Microsoft

Exam Questions DP-100

Designing and Implementing a Data Science Solution on Azure

NEW QUESTION 1
- (Exam Topic 3)
You have an Azure Machine Learning workspace that contains a training cluster and an inference cluster. You plan to create a classification model by using the Azure Machine Learning designer.
You need to ensure that client applications can submit data as HTTP requests and receive predictions as responses.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

A. Mastered
B. Not Mastered

Answer: A

Explanation:

NEW QUESTION 2
- (Exam Topic 3)
You have several machine learning models registered in an Azure Machine Learning workspace. You must use the Fairlearn dashboard to assess fairness in a selected model.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

A. Mastered

B. Not Mastered

Answer: A

Explanation:
Graphical user interface, text, application Description automatically generated
Step 1: Select a model feature to be evaluated.
Step 2: Select a binary classification or regression model.
Register your models within Azure Machine Learning. For convenience, store the results in a dictionary,
which maps the id of the registered model (a string in name:version format) to the predictor itself. Example:
model_dict = {}
lr_reg_id = register_model("fairness_logistic_regression", lr_predictor) model_dict[lr_reg_id] = lr_predictor
svm_reg_id = register_model("fairness_svm", svm_predictor) model_dict[svm_reg_id] = svm_predictor
Step 3: Select a metric to be measured Precompute fairness metrics.
Create a dashboard dictionary using Fairlearn's metrics package. Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-fairness-aml

NEW QUESTION 3
- (Exam Topic 3)
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 create a model to forecast weather conditions based on historical data.
You need to create a pipeline that runs a processing script to load data from a datastore and pass the processed data to a machine learning model training script.
Solution: Run the following code:

```
datastore = ws.get_default_datastore()
data_output = PipelineData("processed_data", datastore=datastore)
process_step = PythonScriptStep(script_name="process.py",
  arguments=["--data_for_train", data_output],
  outputs=[data_output], compute_target=aml_compute,
  source_directory=process_directory)
pipeline = Pipeline(workspace=ws, steps=[process_step])
```

Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:
 train_step is missing. Reference:
https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.pipelinedata?view=azu

NEW QUESTION 4
- (Exam Topic 3)
HOTSPOT
You register the following versions of a model.

Model name	Model version	Tags	Properties
healthcare_model	3	'Training context':'CPU Compute'	value:87.43
healthcare_model	2	'Training context':'CPU Compute'	value:54.98
healthcare_model	1	'Training context':'CPU Compute'	value:23.56

You use the Azure ML Python SDK to run a training experiment. You use a variable named run to reference the experiment run.
After the run has been submitted and completed, you run the following code:

```
run.register_model(model_path='outputs/model.pkl',
  model_name='healthcare_model',
  tags={'Training context':'CPU Compute'} )
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.
NOTE: Each correct selection is worth one point.

	Yes	No
The code will cause a previous version of the saved model to be overwritten.	○	○
The version number will now be 4.	○	○
The latest version of the stored model will have a property of value: 87.43.	○	○

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-and-where

NEW QUESTION 5
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
An IT department creates the following Azure resource groups and resources:

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

The IT department creates an Azure Kubernetes Service (AKS)-based inference compute target named
aks-cluster in the Azure Machine Learning workspace. You have a Microsoft Surface Book computer with a GPU. Python 3.6 and Visual Studio Code are installed.
You need to run a script that trains a deep neural network (DNN) model and logs the loss and accuracy metrics.
Solution: Install the Azure ML SDK on the Surface Book. Run Python code to connect to the workspace. Run the training script as an experiment on the aks-cluster compute target.
Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Need to attach the mlvm virtual machine as a compute target in the Azure Machine Learning workspace.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target

NEW QUESTION 6
- (Exam Topic 3)
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. Azure Container Instance (ACI)
D. attached Azure Databricks cluster

Answer: C

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

NEW QUESTION 7

- (Exam Topic 3)
You are producing a multiple linear regression model in Azure Machine Learning Studio. Several independent variables are highly correlated.
You need to select appropriate methods for conducting effective feature engineering on all the data.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

Action		Answer area
Evaluate the probability function		
Remove duplicate rows		
Use the Filter Based Feature Selection module	◁ ▷	◁ ▽
Test the hypothesis using t-Test		
Compute linear correlation		
Build a counting transform		

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Step 1: Use the Filter Based Feature Selection module
Filter Based Feature Selection identifies the features in a dataset with the greatest predictive power.
The module outputs a dataset that contains the best feature columns, as ranked by predictive power. It also outputs the names of the features and their scores from the selected metric.
Step 2: Build a counting transform
A counting transform creates a transformation that turns count tables into features, so that you can apply the transformation to multiple datasets.
Step 3: Test the hypothesis using t-Test References:
https://docs.microsoft.com/bs-latn-ba/azure/machine-learning/studio-module-reference/filter-based-feature-selec
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/build-counting-transform

NEW QUESTION 8
- (Exam Topic 3)
You use an Azure Machine Learning workspace. You create the following Python code:

```
from azureml.core import ScriptRunConfig
src = ScriptRunConfig(source_directory=project_folder,
                      script='train.py'
                      environment=myenv)
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.
NOTE: Each correct selection is worth one point.

Statements	Yes	No
The default environment will be created	○	○
The training script will run on local compute	○	○
A script run configuration runs a training script named `train.py` located in a directory defined by the `project_folder` variable	○	○

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Graphical user interface, text, application Description automatically generated

Box 1: No
Environment is a required parameter. The environment to use for the run. If no environment is specified, azureml.core.runconfig.DEFAULT_CPU_IMAGE will be used as the Docker image for the run.
The following example shows how to instantiate a new environment. from azureml.core import Environment
myenv = Environment(name="myenv") Box 2: Yes
Parameter compute_target: The compute target where training will happen. This can either be a ComputeTarget object, the name of an existing ComputeTarget, or the string "local". If no compute target is specified, your local machine will be used.
Box 3: Yes
Parameter source_directory. A local directory containing code files needed for a run. Parameter script. The file path relative to the source_directory of the script to be run. Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.scriptrunconfig https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.environment.environment

NEW QUESTION 9
- (Exam Topic 3)
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 analyzing a numerical dataset which contains missing values in several columns.
You must clean the missing values using an appropriate operation without affecting the dimensionality of the feature set.
You need to analyze a full dataset to include all values.
Solution: Calculate the column median value and use the median value as the replacement for any missing value in the column.
Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Use the Multiple Imputation by Chained Equations (MICE) method. References: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

NEW QUESTION 10
- (Exam Topic 3)
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 new experiment in Azure Machine Learning Studio.
One class has a much smaller number of observations than the other classes in the training set. You need to select an appropriate data sampling strategy to compensate for the class imbalance. Solution: You use the Stratified split for the sampling mode.
Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Instead use the Synthetic Minority Oversampling Technique (SMOTE) sampling mode.
Note: SMOTE is used to increase the number of underrepresented cases in a dataset used for machine learning. SMOTE is a better way of increasing the number of rare cases than simply duplicating existing cases.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/smote

NEW QUESTION 10
- (Exam Topic 3)
You are developing a deep learning model by using TensorFlow. You plan to run the model training workload on an Azure Machine Learning Compute Instance.
You must use CUDA-based model training. You need to provision the Compute Instance.
Which two virtual machines sizes can you use? To answer, select the appropriate virtual machine sizes in the answer area.
NOTE: Each correct selection is worth one point.

Virtual machine size

Search by name...

Name ↑	vCPUs	GPUs	RAM	Resource disk
BASIC_A0	1		0.75 GB	20 GB
STANDARD_D3_V2	4		14 GB	200 GB
STANDARD_E64_V3	64		432 GB	1,600 GB
STANDARD_M64LS	64		512 GB	2,000 GB
STANDARD_NC12	12	2	112 GB	680 GB
STANDARD_NC24	24	4	224 GB	1,440 GB

A. Mastered
B. Not Mastered

Answer: A

Explanation:
CUDA is a parallel computing platform and programming model developed by Nvidia for general computing on its own GPUs (graphics processing units). CUDA enables developers to speed up compute-intensive applications by harnessing the power of GPUs for the parallelizable part of the computation.
Reference:
https://www.infoworld.com/article/3299703/what-is-cuda-parallel-programming-for-gpus.html

NEW QUESTION 14
- (Exam Topic 3)
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. Register a new version of the model and update the entry script to load the new version of the model from its registered path.
B. Modify the AKS service deployment configuration to enable application insights and re-deploy to AKS.
C. Create an Azure Container Instances (ACI) web service deployment configuration and deploy the model on ACI.
D. Add a breakpoint to the first line of the entry script and redeploy the service to AKS.
E. Create a local web service deployment configuration and deploy the model to a local Docker container.

Answer: C

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).

NEW QUESTION 19
- (Exam Topic 3)
You are developing deep learning models to analyze semi-structured, unstructured, and structured data types. You have the following data available for model building:
> Video recordings of sporting events
> Transcripts of radio commentary about events
> Logs from related social media feeds captured during sporting events You need to select an environment for creating the model.
Which environment should you use?

