by using the Azure ML SDK. You run the pipeline by using the following code: `from azureml.pipeline.core import Pipeline`
`from azureml.core.experiment import Experiment` `pipeline = Pipeline(workspace=ws, steps=[parallelrun_step])` `pipeline_run = Experiment(ws, 'batch_pipeline').submit(pipeline)` You need to monitor the progress of the pipeline execution. What are two possible ways to achieve this goal? Each correct answer presents a complete solution. NOTE : Each correct selection is worth one point.

☐ A

Run the following code in a notebook:

```
from azureml.widgets import RunDetails
RunDetails(pipeline_run).show()
```

draw343167☐ B

Use the Inference Clusters tab in Machine Learning Studio.

☐ C

Run the following code in a notebook:

```
from azureml.contrib.interpret.explanation.explanation_client import ExplanationClient
client = ExplanationClient.from_run(pipeline_run)
explanation = client.download_model_explanation()
explanation = client.download_model_explanation(top_k=4)
global_importance_values = explanation.get_ranked_global_values()
global_importance_names = explanation.get_ranked_global_names()
print('global importance values: {}'.format(global_importance_values))
print('global importance names: {}'.format(global_importance_names))
```

draw343167☐ D

Run the following code and monitor the console output from the PipelineRun object:

```
pipeline_run.wait_for_completion(show_output=True)
```

draw343167☐ E

Use the Activity log in the Azure portal for the Machine Learning workspace

CORRECT ANSWERS: A,D

KEEP OPEN

EXPLANATION:

Explanation: A batch inference job can take a long time to finish. This example monitors progress by using a Jupyter widget. You can also manage the job's progress by using: Azure Machine Learning Studio. Console output from the PipelineRun object. `from azureml.widgets import RunDetails`
`RunDetails(pipeline_run).show()`
`pipeline_run.wait_for_completion(show_output=True)` Reference:

TEST 5 - QUESTION: 40/50

You create an Azure Machine Learning workspace named ML-workspace. You also create an Azure Databricks workspace named DB-workspace. DB-workspace contains a cluster named DB-cluster. You must use DB-cluster to run experiments from notebooks that you import into DB-workspace. You need to use ML-workspace to track MLflow metrics and artifacts generated by experiments running on DB-cluster. The solution must minimize the need for custom code. What should you do?

☐ A

From ML-workspace, create a compute cluster.

☐ B

From DB-cluster, configure the Advanced Logging option.

☐ C

From DB-workspace, configure the Link Azure ML workspace option.

☐ D

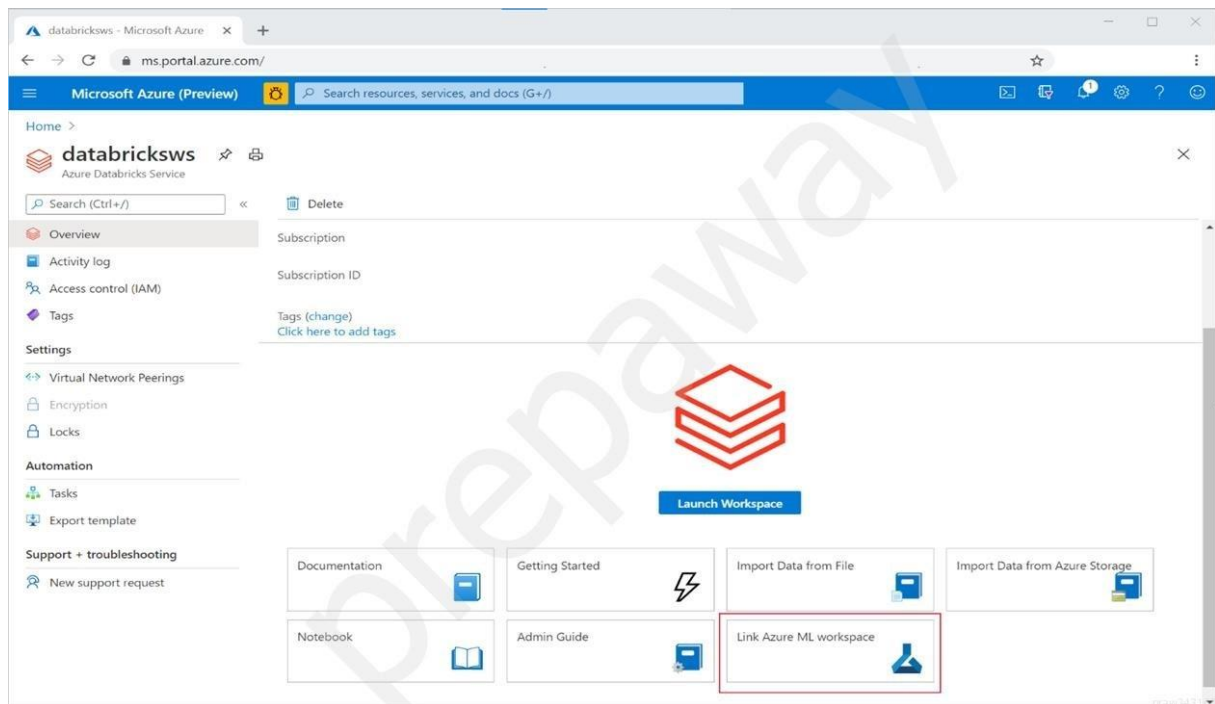
From ML-workspace, create an attached compute.

CORRECT ANSWER: C

KEEP OPEN

EXPLANATION:

Explanation: Connect your Azure Databricks and Azure Machine Learning workspaces: Linking your ADB workspace to your Azure Machine Learning workspace enables you to track your experiment data in the Azure Machine Learning workspace. To link your ADB workspace to a new or existing Azure Machine Learning workspace Sign in to Azure portal. Navigate to your ADB workspace's Overview page. Select the Link Azure Machine Learning workspace button on the bottom right.



Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-mlflow-azure-databricks>

TEST 5 - QUESTION: 41/50

You need to select a feature extraction method. Which method should you use?

SCENARIO

Case study This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided. To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study. At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section. To start the case study To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

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. You add both files to Azure Machine Learning Studio as separate datasets to the starting point for an experiment. Both datasets contain 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 a 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

An initial investigation shows 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. Data issues Missing values 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 datasets also contain 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. 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-parametric statistics to measure relationships. You must use a feature selection algorithm to analyze the relationship between the MedianValue and AvgRoomsInHouse columns. Model training Permutation Feature Importance Given a trained model and a test dataset, you must compute the Permutation Feature Importance scores of feature variables. You must be determined the absolute fit for the model. Hyperparameters You must configure

hyperparameters in the model learning process to speed the learning phase . 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, must 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. Cross-validation 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. You must complete this task before the data goes through the sampling process. Linear regression module When you train a Linear Regression module, you must 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. The distribution of features across multiple training models must be 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.

☐

A

Spearman correlation

☐

B

Pearson's correlation

☐

C

Mutual information

☐

D

Fisher Linear Discriminant Analysis

CORRECT ANSWER: A

KEEP OPEN

EXPLANATION:

Explanation: Spearman's rank correlation coefficient assesses how well the relationship between two variables can be described using a monotonic function. Note: Both Spearman's and Kendall's can be formulated as special cases of a more general correlation coefficient, and they are both appropriate in this scenario. Scenario: 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. Incorrect Answers: B: The Spearman correlation between two variables is equal to the Pearson correlation between the rank values of those two variables; while Pearson's correlation assesses linear relationships, Spearman's correlation assesses monotonic relationships (whether linear or not). Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/feature-selection-modules>

TEST 5 - QUESTION: 42/50

You use Azure Machine Learning designer to create a real-time service endpoint. You have a single Azure Machine Learning service compute resource. You train the model and prepare the real-time pipeline for deployment. You need to publish the inference pipeline as a web service. Which compute type should you use?