A. Azure Cognitive Services
B. Azure Data Lake Analytics
C. Azure HDInsight with Spark MLib
D. Azure Machine Learning Studio

Answer: A

Explanation:
Azure Cognitive Services expand on Microsoft's evolving portfolio of machine learning APIs and enable developers to easily add cognitive features – such as emotion and video detection; facial, speech, and vision recognition; and speech and language understanding – into their applications. The goal of Azure Cognitive Services is to help developers create applications that can see, hear, speak, understand, and even begin to reason. The catalog of services within Azure Cognitive Services can be categorized into five main pillars - Vision, Speech, Language, Search, and Knowledge.
References:
https://docs.microsoft.com/en-us/azure/cognitive-services/welcome

NEW QUESTION 24
- (Exam Topic 3)
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 plan to use a Python script to run an Azure Machine Learning experiment. The script creates a reference to the experiment run context, loads data from a file, identifies the set of unique values for the label column, and completes the experiment run:

```
from azureml.core import Run
import pandas as pd


run = Run.get_context()
data = pd.read_csv('data.csv')
label_vals = data['label'].unique()
# Add code to record metrics here
run.complete()
```

The experiment must record the unique labels in the data as metrics for the run that can be reviewed later. You must add code to the script to record the unique label values as run metrics at the point indicated by the comment.
Solution: Replace the comment with the following code:
run.log_list('Label Values', label_vals) Does the solution meet the goal?

A. Yes
B. No

Answer: A

Explanation:
run.log_list log a list of values to the run with the given name using log_list. Example: run.log_list("accuracies", [0.6, 0.7, 0.87])
Note:
Data= pd.read_csv('data.csv')
Data is read into a pandas.DataFrame, which is a two-dimensional, size-mutable, potentially heterogeneous tabular data.
label_vals =data['label'].unique
label_vals contains a list of unique label values. Reference:
https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run(class) https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html

NEW QUESTION 29
- (Exam Topic 3)
A set of CSV files contains sales records. All the CSV files have the same data schema.
Each CSV file contains the sales record for a particular month and has the filename sales.csv. Each file in stored in a folder that indicates the month and year when the data was recorded. The folders are in an Azure blob container for which a datastore has been defined in an Azure Machine Learning workspace. The folders are organized in a parent folder named sales to create the following hierarchical structure:

```
/sales
  /01-2019
    /sales.csv
  /02-2019
    /sales.csv
  /03-2019
    /sales.csv
  ...
```

At the end of each month, a new folder with that month's sales file is added to the sales folder.
You plan to use the sales data to train a machine learning model based on the following requirements:
➤ You must define a dataset that loads all of the sales data to date into a structure that can be easily converted to a dataframe.
➤ You must be able to create experiments that use only data that was created before a specific previous month, ignoring any data that was added after that month.
➤ You must register the minimum number of datasets possible.
You need to register the sales data as a dataset in Azure Machine Learning service workspace. What should you do?

A. Create a tabular dataset that references the datastore and explicitly specifies each 'sales/mm-yyyy/ sales.csv' file every mont
B. Register the dataset with the name sales_dataset each month, replacing the existing dataset and specifying a tag named month indicating the month and year it was registere
C. Use this dataset for all experiments.
D. Create a tabular dataset that references the datastore and specifies the path 'sales/*/sales.csv', register the dataset with the name sales_dataset and a tag named month indicating the month and year it was registered, and use this dataset for all experiments.
E. Create a new tabular dataset that references the datastore and explicitly specifies each 'sales/mm-yyyy/ sales.csv' file every mont
F. Register the dataset with the name sales_dataset_MM-YYYY each month vith appropriate MM and YYYY values for the month and yea
G. Use the appropriate month-specific dataset for experiments.
H. Create a tabular dataset that references the datastore and explicitly specifies each 'sales/mm-yyyy/ sales.csv' fil
I. Register the dataset with the name sales_dataset each month as a new version and with a tag named month indicating the month and year it was registere
J. Use this dataset for all experiments, identifying the version to be used based on the month tag as necessary.

Answer: B

Explanation:
Specify the path. Example:
The following code gets the workspace existing workspace and the desired datastore by name. And then passes the datastore and file locations to the path parameter to create a new TabularDataset, weather_ds.
from azureml.core import Workspace, Datastore, Dataset datastore_name = 'your datastore name'

```
# get existing workspace
workspace = Workspace.from_config()
# retrieve an existing datastore in the workspace by name datastore = Datastore.get(workspace, datastore_name)
# create a TabularDataset from 3 file paths in datastore datastore_paths = [(datastore, 'weather/2018/11.csv'), (datastore, 'weather/2018/12.csv'),
(datastore, 'weather/2019/*.csv')]
weather_ds = Dataset.Tabular.from_delimited_files(path=datastore_paths)
```

NEW QUESTION 31
- (Exam Topic 3)
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. Create a batch inference pipeline
B. Set the compute target for the pipeline to an inference cluster
C. Create a real-time inference pipeline
D. Clone the pipeline

Answer: A

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-pipeli

NEW QUESTION 34
- (Exam Topic 3)
You are training machine learning models in Azure Machine Learning. You use Hyperdrive to tune the hyperparameters. In previous model training and tuning runs, many models showed similar performance. You need to select an early termination policy that meets the following requirements:
• accounts for the performance of all previous runs when evaluating the current run
• avoids comparing the current run with only the best performing run to date
Which two early termination policies should you use? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

A. Bandit
B. Median stopping
C. Default
D. Truncation selection

Answer: BC

Explanation:
The Median Stopping policy computes running averages across all runs and cancels runs whose best performance is worse than the median of the running averages.
If no policy is specified, the hyperparameter tuning service will let all training runs execute to completion. Reference:
https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.medianstoppingpolicy https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.truncationselectionpoli https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive.banditpolicy

NEW QUESTION 38
- (Exam Topic 3)
You create a new Azure subscription. No resources are provisioned in the subscription. You need to create an Azure Machine Learning workspace.
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. Run Python code that uses the Azure ML SDK library and calls the Workspace.create method with name, subscription_id, resource_group, and location parameters.
B. Use an Azure Resource Management template that includes a Microsoft.MachineLearningServices/ workspaces resource and its dependencies.
C. Use the Azure Command Line Interface (CLI) with the Azure Machine Learning extension to call the az group create function with --name and --location parameters, and then the az ml workspace create function, specifying –w and –g parameters for the workspace name and resource group.
D. Navigate to Azure Machine Learning studio and create a workspace.
E. Run Python code that uses the Azure ML SDK library and calls the Workspace.get method with name, subscription_id, and resource_group parameters.

Answer: BCD

Explanation:
B: You can use an Azure Resource Manager template to create a workspace for Azure Machine Learning. Example:
{"type": "Microsoft.MachineLearningServices/workspaces",
…
C: You can create a workspace for Azure Machine Learning with Azure CLI Install the machine learning extension.
Create a resource group: az group create --name <resource-group-name> --location <location>
To create a new workspace where the services are automatically created, use the following command: az ml workspace create -w <workspace-name> -g <resource-group-name>
D: You can create and manage Azure Machine Learning workspaces in the Azure portal.

➢ Sign in to the Azure portal by using the credentials for your Azure subscription.

➢ In the upper-left corner of Azure portal, select + Create a resource.

➢ Use the search bar to find Machine Learning.

➢ Select Machine Learning.

➢ In the Machine Learning pane, select Create to begin.

Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-workspace-template https://docs.microsoft.com/en-us/azure/machine-learning/how-to-manage-workspace-cli https://docs.microsoft.com/en-us/azure/machine-learning/how-to-manage-workspace

NEW QUESTION 41
- (Exam Topic 3)
You plan to use a Deep Learning Virtual Machine (DLVM) to train deep learning models using Compute Unified Device Architecture (CUDA) computations.
You need to configure the DLVM to support CUDA. What should you implement?

A. Intel Software Guard Extensions (Intel SGX) technology
B. Solid State Drives (SSD)
C. Graphic Processing Unit (GPU)
D. Computer Processing Unit (CPU) speed increase by using overcloking
E. High Random Access Memory (RAM) configuration

Answer: C

Explanation:
A Deep Learning Virtual Machine is a pre-configured environment for deep learning using GPU instances.
References:
https://azuremarketplace.microsoft.com/en-au/marketplace/apps/microsoft-ads.dsvm-deep-learning

NEW QUESTION 45
- (Exam Topic 3)
You are hired as a data scientist at a winery. The previous data scientist used Azure Machine Learning. You need to review the models and explain how each model makes decisions.
Which explainer modules should you use? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Model type	Explainer
A random forest model for predicting the alcohol content in wine given a set of covariates	Tabular / HAN / Text / Image
A natural language processing model for analyzing field reports	Tree / HAN / Text / Image
An image classifier that determines the quality of the grape based upon its physical characteristics.	Kernel / HAN / Text / Image

A. Mastered
B. Not Mastered

Answer: A

Explanation:
 info based on
the given model and data sets. The meta explainers leverage all the libraries (SHAP, LIME, Mimic, etc.) that we have integrated or developed. The following are the meta explainers available in the SDK:
Tabular Explainer: Used with tabular datasets. Text Explainer: Used with text datasets. Image Explainer: Used with image datasets. Box 1: Tabular
Box 2: Text
Box 3: Image Reference:
https://medium.com/microsoftazure/automated-and-interpretable-machine-learning-d07975741298

NEW QUESTION 48
- (Exam Topic 3)
You are solving a classification task.
You must evaluate your model on a limited data sample by using k-fold cross validation. You start by configuring a k parameter as the number of splits.
You need to configure the k parameter for the cross-validation. Which value should you use?

A. k=0.5
B. k=0
C. k=5
D. k=1

Answer: C

Explanation:
Leave One Out (LOO) cross-validation
Setting K = n (the number of observations) yields n-fold and is called leave-one out cross-validation (LOO), a special case of the K-fold approach.
LOO CV is sometimes useful but typically doesn't shake up the data enough. The estimates from each fold are highly correlated and hence their average can have high variance.
This is why the usual choice is K=5 or 10. It provides a good compromise for the bias-variance tradeoff.

NEW QUESTION 53
- (Exam Topic 3)
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 analyzing a numerical dataset which contains missing values in several columns.
You must clean the missing values using an appropriate operation without affecting the dimensionality of the feature set.
You need to analyze a full dataset to include all values.
Solution: Remove the entire column that contains the missing data point. Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Use the Multiple Imputation by Chained Equations (MICE) method. References: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

NEW QUESTION 55
- (Exam Topic 3)
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. Relative Absolute Error
B. AUC Curve
C. Mean Absolute Error
D. Relative Squared Error

Answer: B

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.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio/evaluate-model-performance#evaluating-a-bina

NEW QUESTION 57
- (Exam Topic 3)
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.

Context | **Resource**

Test
- ds-workstation notebook VM
- aks-compute cluster
- cpu-compute cluster
- gpu-compute cluster

Production
- ds-workstation notebook VM
- aks-compute cluster
- cpu-compute cluster
- gpu-compute cluster

A. Mastered
B. Not Mastered

Answer: A

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

NEW QUESTION 61
- (Exam Topic 3)
You are creating a machine learning model. You need to identify outliers in the data.
Which two visualizations can you use? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point. NOTE: Each correct selection is worth one point.

A. box plot
B. scatter
C. random forest diagram

D. Venn diagram
E. ROC curve

Answer: AB

Explanation:
The box-plot algorithm can be used to display outliers.
One other way to quickly identify Outliers visually is to create scatter plots. References:
https://blogs.msdn.microsoft.com/azuredev/2017/05/27/data-cleansing-tools-in-azure-machine-learning/

NEW QUESTION 65
- (Exam Topic 3)
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 using Azure Machine Learning to run an experiment that trains a classification model.
You want to use Hyperdrive to find parameters that optimize the AUC metric for the model. You configure a HyperDriveConfig for the experiment by running the following code:

```
hyperdrive = HyperDriveConfig(estimator=your_estimator,
  hyperparameter_sampling=your_params,
  policy=policy,
  primary_metric_name='AUC',
  primary_metric_goal=PrimaryMetricGoal.MAXIMIZE,
  max_total_runs=6,
  max_concurrent_runs=4)
```

You plan to use this configuration to run a script that trains a random forest model and then tests it with validation data. The label values for the validation data are stored in a variable named y_test variable, and the predicted probabilities from the model are stored in a variable named y_predicted.
You need to add logging to the script to allow Hyperdrive to optimize hyperparameters for the AUC metric. Solution: Run the following code:

```
import numpy as np
from sklearn.metrics import roc_auc_score
# code to train model omitted
auc = roc_auc_score(y_test, y_predicted)
print(np.float(auc))
```

Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Use a solution with logging.info(message) instead. Note: Python printing/logging example: logging.info(message)
Destination: Driver logs, Azure Machine Learning designer Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-debug-pipelines

NEW QUESTION 68
- (Exam Topic 3)
You are evaluating a completed binary classification machine learning model. You need to use the precision as the valuation metric.
Which visualization should you use?

A. Binary classification confusion matrix
B. box plot
C. Gradient descent
D. coefficient of determination

Answer: A

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

NEW QUESTION 70
- (Exam Topic 3)
You have a dataset that includes home sales data for a city. The dataset includes the following columns.

Name	Description
Price	The sales price for the house.
Bedrooms	The number of bedrooms in the house.
Size	The size of the house in square feet.
HasGarage	A binary value indicating whether or not the house has a garage.
HomeType	The category of home, for example, apartment, townhouse, single-family home.

Each row in the dataset corresponds to an individual home sales transaction.
You need to use automated machine learning to generate the best model for predicting the sales price based on the features of the house.

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

Setting	Value
Prediction task	▼
	Classification
	Forecasting
	Regression
	Outlier
Target column	▼
	Price
	Bedrooms
	Size
	HasGarage
	HomeType

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Regression
Regression is a supervised machine learning technique used to predict numeric values. Box 2: Price
Reference:
https://docs.microsoft.com/en-us/learn/modules/create-regression-model-azure-machine-learning-designer