☐

A

the existing Machine Learning Compute resource

☐

B

a new Machine Learning Compute resource

☐

C

HDInsight

☐

D

Azure Databricks

☐

E

Azure Kubernetes Services

CORRECT ANSWER: E

KEEP OPEN

EXPLANATION:

Explanation: Azure Kubernetes Service (AKS) can be used real-time inference.

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

TEST 5 - QUESTION: 43/50

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. You train a classification model by using a logistic regression algorithm. You must be able to explain the model's predictions by calculating the importance of each feature, both as an overall global relative importance value and as a measure of local importance for a specific set of predictions. You need to create an explainer that you can use to retrieve the required global and local feature importance values. Solution: Create a MimicExplainer. Does the solution meet the goal?

☐ A

Yes

☐ B

No

CORRECT ANSWER: B

KEEP OPEN

EXPLANATION:

Explanation: Instead use Permutation Feature Importance Explainer (PFI). Note 1: Mimic explainer is based on the idea of training global surrogate models to mimic blackbox models. A global surrogate model is an intrinsically interpretable model that is trained to approximate the predictions of any black box model as accurately as possible. Data scientists can interpret the surrogate model to draw conclusions about the black box model. Note 2: Permutation Feature Importance Explainer (PFI): Permutation Feature Importance is a technique used to explain classification and regression models. At a high level, the way it works is by randomly shuffling data one feature at a time for the entire dataset and calculating how much the performance metric of interest changes. The larger the change, the more important that feature is. PFI can explain the overall behavior of any underlying model but does not explain individual predictions. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability>

TEST 5 - QUESTION: 44/50

You use the following Python code in a notebook to deploy a model as a web service:

```
from azureml.core.webservice import AciWebservice
from azureml.core.model import InferenceConfig
inference_config = InferenceConfig(runtime='python',
                                   source_directory='model_files',
                                   entry_script='score.py',
                                   conda_file='env.yml')
deployment_config = AciWebservice.deploy_configuration(cpu_cores=1, memory_gb=1)
service = Model.deploy(ws, 'my-service', [model], inference_config, deployment_config)
service.wait_for_deployment(True)
```

The deployment fails. You need to use the Python SDK in the notebook to determine the events that occurred during service deployment an initialization. Which code segment should you use?

☐ A
service.state

☐ B
service.environment

☐ C
service.get_logs()

☐ D
service.serialize()

CORRECT ANSWER: C

KEEP OPEN

EXPLANATION:

Explanation: The first step in debugging errors is to get your deployment logs. In Python: `service.get_logs()` Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-troubleshoot-deployment>

TEST 5 - QUESTION: 45/50

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. You are creating a model to predict the price of a student's artwork depending on the following variables: the student's length of education, degree type, and art form. You start by creating a linear regression model. You need to evaluate the linear regression model. Solution: Use the following metrics: Relative Squared Error, Coefficient of Determination, Accuracy, Precision, Recall, F1 score, and AUC. Does the solution meet the goal?

☐ A

No

☐ B

Yes

CORRECT ANSWER: A

KEEP OPEN

EXPLANATION:

Explanation: Relative Squared Error, Coefficient of Determination are good metrics to evaluate the linear regression model, but the others are metrics for classification models. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model>

TEST 5 - QUESTION: 46/50

You have a Python script that executes a pipeline. The script includes the following code: `from azureml.core import Experiment pipeline_run = Experiment(ws, 'pipeline_test').submit(pipeline)` You want to test the pipeline before deploying the script. You need to display the pipeline run details written to the STDOUT output when the pipeline completes. Which code segment should you add to the test script?

- ☐ A
`pipeline_run.wait_for_completion(show_output=True)`
- ☐ B
`pipeline_run.get.metrics()`
- ☐ C
`pipeline_run.get_status()`
- ☐ D
`pipeline_param = PipelineParameter(name="stdout", default_value="console")`

CORRECT ANSWER: A

KEEP OPEN

EXPLANATION:

Explanation: `wait_for_completion`: Wait for the completion of this run. Returns the status object after the wait. Syntax: `wait_for_completion(show_output=False, wait_post_processing=False, raise_on_error=True)` Parameter: `show_output` Indicates whether to show the run output on `sys.stdout`.

TEST 5 - QUESTION: 47/50

You train a model and register it in your Azure Machine Learning workspace. You are ready to deploy the model as a real-time web service. You deploy the model to an Azure Kubernetes Service (AKS) inference cluster, but the deployment fails because an error occurs when the service runs the entry script that is associated with the model deployment. You need to debug the error by iteratively modifying the code and reloading the service, without requiring a re-deployment of the service for each code update. What should you do?

☐ A

Modify the AKS service deployment configuration to enable application insights and re-deploy to AKS.

☐ B

Create an Azure Container Instances (ACI) web service deployment configuration and deploy the model on ACI.

☐ C

Register a new version of the model and update the entry script to load the new version of the model from its registered path.

☐ D

Create a local web service deployment configuration and deploy the model to a local Docker container.

☐ E

Add a breakpoint to the first line of the entry script and redeploy the service to AKS.

CORRECT ANSWER: B

KEEP OPEN

EXPLANATION:

Explanation: How to work around or solve common Docker deployment errors with Azure Container Instances (ACI) and Azure Kubernetes Service (AKS) using Azure Machine Learning. The recommended and the most up to date approach for model deployment is via the `Model.deploy()` API using an `Environment` object as an input parameter. In this case our service will create a base docker image for you during deployment stage and mount the required models all in one call. The basic deployment tasks are: 1. Register the model in the workspace model registry. 2. Define Inference Configuration: a) Create an `Environment` object based on the dependencies you specify in the environment yaml file or use one of our procured environments. b) Create an inference configuration (`InferenceConfig` object) based on the environment and the scoring script. 3. Deploy the model to Azure Container Instance (ACI) service or to Azure Kubernetes Service (AKS).

TEST 5 - QUESTION: 48/50

You are performing feature engineering on a dataset. You must add a feature named CityName and populate the column value with the text London . You need to add the new feature to the dataset. Which Azure Machine Learning Studio module should you use?

☐ A

Preprocess Text

☐ B

Edit Metadata

☐ C

Apply SQL Transformation

☐ D

Extract N-Gram Features from Text

CORRECT ANSWER: B

KEEP OPEN

EXPLANATION:

Explanation: Typical metadata changes might include marking columns as features. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/edit-metadata>

TEST 5 - QUESTION: 49/50

You develop and train a machine learning model to predict fraudulent transactions for a hotel booking website. Traffic to the site varies considerably. The site experiences heavy traffic on Monday and Friday and much lower traffic on other days. Holidays are also high web traffic days. You need to deploy the model as an Azure Machine Learning real-time web service endpoint on compute that can dynamically scale up and down to support demand. Which deployment compute option should you use?

- ☐ A
attached Azure Databricks cluster
- ☐ B
attached virtual machine in a different region
- ☐ C
Azure Machine Learning Compute Instance
- ☐ D
Azure Container Instance (ACI)
- ☐ E
Azure Kubernetes Service (AKS) inference cluster

CORRECT ANSWER: C

KEEP OPEN

EXPLANATION:

Explanation: Azure Machine Learning compute cluster is a managed-compute infrastructure that allows you to easily create a single or multi-node compute. The compute is created within your workspace region as a resource that can be shared with other users in your workspace. The compute scales up automatically when a job is submitted, and can be put in an Azure Virtual Network. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-compute-sdk>

TEST 5 - QUESTION: 50/50

HOTSPOT You write code to retrieve an experiment that is run from your Azure Machine Learning workspace. The run used the model interpretation support in Azure Machine Learning to generate and upload a model explanation. Business managers in your organization want to see the importance of the features in the model. You need to print out the model features and their relative importance in an output that looks similar to the following.