NEW QUESTION 75
- (Exam Topic 3)
You have a multi-class image classification deep learning model that uses a set of labeled photographs. You create the following code to select hyperparameter values when training the model.

```
from azureml.train.hyperdrive import BayesianParameterSampling

param_sampling = BayesianParametersSampling ({
 "learning_rate": uniform(0.01, 0.1),
 "batch_size": choice(16, 32, 64, 128)}
)
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.
NOTE: Each correct selection is worth one point.

	Yes	No
Hyperparameter combinations for the runs are selected based on how previous samples performed in the previous experiment run.	○	○
The learning rate value 0.09 might be used during model training.	○	○
You can define an early termination policy for this hyperparameter tuning run.	○	○

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Yes
Hyperparameters are adjustable parameters you choose to train a model that govern the training process itself. Azure Machine Learning allows you to automate hyperparameter exploration in an efficient manner, saving you significant time and resources. You specify the range of hyperparameter values and a maximum number of training runs. The system then automatically launches multiple simultaneous runs with different parameter configurations and finds the configuration that results in the best performance, measured by the metric you choose. Poorly performing training runs are automatically early terminated, reducing wastage of compute resources. These resources are instead used to explore other hyperparameter configurations.
Box 2: Yes
uniform(low, high) - Returns a value uniformly distributed between low and high Box 3: No
Bayesian sampling does not currently support any early termination policy. Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters

NEW QUESTION 80
- (Exam Topic 3)
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. Create a real-time inference pipeline from the training pipeline.
D. Replace the dataset in the training pipeline with an Enter Data Manually module.

Answer: C

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.
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

NEW QUESTION 82
- (Exam Topic 3)
You create an Azure Machine Learning workspace and a new Azure DevOps organization. You register a
model in the workspace and deploy the model to the target environment.
All new versions of the model registered in the workspace must automatically be deployed to the target environment.
You need to configure Azure Pipelines to deploy the model.
Which four actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Graphical user interface, text, application, email Description automatically generated
Step 1: Create an Azure DevOps project
Step 2: Create a release pipeline
➤ Sign in to your Azure DevOps organization and navigate to your project.
➤ Go to Pipelines, and then select New pipeline.
Step 3: Install the Machine Learning extension for Azure Pipelines
You must install and configure the Azure CLI and ML extension.
Step 4: Create a service connection
How to set up your service connection
Graphical user interface, text, application, email Description automatically generated
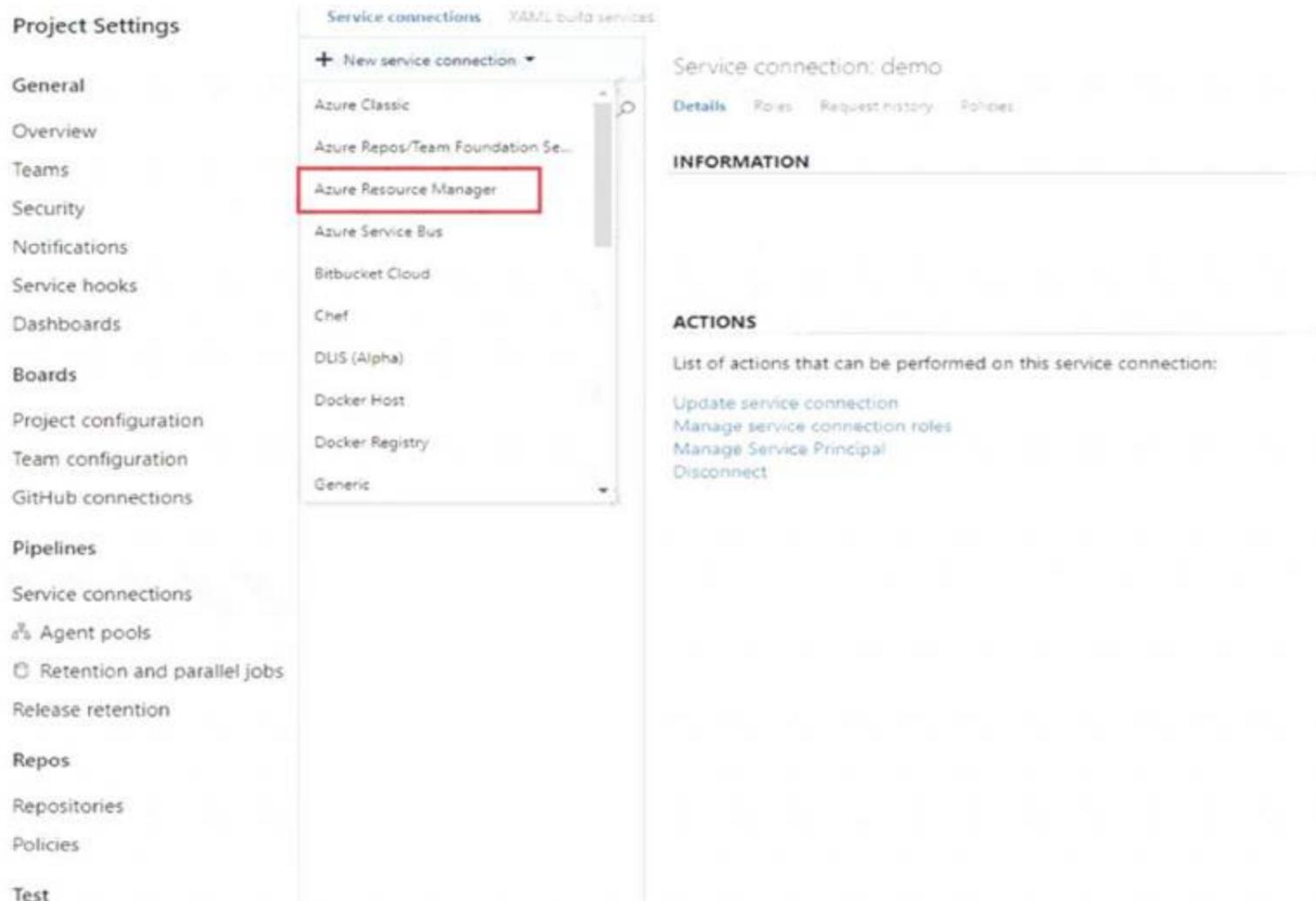
Select AzureMLWorkspace for the scope level, then fill in the following subsequent parameters. Graphical user interface, text, application Description automatically generated

Note: How to enable model triggering in a release pipeline

⟩ Go to your release pipeline and add a new artifact. Click on AzureML Model artifact then select the appropriate AzureML service connection and select from the available models in your workspace.

⟩ Enable the deployment trigger on your model artifact as shown here. Every time a new version of that model is registered, a release pipeline will be triggered.
Reference:
https://marketplace.visualstudio.com/items?itemName=ms-air-aiagility.vss-services-azureml https://docs.microsoft.com/en-us/azure/devops/pipelines/targets/azure-machine-learning

NEW QUESTION 86
- (Exam Topic 3)
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. Register the model with appropriate tags and properties.
C. Create a Conda environment for the web service compute and install the necessary Python packages.
D. Load the model and use it to predict labels from input data.
E. Specify the number of cores and the amount of memory required for the inference compute.

Answer: D

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

NEW QUESTION 89
- (Exam Topic 3)
You create a Python script that runs a training experiment in Azure Machine Learning. The script uses the Azure Machine Learning SDK for Python.
You must add a statement that retrieves the names of the logs and outputs generated by the script. You need to reference a Python class object from the SDK for the statement.
Which class object should you use?

A. Run
B. ScripcRunConfig
C. Workspace
D. Experiment

Answer: A

Explanation:
A run represents a single trial of an experiment. Runs are used to monitor the asynchronous execution of a trial, log metrics and store output of the trial, and to analyze results and access artifacts generated by the trial.
The run Class get_all_logs method downloads all logs for the run to a directory. Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run(class)

NEW QUESTION 90
- (Exam Topic 3)
You create a datastore named training_data that references a blob container in an Azure Storage account. The blob container contains a folder named csv_files in which multiple comma-separated values (CSV) files are stored.
You have a script named train.py in a local folder named ./script that you plan to run as an experiment using an estimator. The script includes the following code to read data from the csv_files folder:

```
import os
import argparse
import pandas as pd

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from azureml.core import Run

run = Run.get_context()
parser = argparse.ArgumentParser()
parser.add_argument('--data-folder', type=str, dest='data_folder', help='data reference')
args = parser.parse_args()

data_folder = args.data_folder
csv_files = os.listdir(data_folder)
training_data = pd.concat((pd.read_csv(os.path.join(data_folder,csv_file)) for csv_file in csv_files))

# Code goes on to split the training data and train a logistic regression model
```

You have the following script.

```
from azureml.core import Workspace, Datastore, Experiment
from azureml.train.sklearn import SKLearn

ws = Workspace.from_config()
exp = Experiment(workspace=ws, name='csv_training')
ds = Datastore.get(ws, datastore_name='training_data')
data_ref = ds.path('csv_files')

# Code to define estimator goes here

run = exp.submit(config=estimator)
run.wait_for_completion(show_output=True)
```

You need to configure the estimator for the experiment so that the script can read the data from a data reference named data_ref that references the csv_files folder in the training_data datastore.
Which code should you use to configure the estimator?

```
A. estimator = SKLearn(source_directory='./script',
       inputs=[data_ref.as_named_input('data-folder').to_pandas_dataframe()],
       compute_target='local',
       entry_script='train.py')

B. script_params = {
         '--data-folder': data_ref.as_mount()
    }
    estimator = SKLearn(source_directory='./script',
       script_params=script_params,
       compute_target='local',
       entry_script='train.py'

C. estimator = SKLearn(source_directory='./script',
       inputs=[data_ref.as_named_input('data-folder').as_mount()],
       compute_target='local',
       entry_script='train.py')

D. script_params = {
         '--data-folder': data_ref.as_download(path_on_compute='csv_files')
    }
    estimator = SKLearn(source_directory='./script',
       script_params=script_params,
       compute_target='local',
       entry_script='train.py'

E. estimator = SKLearn(source_directory='./script',
       inputs=[data_ref.as_named_input('data-folder').as_download(path_on_compute='csv_files')],
       compute_target='local',
       entry_script='train.py')
```

A. Option A
B. Option B
C. Option C
D. Option D
E. Option E

Answer: B

Explanation:
Besides passing the dataset through the inputs parameter in the estimator, you can also pass the dataset through script_params and get the data path (mounting point) in your training script via arguments. This way, you can keep your training script independent of azureml-sdk. In other words, you will be able use the same training script for local debugging and remote training on any cloud platform.
Example:
from azureml.train.sklearn import SKLearn script_params = {
# mount the dataset on the remote compute and pass the mounted path as an argument to the training script '--data-folder':
mnist_ds.as_named_input('mnist').as_mount(),
'--regularization': 0.5
}
est = SKLearn(source_directory=script_folder, script_params=script_params, compute_target=compute_target, environment_definition=env, entry_script='train_mnist.py')
# Run the experiment
run = experiment.submit(est) run.wait_for_completion(show_output=True) Reference:
https://docs.microsoft.com/es-es/azure/machine-learning/how-to-train-with-datasets

NEW QUESTION 95
- (Exam Topic 3)
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.

```
# 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)
```

A. Mastered
B. Not Mastered

Answer: A

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.
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

NEW QUESTION 100
- (Exam Topic 3)
A coworker registers a datastore in a Machine Learning services workspace by using the following code:

```
Datastore.register_azure_blob_container(workspace=ws,
    datastore_name='demo_datastore',
    container_name='demo_datacontainer',
    account_name='demo_account',
    account_key='0A0A0A-0A0A00A-0A00A0A0A0A0A',
    create_if_not_exists=True)
```

You need to write code to access the datastore from a notebook.

Answer Area

```
import azureml.core
from azureml.core import Workspace, Datastore
ws = Workspace.from_config()

datastore =    [Workspace / Datastore / Experiment / Run]    .get(    [ws / run / experiment / log]    ,    '    [demo_datastore / demo_datacontainer / demo_account / Datastore]
```
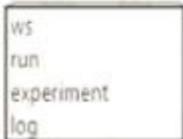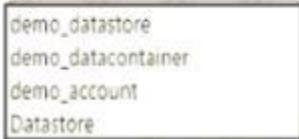
A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: DataStore
To get a specific datastore registered in the current workspace, use the get() static method on the Datastore class:
# Get a named datastore from the current workspace
datastore = Datastore.get(ws, datastore_name='your datastore name') Box 2: ws
Box 3: demo_datastore Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-access-data

NEW QUESTION 102
- (Exam Topic 3)
You are building a machine learning model for translating English language textual content into French language textual content.
You need to build and train the machine learning model to learn the sequence of the textual content. Which type of neural network should you use?

A. Multilayer Perceptions (MLPs)
B. Convolutional Neural Networks (CNNs)
C. Recurrent Neural Networks (RNNs)
D. Generative Adversarial Networks (GANs)

Answer: C

Explanation:
To translate a corpus of English text to French, we need to build a recurrent neural network (RNN).
Note: RNNs are designed to take sequences of text as inputs or return sequences of text as outputs, or both. They're called recurrent because the network's hidden layers have a loop in which the output and cell state from each time step become inputs at the next time step. This recurrence serves as a form of memory. It allows contextual information to flow through the network so that relevant outputs from previous time steps can be applied to network operations at the current time step.
References:
https://towardsdatascience.com/language-translation-with-rnns-d84d43b40571

NEW QUESTION 104
- (Exam Topic 3)
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 explanation for the registered model in Azure ML studio.
B. Enable Azure Application Insights for the service endpoint and view logged data in the Azure portal.
C. Create an ML Flow tracking URI that references the endpoint, and view the data logged by ML Flow.
D. View the log files generated by the experiment used to train the model.

Answer: B

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:
* 1. Sign in to the studio at https://ml.azure.com.
* 2. Go to Models and select the model you want to deploy.
* 3. Select +Deploy.
* 4. Populate the Deploy model form.

* 5. Expand the Advanced menu.
* 6. 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
  ┌──────────────────────────────────────────────┐
  │ 1                                            │
  └──────────────────────────────────────────────┘

CPU reserve capacity ⓘ
  ┌──────────────────────────────────────────────┐
  │ 0.1                                          │
  └──────────────────────────────────────────────┘

Memory reserve capacity ⓘ
  ┌──────────────────────────────────────────────┐
  │ 0.5                                          │
  └──────────────────────────────────────────────┘


                              Deploy        Cancel
```

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

NEW QUESTION 109
- (Exam Topic 3)
You are the owner of an Azure Machine Learning workspace.
You must prevent the creation or deletion of compute resources by using a custom role. You must allow all other operations inside the workspace.
You need to configure the custom role.
How should you complete the configuration? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area

```
{
  "Name": "Data Scientist Custom",
  "IsCustom": true,
  "Description": "Description"
  "Actions": [
    ┌────────────────────────────────────────────────────────────────────▼─┐
    │ Microsoft.MachineLearningServices/workspaces/*/read                   │
    │ Microsoft.MachineLearningServices/workspaces/computes/*/write         │
    │ Microsoft.MachineLearningServices/workspaces/delete                   │
    └───────────────────────────────────────────────────────────────────────┘
    ┌────────────────────────────────────────────────────────────────────▼─┐
    │ Microsoft.MachineLearningServices/workspaces/*/write                  │
    │ Microsoft.MachineLearningServices/workspaces/computes/*/write         │
    │ Microsoft.MachineLearningServices/workspaces/delete                   │
    └───────────────────────────────────────────────────────────────────────┘
  ],
  "NotActions": [
    ┌────────────────────────────────────────────────────────────────────▼─┐
    │ Microsoft.MachineLearningServices/workspaces/*/read                   │
    │ Microsoft.MachineLearningServices/workspaces/*/write                  │
    │ Microsoft.MachineLearningServices/workspaces/computes/*/delete        │
    └───────────────────────────────────────────────────────────────────────┘
    ┌────────────────────────────────────────────────────────────────────▼─┐
    │ Microsoft.MachineLearningServices/workspaces/*/read                   │
    │ Microsoft.MachineLearningServices/workspaces/*/write                  │
    │ Microsoft.MachineLearningServices/workspaces/computes/*/write         │
    └───────────────────────────────────────────────────────────────────────┘
  ],
  "AssignableScopes": [
    "/subscriptions/<subscription_id>"
  ]
}
```

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Graphical user interface, application Description automatically generated
Graphical user interface, application Description automatically generated
Box 1: Microsoft.MachineLearningServices/workspaces/*/read
Reader role: Read-only actions in the workspace. Readers can list and view assets, including datastore credentials, in a workspace. Readers can't create or update these assets.
Box 2: Microsoft.MachineLearningServices/workspaces/*/write
If the roles include Actions that have a wildcard (*), the effective permissions are computed by subtracting the NotActions from the allowed Actions.
Box 3: Box 2: Microsoft.MachineLearningServices/workspaces/computes/*/delete
Box 4: Microsoft.MachineLearningServices/workspaces/computes/*/write Reference:
https://docs.microsoft.com/en-us/azure/role-based-access-control/overview#how-azure-rbac-determines-if-a-use

NEW QUESTION 110
- (Exam Topic 3)
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 DB-cluster, configure the Advanced Logging option.
B. From DB-workspac
C. configure the Link Azure ML workspace option.
D. From ML-workspac
E. create an attached compute.
F. From ML-workspac
G. create a compute cluster.

Answer: B

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

NEW QUESTION 112
- (Exam Topic 3)
You create and register a model in an Azure Machine Learning workspace.
You must use the Azure Machine Learning SDK to implement a batch inference pipeline that uses a ParallelRunStep to score input data using the model. You must specify a value for the ParallelRunConfig compute_target setting of the pipeline step.
You need to create the compute target. Which class should you use?

A. BatchCompute
B. AdlaCompute
C. AmlCompute
D. Aks Compute

Answer: C

Explanation:
Compute target to use for ParallelRunStep. This parameter may be specified as a compute target object or the string name of a compute target in the workspace.
The compute_target target is of AmlCompute or string.
Note: An Azure Machine Learning Compute (AmlCompute) 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
Reference:
https://docs.microsoft.com/en-us/python/api/azureml-contrib-pipeline-steps/azureml.contrib.pipeline.steps.parall https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.compute.amlcompute(class)

NEW QUESTION 116
- (Exam Topic 3)
You are using a Git repository to track work in an Azure Machine Learning workspace. You need to authenticate a Git account by using SSH.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

Actions

Answer Area

Generate a public/private key pair

Add the private key to the Git account

Clone the Git repository by using an SSH repository URL

Add the public key to the Git account

Create a new Azure Key Vault resource

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Graphical user interface, text, application, chat or text message Description automatically generated
Authenticate your Git Account with SSH: Step 1: Generating a public/private key pair Generate a new SSH key
* 1. Open the terminal window in the Azure Machine Learning Notebook Tab.
* 2. Paste the text below, substituting in your email address. ssh-keygen -t rsa -b 4096 -C "your_email@example.com"
This creates a new ssh key, using the provided email as a label.
> Generating public/private rsa key pair.
Step 2: Add the public key to the Git Account
In your terminal window, copy the contents of your public key file. Step 3: Clone the Git repository by using an SSH repository URL
* 1. Copy the SSH Git clone URL from the Git repo.
* 2. Paste the url into the git clone command below, to use your SSH Git repo URL. This will look something like:
git clone git@example.com:GitUser/azureml-example.git Cloning into 'azureml-example'.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/concept-train-model-git-integration

NEW QUESTION 119
- (Exam Topic 3)
You are performing clustering by using the K-means algorithm. You need to define the possible termination conditions.
Which three conditions can you use? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

A. A fixed number of iterations is executed.
B. The residual sum of squares (RSS) rises above a threshold.
C. The sum of distances between centroids reaches a maximum.
D. The residual sum of squares (RSS) falls below a threshold.
E. Centroids do not change between iterations.

Answer: ADE

Explanation:
 References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/k-means-clustering https://nlp.stanford.edu/IR-book/html/htmledition/k-means-1.html

NEW QUESTION 124
- (Exam Topic 3)
You plan to create a speech recognition deep learning model. The model must support the latest version of Python.
You need to recommend a deep learning framework for speech recognition to include in the Data Science Virtual Machine (DSVM).
What should you recommend?

A. Apache Drill
B. Tensorflow
C. Rattle
D. Weka

Answer: B

Explanation:
TensorFlow is an open source library for numerical computation and large-scale machine learning. It uses Python to provide a convenient front-end API for building applications with the framework
TensorFlow can train and run deep neural networks for handwritten digit classification, image recognition, word embeddings, recurrent neural networks, sequence-to-sequence models for machine translation, natural language processing, and PDE (partial differential equation) based simulations.
References:
https://www.infoworld.com/article/3278008/what-is-tensorflow-the-machine-learning-library-explained.html

NEW QUESTION 128
- (Exam Topic 3)
You plan to deliver a hands-on workshop to several students. The workshop will focus on creating data visualizations using Python. Each student will use a device that has internet access.
Student devices are not configured for Python development. Students do not have administrator access to install software on their devices. Azure subscriptions are not available for students.
You need to ensure that students can run Python-based data visualization code. Which Azure tool should you use?

A. Anaconda Data Science Platform
B. Azure BatchAI
C. Azure Notebooks
D. Azure Machine Learning Service

Answer: C

Explanation:
 References:
https://notebooks.azure.com/

NEW QUESTION 133
- (Exam Topic 3)
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 analyzing a numerical dataset which contains missing values in several columns.
You must clean the missing values using an appropriate operation without affecting the dimensionality of the feature set.
You need to analyze a full dataset to include all values.
Solution: Replace each missing value using the Multiple Imputation by Chained Equations (MICE) method. Does the solution meet the goal?