Feature	Importance
0	1.5627435610083558
2	0.6077689312583112
4	0.5574002432900718
3	0.42858759955671777
1	0.3501361539771977

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

☐ CHECK BELOW THE RIGHT ANSWER

Answer Area

Assume required modules are imported

```
ws = Workspace.from_config()
```

```
feature_importances = explanation.
```

(workspace = ws,
experiment_name='train_and_explain',
run_id='train_and_explain_12345')

- from_run
- list_model_explanations
- from_run_id
- download_model_explanation

```
explanation = client.
```

()

- upload_model_explanation
- list_model_explanations
- run
- download_model_explanation

```
feature_importances = explanation.
```

()

- explanation
- explanation_client
- get_feature_importance
- download_model_explanation

```
for key, value in feature_importances.items():  
    print(key, "\t", value)
```

CORRECT ANSWER:

KEEP OPEN

EXPLANATION:

Explanation: Box 1: from_run_id from_run_id(workspace, experiment_name, run_id) Create the client with factory method given a run ID. Returns an instance of the ExplanationClient. Parameters Workspace Workspace - An object that

represents a workspace. experiment_name str - The name of an experiment.
run_id str - A GUID that represents a run. Box 2: list_model_explanations
list_model_explanations returns a dictionary of metadata for all model
explanations available. Returns A dictionary of explanation metadata such as
id, data type, explanation method, model type, and upload time, sorted by upload
time Box 3: explanation Reference: https://docs.microsoft.com/en-us/python/api/azureml-contrib-interpret/azureml.contrib.interpret.explanation.explanation_client.explanation_client?view=azure-ml-py

TEST 6 - QUESTION: 1/10 SELECT MULTIPLE

You are building a binary classification model by using a supplied training set. The training set is imbalanced between two classes. You need to resolve the data imbalance. What are three possible ways to achieve this goal? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

☐ A

Normalize the training feature set

☐ B

Penalize the classification

☐ C

Generate synthetic samples in the minority class

☐ D

Use accuracy as the evaluation metric of the model

☐ E

Resample the dataset using undersampling or oversampling

CORRECT ANSWERS: B,C,E

KEEP OPEN

EXPLANATION:

Explanation: A: Try Penalized Models You can use the same algorithms but give them a different perspective on the problem. Penalized classification imposes an additional cost on the model for making classification mistakes on the minority class during training. These penalties can bias the model to pay more attention to the minority class. B: You can change the dataset that you use to build your predictive model to have more balanced data. This change is called sampling your dataset and there are two main methods that you can use to even-up the classes: Consider testing under-sampling when you have a lot data (tens- or hundreds of thousands of instances or more) Consider testing over-sampling when you don't have a lot of data (tens of thousands of records or less) D: Try Generate Synthetic Samples A simple way to generate synthetic samples is to randomly sample the attributes from instances in the minority class. Reference: <https://machinelearningmastery.com/tactics-to-combat-imbalanced-classes-in-your-machine-learning-dataset/>

TEST 6 - QUESTION: 2/10

HOTSPOT A biomedical research company plans to enroll people in an experimental medical treatment trial. You create and train a binary classification model to support selection and admission of patients to the trial. The model includes the following features: Age, Gender, and Ethnicity. The model returns different performance metrics for people from different ethnic groups. You need to use Fairlearn to mitigate and minimize disparities for each category in the Ethnicity feature. Which technique and constraint should you use? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

☐ CHECK BELOW THE RIGHT ANSWER

Answer Area

Option	Value
Technique	<div><div></div><div>Grid search</div><div>Outlier detection</div><div>Dimensionality reduction</div></div>
Constraint	<div><div></div><div>Demographic parity</div><div>False-positive rate parity</div></div>

CORRECT ANSWER:

KEEP OPEN

EXPLANATION:

Explanation: Box 1: Grid Search Fairlearn open-source package provides postprocessing and reduction unfairness mitigation algorithms: ExponentiatedGradient, GridSearch, and ThresholdOptimizer. Note: The Fairlearn open-source package provides postprocessing and reduction unfairness mitigation algorithms types: Reduction: These algorithms take a standard black-box machine learning estimator (e.g., a LightGBM model) and generate a set of retrained models using a sequence of re-weighted training datasets. Post-processing: These algorithms take an existing classifier and the sensitive feature as input. Box 2: Demographic parity The Fairlearn open-source package supports the following types of parity constraints: Demographic parity, Equalized odds, Equal opportunity, and Bounded group loss. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/concept-fairness-ml>

TEST 6 - QUESTION: 3/10

You create a binary classification model. The model is registered in an Azure Machine Learning workspace. You use the Azure Machine Learning Fairness SDK to assess the model fairness. You develop a training script for the model on a local machine. You need to load the model fairness metrics into Azure Machine Learning studio. What should you do?

☐ A

Upload the training script

☐ B

Implement the `upload_dashboard_dictionary` function

☐ C

Implement the `download_dashboard_by_upload_id` function

☐ D

Implement the `create_group_metric_set` function

CORRECT ANSWER: B

KEEP OPEN

EXPLANATION:

Explanation: import `azureml.contrib.fairness` package to perform the upload:
from `azureml.contrib.fairness` import `upload_dashboard_dictionary`,
`download_dashboard_by_upload_id` Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-fairness-aml>

TEST 6 - QUESTION: 4/10

You have a dataset that includes confidential data. You use the dataset to train a model. You must use a differential privacy parameter to keep the data of individuals safe and private. You need to reduce the effect of user data on aggregated results. What should you do?

☐ A

Increase the value of the epsilon parameter to decrease privacy and increase accuracy

☐ B

Decrease the value of the epsilon parameter to increase privacy and reduce accuracy

☐ C

Decrease the value of the epsilon parameter to reduce the amount of noise added to the data

☐ D

Set the value of the epsilon parameter to 1 to ensure maximum privacy

CORRECT ANSWER: B

KEEP OPEN

EXPLANATION:

Explanation: Differential privacy tries to protect against the possibility that a user can produce an indefinite number of reports to eventually reveal sensitive data. A value known as epsilon measures how noisy, or private, a report is. Epsilon has an inverse relationship to noise or privacy. The lower the epsilon, the more noisy (and private) the data is. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/concept-differential-privacy>

TEST 6 - QUESTION: 5/10

HOTSPOT You are using a decision tree algorithm. You have trained a model that generalizes well at a tree depth equal to 10. You need to select the bias and variance properties of the model with varying tree depth values. Which properties should you select for each tree depth? To answer, select the appropriate options in the answer area.

☐ CHECK BELOW THE RIGHT ANSWER

Answer Area

Tree Depth	Bias	Variance
5	<div><div></div><div>High</div><div>Low</div><div>Identical</div></div>	<div><div></div><div>High</div><div>Low</div><div>Identical</div></div>
15	<div><div></div><div>High</div><div>Low</div><div>Identical</div></div>	<div><div></div><div>High</div><div>Low</div><div>Identical</div></div>

CORRECT ANSWER:

KEEP OPEN

EXPLANATION:

Explanation: In decision trees, the depth of the tree determines the variance. A complicated decision tree (e.g. deep) has low bias and high variance. Note: In statistics and machine learning, the bias–variance tradeoff is the property of a set of predictive models whereby models with a lower bias in parameter estimation have a higher variance of the parameter estimates across samples, and vice versa. Increasing the bias will decrease the variance. Reference: <https://machinelearningmastery.com/gentle-introduction-to-the-bias-variance-trade-off-in-machine-learning/>