A. Yes
B. NO

Answer: A

Explanation:
Replace using MICE: For each missing value, this option assigns a new value, which is calculated by using a method described in the statistical literature as "Multivariate Imputation using Chained Equations" or "Multiple Imputation by Chained Equations". With a multiple imputation method, each variable with missing data is modeled conditionally using the other variables in the data before filling in the missing values.
Note: Multivariate imputation by chained equations (MICE), sometimes called "fully conditional specification" or "sequential regression multiple imputation" has emerged in the statistical literature as one principled method of addressing missing data. Creating multiple imputations, as opposed to single imputations, accounts for the statistical uncertainty in the imputations. In addition, the chained equations approach is very flexible and can handle variables of varying types (e.g., continuous or binary) as well as complexities such as bounds or survey skip patterns.
References: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data

NEW QUESTION 138
- (Exam Topic 3)
You plan to run a Python script as an Azure Machine Learning experiment.
The script must read files from a hierarchy of folders. The files will be passed to the script as a dataset argument.
You must specify an appropriate mode for the dataset argument.
Which two modes can you use? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

A. to_pandas_dataframe ()
B. as_download()
C. as_upload()
D. as mount ()

Answer: B

Explanation:
Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.data.filedataset?view=azure-ml-py

NEW QUESTION 142
- (Exam Topic 3)
You are creating an experiment by using Azure Machine Learning Studio.
You must divide the data into four subsets for evaluation. There is a high degree of missing values in the data. You must prepare the data for analysis.
You need to select appropriate methods for producing the experiment.
Which three modules should you run in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.
NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.

Actions

Answer Area

Build Counting Transform

Missing Values Scrubber

Feature Hashing

Clean Missing Data

Replace Discrete Values

Import Data

Latent Dirichlet Transformation

Partition and Sample

A. Mastered
B. Not Mastered

Answer: A

Explanation:
The Clean Missing Data module in Azure Machine Learning Studio, to remove, replace, or infer missing values.

NEW QUESTION 143
- (Exam Topic 3)
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 have a Python script named train.py in a local folder named scripts. The script trains a regression model by using scikit-learn. The script includes code to load a training data file which is also located in the scripts folder.
You must run the script as an Azure ML experiment on a compute cluster named aml-compute.
You need to configure the run to ensure that the environment includes the required packages for model training. You have instantiated a variable named aml-compute that references the target compute cluster.
Solution: Run the following code:

```
from azureml.train.estimator import Estimator
sk_est = Estimator(source_directory='./scripts',
  compute_target=aml-compute,
  entry_script='train.py',
  conda_packages=['scikit-learn'])
```

Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:
The scikit-learn estimator provides a simple way of launching a scikit-learn training job on a compute target. It is implemented through the SKLearn class, which can be used to support single-node CPU training.
Example:
from azureml.train.sklearn import SKLearn
}
estimator = SKLearn(source_directory=project_folder, compute_target=compute_target, entry_script='train_iris.py'
)
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-train-scikit-learn

NEW QUESTION 146
- (Exam Topic 3)
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 AksWebservice 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

Answer: A

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.github.io/azureml-sdk-for-r/articles/deploying-models.html

NEW QUESTION 151
- (Exam Topic 3)
You register a file dataset named csvjolder that references a folder. The folder includes multiple com
ma-separated values (CSV) files in an Azure storage blob container. You plan to use the following code to run a script that loads data from the file dataset. You
create and instantiate the following variables:

Variable	Description
remote_cluster	References the Azure Machine Learning compute cluster
ws	References the Azure Machine Learning workspace

You have the following code:

```
from azureml.train.estimator import Estimator
file_dataset = ws.datasets.get('csv_folder')
estimator = Estimator(source_directory=script_folder,

    compute_target = remote_cluster,
    entry_script='script.py')
run = experiment.submit(config=estimator)
run.wait_for_completion(show_output=True)
```

You need to pass the dataset to ensure that the script can read the files it references. Which code segment should you insert to replace the code comment?
A)
```
inputs=[file_dataset.as_named_input('training_files').to_pandas_dataframe()],
```

B)
```
inputs=[file_dataset.as_named_input('training_files').as_mount()],
```

C)
```
script_params={'--training_files': file_dataset},
```

D)
```
inputs=[file_dataset.as_named_input('training_files')],
```

A. Option A
B. Option B
C. Option C
D. Option D

Answer: D

Explanation:
 Example:
from azureml.train.estimator import Estimator script_params = {
# to mount files referenced by mnist dataset
'--data-folder': mnist_file_dataset.as_named_input('mnist_opendataset').as_mount(),
'--regularization': 0.5
}
est = Estimator(source_directory=script_folder, script_params=script_params, compute_target=compute_target, environment_definition=env,
entry_script='train.py')
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-train-models-with-aml

NEW QUESTION 152
- (Exam Topic 3)
You are developing a data science workspace that uses an Azure Machine Learning service. You need to select a compute target to deploy the workspace.
What should you use?

A. Azure Data Lake Analytics
B. Azure Databrick .

C. Apache Spark for HDInsight.
D. Azure Container Service

Answer: D

Explanation:
Azure Container Instances can be used as compute target for testing or development. Use for low-scale CPU-based workloads that require less than 48 GB of RAM.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/service/how-to-deploy-and-where

NEW QUESTION 153
- (Exam Topic 3)
You need to select a pre built development environment for a series of data science experiments. You must use the R language for the experiments.
Which three environments can you use? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

A. MI.NET Library on a local environment
B. Azure Machine Learning Studio
C. Data Science Virtual Machine (OSVM)
D. Azure Data bricks
E. Azure Cognitive Services

Answer: ABD

NEW QUESTION 156
- (Exam Topic 3)
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 have a Python script named train.py in a local folder named scripts. The script trains a regression model by using scikit-learn. The script includes code to load a training data file which is also located in the scripts folder.
You must run the script as an Azure ML experiment on a compute cluster named aml-compute.
You need to configure the run to ensure that the environment includes the required packages for model training. You have instantiated a variable named aml-compute that references the target compute cluster.
Solution: Run the following code:

```
from azureml.train.sklearn import SKLearn
sk_est = SKLearn(source_directory='./scripts',
  compute_target=aml-compute,
  entry_script='train.py')
```

Does the solution meet the goal?

A. Yes
B. No

Answer: A

Explanation:
The scikit-learn estimator provides a simple way of launching a scikit-learn training job on a compute target. It is implemented through the SKLearn class, which can be used to support single-node CPU training.
Example:
from azureml.train.sklearn import SKLearn
}
estimator = SKLearn(source_directory=project_folder, compute_target=compute_target, entry_script='train_iris.py'
)
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-train-scikit-learn

NEW QUESTION 159
- (Exam Topic 3)
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

Answer: A

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: ZeroOneLoss(x,y) = 1 when x!=y; otherwise 0.

Coefficient of determination, often referred to as R2, 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 R2 values, as low values can be entirely normal and high values can be suspect.

AUC.

References:

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

NEW QUESTION 163
- (Exam Topic 3)
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

Answer: B

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

NEW QUESTION 165
- (Exam Topic 3)
You train and register a model by using the Azure Machine Learning SDK on a local workstation. Python 3.6 and Visual Studio Code are installed on the workstation.

When you try to deploy the model into production as an Azure Kubernetes Service (AKS)-based web service, you experience an error in the scoring script that causes deployment to fail.

You need to debug the service on the local workstation before deploying the service to production.

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

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Graphical user interface, text, application, email Description automatically generated

Step 1: Install Docker on the workstation
Prerequisites include having a working Docker installation on your local system. Build or download the dockerfile to the compute node.
Step 2: Create an AksWebservice deployment configuration and deploy the model to it
To deploy a model to Azure Kubernetes Service, create a deployment configuration that describes the compute resources needed.
# 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) service.wait_for_deployment(show_output = True)
print(service.state) print(service.get_logs())
Step 3: Create a LocalWebservice deployment configuration for the service and deploy the model to it
To deploy locally, modify your code to use LocalWebservice.deploy_configuration() to create a deployment configuration. Then use Model.deploy() to deploy the service.
Step 4: Debug and modify the scoring script as necessary. Use the reload() method of the service after each modification.
During local testing, you may need to update the score.py file to add logging or attempt to resolve any problems that you've discovered. To reload changes to the score.py file, use reload(). For example, the following code reloads the script for the service, and then sends data to it.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-azure-kubernetes-service https://docs.microsoft.com/en-us/azure/machine-learning/how-to-troubleshoot-deployment-local

---

NEW QUESTION 170
- (Exam Topic 3)
You are building a recurrent neural network to perform a binary classification. You review the training loss, validation loss, training accuracy, and validation accuracy for each training epoch.
You need to analyze model performance.
Which observation indicates that the classification model is over fitted?

A. The training loss .stays constant and the validation loss stays on a constant value and close to the training loss value when training the model.
B. The training loss increases while the validation loss decreases when training the model.
C. The training loss decreases while the validation loss increases when training the model.
D. The training loss stays constant and the validation loss decreases when training the model.

Answer: B

NEW QUESTION 173
- (Exam Topic 3)
You create a binary classification model using Azure Machine Learning Studio.
You must use a Receiver Operating Characteristic (RO C) curve and an F1 score to evaluate the model. You need to create the required business metrics.
How should you complete the experiment? To answer, select the appropriate options in the dialog box in the answer area.
NOTE: Each correct selection is worth one point.
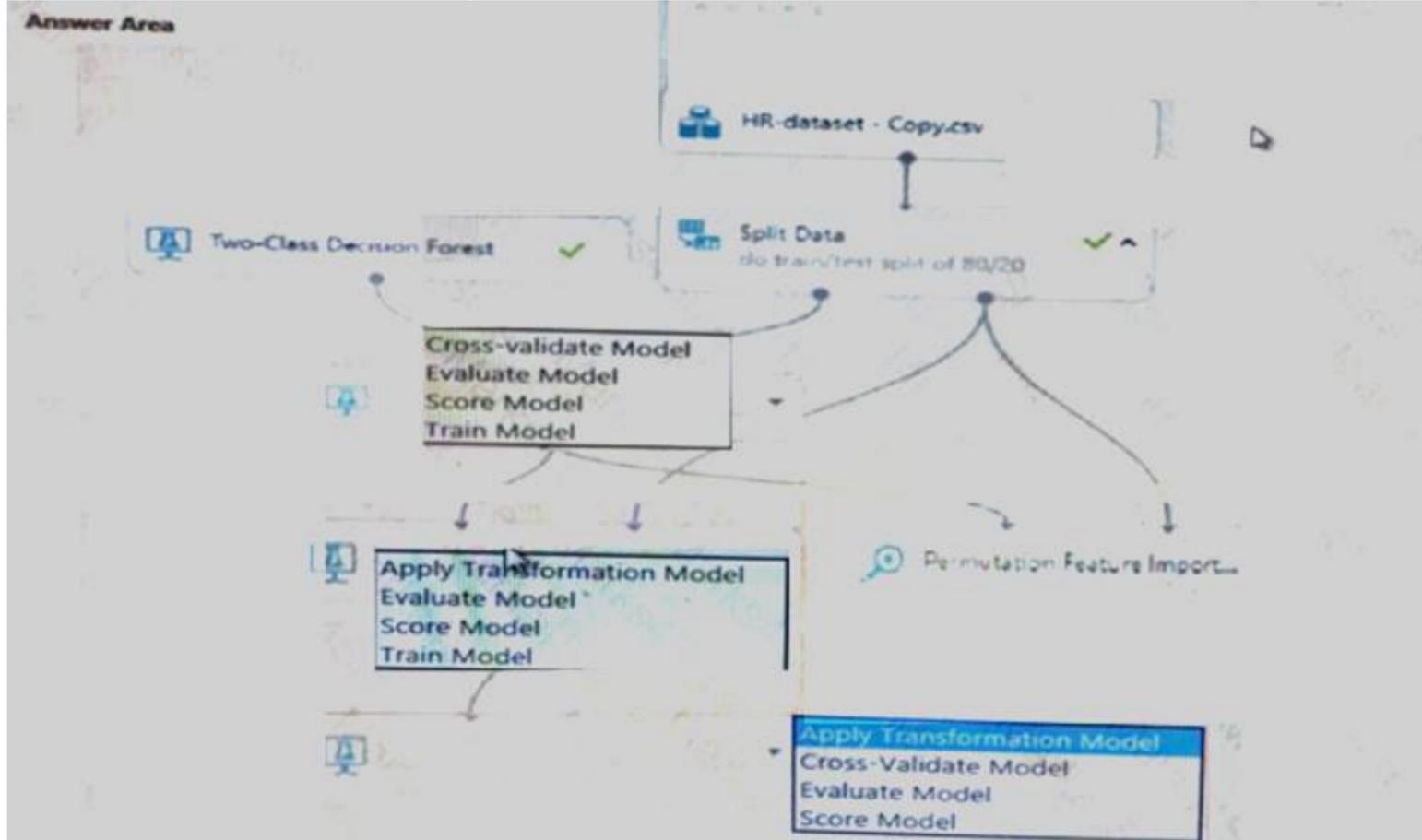
A. Mastered
B. Not Mastered

Answer: A

Explanation:
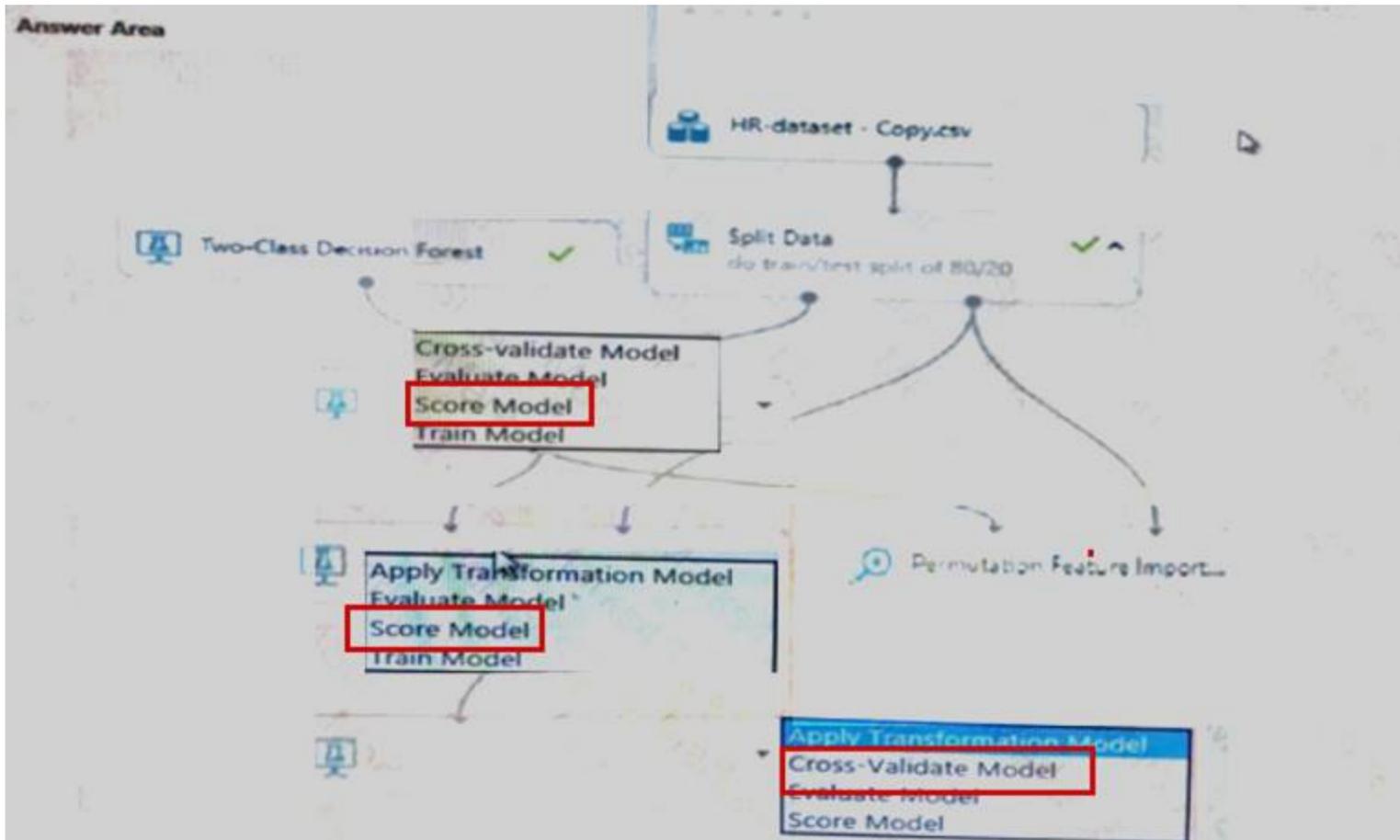
NEW QUESTION 176
- (Exam Topic 3)
You plan to explore demographic data for home ownership in various cities. The data is in a CSV file with the following format:
age,city,income,home_owner 21,Chicago,50000,0 35,Seattle,120000,1 23,Seattle,65000,0 45,Seattle,130000,1 18,Chicago,48000,0
You need to run an experiment in your Azure Machine Learning workspace to explore the data and log the results. The experiment must log the following information:

➤ the number of observations in the dataset

➤ a box plot of income by home_owner

➤ a dictionary containing the city names and the average income for each city

You need to use the appropriate logging methods of the experiment's run object to log the required information.
How should you complete the code? To answer, drag the appropriate code segments to the correct locations. Each code segment may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.