TEST 6 - QUESTION: 6/10

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. You are creating a model to predict the price of a student's artwork depending on the following variables: the student's length of education, degree type, and art form. You start by creating a linear regression model. You need to evaluate the linear regression model. Solution: Use the following metrics: Mean Absolute Error, Root Mean Absolute Error, Relative Absolute Error, Accuracy, Precision, Recall, F1 score, and AUC. Does the solution meet the goal?

☐ A

Yes

☐ B

No

CORRECT ANSWER: B

KEEP OPEN

EXPLANATION:

Explanation: Accuracy, Precision, Recall, F1 score, and AUC are metrics for evaluating classification models. Note: Mean Absolute Error, Root Mean Absolute Error, Relative Absolute Error are OK for the linear regression model.

Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model>

TEST 6 - QUESTION: 7/10

HOTSPOT You are developing a linear regression model in Azure Machine Learning Studio. You run an experiment to compare different algorithms. The following image displays the results dataset output:

Algorithm	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Relative Squared Error
Bayesian Liner	3.276025	4.655442	0.511436	0.282138
Neural Network	2.676538	3.621476	0.417847	0.17073
Boosted Decision Tree	2.168847	2.878077	0.338589	0.107831
Linear	6.350005	8.720718	0.99133	0.99002
Decision Forest	2.390206	3.315 164	0.373146	0.14307

Use the drop-down menus to select the answer choice that answers each question based on the information presented in the image. NOTE: Each correct selection is worth one point.

☐ CHECK BELOW THE RIGHT ANSWER

Answer Area

Which algorithm minimizes differences between actual and predicted values?

▼

Bayesian Linear Regression
Neutral Network Regression
Boosted Decision Tree Regression
Linear Regression
Decision Forest Regression

Which approach should you use to find the best parameters for a Linear Regression model for the Online Gradient Descent method?

▼

Set the Decrease learning rate option to True.
Set the Decrease learning rate option to False.
Set the Create trainer mode option to Parameter Range.
Increase the number of epochs.
Decrease the number of epochs.

CORRECT ANSWER:

KEEP OPEN

EXPLANATION:

Explanation: Box 1: Boosted Decision Tree Regression Mean absolute error (MAE) measures how close the predictions are to the actual outcomes; thus, a lower score is better. Box 2: Online Gradient Descent: If you want the algorithm to find the best parameters for you, set Create trainer mode option to Parameter Range. You can then specify multiple values for the algorithm to try. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model> <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/linear-regression>

TEST 6 - QUESTION: 8/10

HOTSPOT You train a classification model by using a decision tree algorithm. You create an estimator by running the following Python code. The variable `feature_names` is a list of all feature names, and `class_names` is a list of all class names.

```
from interpret.ext.blackbox import TabularExplainer
explainer = TabularExplainer(model, x_train, features=feature_names, classes=class_names)
```

You need to explain the predictions made by the model for all classes by determining the importance of all features. For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

☐ CHECK BELOW THE RIGHT ANSWER

Answer Area

	Yes	No
The SHAP TreeExplainer will be used to interpret the model.	<input type="radio"/>	<input type="radio"/>
If you omit the features and classes parameters in the TabularExplainer instantiation, the explainer still works as expected.	<input checked="" type="radio"/>	<input type="radio"/>
You could interpret the model by using a MimicExplainer instead of a TabularExplainer.	<input type="radio"/>	<input checked="" type="radio"/>

CORRECT ANSWER:

KEEP OPEN

EXPLANATION:

Explanation: Box 1: Yes TabularExplainer calls one of the three SHAP explainers underneath (TreeExplainer, DeepExplainer, or KernelExplainer). Box 2: Yes To make your explanations and visualizations more informative, you can choose to pass in feature names and output class names if doing classification. Box 3: No TabularExplainer automatically selects the most appropriate one for your use case, but you can call each of its three underlying explainers underneath (TreeExplainer, DeepExplainer, or KernelExplainer) directly. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability-aml>

TEST 6 - QUESTION: 9/10 SELECT MULTIPLE

You are a data scientist building a deep convolutional neural network (CNN) for image classification. The CNN model you build shows signs of overfitting. You need to reduce overfitting and converge the model to an optimal fit. Which two actions should you perform? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

☐ A

Add an additional dense layer with 512 input units.

☐ B

Use training data augmentation.

☐ C

Add L1/L2 regularization.

☐ D

Add an additional dense layer with 64 input units.

☐ E

Reduce the amount of training data.

CORRECT ANSWERS: C,E

KEEP OPEN

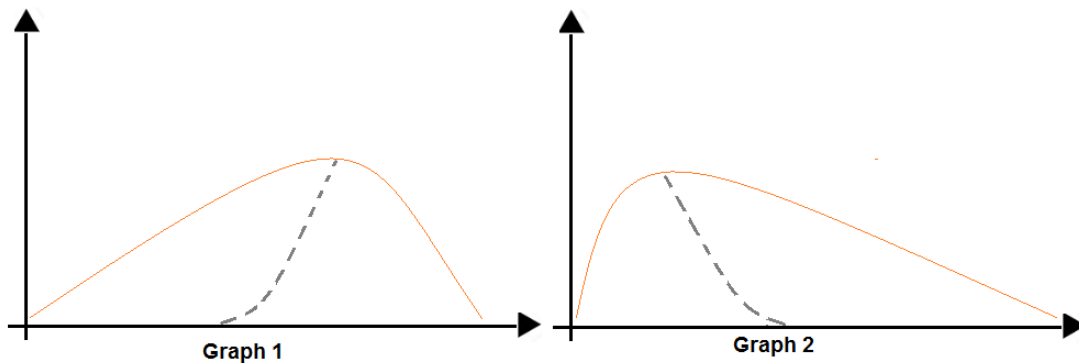
EXPLANATION:

Explanation: B: Weight regularization provides an approach to reduce the overfitting of a deep learning neural network model on the training data and improve the performance of the model on new data, such as the holdout test set. Keras provides a weight regularization API that allows you to add a penalty for weight size to the loss function. Three different regularizer instances are provided; they are: L1: Sum of the absolute weights. L2: Sum of the squared weights. L1L2: Sum of the absolute and the squared weights. D: Because a fully connected layer occupies most of the parameters, it is prone to overfitting. One method to reduce overfitting is dropout. At each training stage, individual nodes are either "dropped out" of the net with probability $1-p$ or kept with probability p , so that a reduced network is left; incoming and outgoing edges to a dropped-out node are also removed. By avoiding training all nodes on all training data, dropout decreases overfitting. Reference:

CORRECT ANSWER:

TEST 6 - QUESTION: 10/10

HOTSPOT You are analyzing the asymmetry in a statistical distribution. The following image contains two density curves that show the probability distribution of two datasets.



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.



CHECK BELOW THE RIGHT ANSWER

Answer Area

Question	Answer choice
Which type of distribution is shown for the dataset density curve of Graph 1?	<div><div></div><div>Negative skew</div><div>Positive skew</div><div>Normal distribution</div><div>Bimodal distribution</div></div>
Which type of distribution is shown for the dataset density curve of Graph 2?	<div><div></div><div>Negative skew</div><div>Positive skew</div><div>Normal distribution</div><div>Bimodal distribution</div></div>

KEEP OPEN

EXPLANATION:

Explanation: Box 1: Positive skew Positive skew values means the distribution is skewed to the right. Box 2: Negative skew Negative skewness values mean the distribution is skewed to the left. Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/compute-elementary-statistics>