Code segments

| log |
| log_list |
| log_row |
| log_table |
| log_image |

Answer Area

```
from azureml.core import Experiment, Run
import pandas as pd
import matplotlib.pyplot as plt
# Create an Azure ML experiment in workspace
experiment = Experiment(workspace = ws, name = "demo-experiment")
# Start logging data from the experiment
run = experiment.start_logging()
# load the dataset
data = pd.read_csv('research/demographics.csv')
# Log the number of observations
row_count = (len(data))
run.  Segment  ("observations", row_count)
# Log box plot for income by home_owner
fig = plt.figure(figsize=(9, 6))
ax = fig.gca()
data.boxplot(column = 'income', by = "home_owner", ax = ax)
ax.set_title('income by home_owner')
ax.set_ylabel('income')
run.  Segment  (name = 'income_by_home_owner', plot = fig)
# Create a dataframe of mean income per city
mean_inc_df = data.groupby('city')['income'].agg(np.mean).to_frame().reset_index()
# Convert to a dictionary
mean_inc_dict = mean_inc_df.to_dict('dict')
# Log city names and average income dictionary
run.  Segment  (name="mean_income_by_city", value= mean_inc_dict)
# Complete tracking and get link to details
run.complete()
```

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: log
The number of observations in the dataset. run.log(name, value, description='')
Scalar values: Log a numerical or string value to the run with the given name. Logging a metric to a run causes that metric to be stored in the run record in the experiment. You can log the same metric multiple times within a run, the result being considered a vector of that metric.
Example: run.log("accuracy", 0.95)
Box 2: log_image
A box plot of income by home_owner.
log_image Log an image to the run record. Use log_image to log a .PNG image file or a matplotlib plot to the run. These images will be visible and comparable in the run record.
Example: run.log_image("ROC", plot=plt) Box 3: log_table
A dictionary containing the city names and the average income for each city. log_table: Log a dictionary object to the run with the given name.

NEW QUESTION 177
- (Exam Topic 3)
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.

```
# Assume required modules are imported

ws = Workspace.from_config()
feature_importances = explanation.[                ▼] ( workspace = ws,
                          from_run                  experiment_name='train_and_explain',
                          list_model_explanations   run_id='train_and_explain_12345')
                          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_important_dict
                          download_model_explanation

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

A. Mastered
B. Not Mastered

Answer: A

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 explanations Client.
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.

NEW QUESTION 181

- (Exam Topic 3)
HOTSPOT
You have an Azure blob container that contains a set of TSV files. The Azure blob container is registered as a datastore for an Azure Machine Learning service workspace. Each TSV file uses the same data schema.
You plan to aggregate data for all of the TSV files together and then register the aggregated data as a dataset in an Azure Machine Learning workspace by using the Azure Machine Learning SDK for Python.
You run the following code.

```
from azureml.core.workspace import Workspace
from azureml.core.datastore import Datastore
from azureml.core.dataset import Dataset
import pandas as pd
datastore_paths = (datastore, './data/*.tsv')
myDataset_1 = Dataset.File.from_files(path=datastore_paths)
myDataset_2 = Dataset.Tabular.from_delimited_files(path=datastore_paths, separator='\t')
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.
NOTE: Each correct selection is worth one point.

	Yes	No
The myDataset_1 dataset can be converted into a pandas dataframe by using the following method: using myDataset_1.to_pandas_dataframe()	O	O
The myDataset_1.to_path() method returns an array of file paths for all of the TSV files in the dataset.	O	O
The myDataset_2 dataset can be converted into a pandas dataframe by using the following method: myDataset_2.to_pandas_dataframe()	O	O

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: No
FileDataset references single or multiple files in datastores or from public URLs. The TSV files need to be parsed.
Box 2: Yes
to_path() gets a list of file paths for each file stream defined by the dataset. Box 3: Yes
TabularDataset.to_pandas_dataframe loads all records from the dataset into a pandas DataFrame. TabularDataset represents data in a tabular format created by parsing the provided file or list of files.
Note: TSV is a file extension for a tab-delimited file used with spreadsheet software. TSV stands for Tab Separated Values. TSV files are used for raw data and can be imported into and exported from spreadsheet software. TSV files are essentially text files, and the raw data can be viewed by text editors, though they are often used when moving raw data between spreadsheets.
Reference:
https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.data.tabulardataset

NEW QUESTION 184
- (Exam Topic 3)
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.get.metrics()
B. pipeline_run.wait_for_completion(show_output=True)
C. pipeline_param = PipelineParameter(name="stdout", default_value="console")
D. pipeline_run.get_status()

Answer: B

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.

NEW QUESTION 187
- (Exam Topic 3)
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. cags
D. properties

Answer: D

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

NEW QUESTION 191
- (Exam Topic 3)
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 using Azure Machine Learning to run an experiment that trains a classification model.
You want to use Hyperdrive to find parameters that optimize the AUC metric for the model. You configure a HyperDriveConfig for the experiment by running the following code:

```
hyperdrive = HyperDriveConfig(estimator=your_estimator,
    hyperparameter_sampling=your_params,
    policy=policy,
    primary_metric_name='AUC',
    primary_metric_goal=PrimaryMetricGoal.MAXIMIZE,
    max_total_runs=6,
    max_concurrent_runs=4)
```

variable named y_test variable, and the predicted probabilities from the model are stored in a variable named y_predicted. You need to add logging to the script to allow Hyperdrive to optimize hyperparameters for the AUC metric. Solution: Run the following code:

```
from sklearn.metrics import roc_auc_score
import logging
# code to train model omitted
auc = roc_auc_score(y_test, y_predicted)
logging.info("AUC: " + str(auc))
```

Does the solution meet the goal?

A. Yes
B. No

Answer: A

Explanation:
Python printing/logging example: logging.info(message)
Destination: Driver logs, Azure Machine Learning designer
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-debug-pipelines

NEW QUESTION 195
- (Exam Topic 3)
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 a different name for the model each time you register it.
B. 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.
C. Specify the model framework version when registering the model, and only register subsequent models if this value is higher.
D. 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 thecurrently registered model.
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 currentlyregistered mode

Answer: CE

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/

NEW QUESTION 196
- (Exam Topic 3)
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 create a model to forecast weather conditions based on historical data.
You need to create a pipeline that runs a processing script to load data from a datastore and pass the processed data to a machine learning model training script.
Solution: Run the following code:

```
datastore = ws.get_default_datastore()
data_output = pd.read_csv("traindata.csv")
process_step = PythonScriptStep(script_name="process.py",
  arguments=["--data_for_train", data_output],
  outputs=[data_output],compute_target=aml_compute,
  source_directory=process_directory)
train_step = PythonScriptStep(script_name="train.py",
  arguments=["--data_for_train", data_output],
  inputs=[data_output],compute_target=aml_compute,
  source_directory=train_directory)
pipeline = Pipeline(workspace=ws, steps=[process_step, train_step])
```

Does the solution meet the goal?

A. Yes
B. No

Answer: A

Explanation:
The two steps are present: process_step and train_step Note:
Data used in pipeline can be produced by one step and consumed in another step by providing a PipelineData object as an output of one step and an input of one or more subsequent steps.
PipelineData objects are also used when constructing Pipelines to describe step dependencies. To specify that a step requires the output of another step as input, use a PipelineData object in the constructor of both steps.
For example, the pipeline train step depends on the process_step_output output of the pipeline process step: from azureml.pipeline.core import Pipeline, PipelineData
from azureml.pipeline.steps import PythonScriptStep datastore = ws.get_default_datastore()
process_step_output = PipelineData("processed_data", datastore=datastore) process_step = PythonScriptStep(script_name="process.py",
arguments=["--data_for_train", process_step_output], outputs=[process_step_output], compute_target=aml_compute, source_directory=process_directory)
train_step = PythonScriptStep(script_name="train.py", arguments=["--data_for_train", process_step_output], inputs=[process_step_output],
compute_target=aml_compute, source_directory=train_directory)
pipeline = Pipeline(workspace=ws, steps=[process_step, train_step]) Reference:
https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.pipelinedata?view=azu

NEW QUESTION 197
- (Exam Topic 3)
You use the following code to define the steps for a pipeline: from azureml.core import Workspace, Experiment, Run from azureml.pipeline.core import Pipeline
from azureml.pipeline.steps import PythonScriptStep ws = Workspace.from_config()
. . .
step1 = PythonScriptStep(name="step1", ...) step2 = PythonScriptsStep(name="step2", ...) pipeline_steps = [step1, step2]
You need to add code to run the steps.
Which two code segments can you use to achieve this goal? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

A. experiment = Experiment(workspace=ws, name='pipeline-experiment')run = experiment.submit(config=pipeline_steps)
B. run = Run(pipeline_steps)
C. pipeline = Pipeline(workspace=ws, steps=pipeline_steps) experiment = Experiment(workspace=ws, name='pipeline-experiment')run = experiment.submit(pipeline)
D. pipeline = Pipeline(workspace=ws, steps=pipeline_steps)run = pipeline.submit(experiment_name='pipeline-experiment')

Answer: CD

Explanation:
After you define your steps, you build the pipeline by using some or all of those steps.
# Build the pipeline. Example:
pipeline1 = Pipeline(workspace=ws, steps=[compare_models])
# Submit the pipeline to be run
pipeline_run1 = Experiment(ws, 'Compare_Models_Exp').submit(pipeline1) Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-machine-learning-pipelines

NEW QUESTION 198
- (Exam Topic 3)
You are analyzing a dataset by using Azure Machine Learning Studio.
YOU need to generate a statistical summary that contains the p value and the unique value count for each feature column.
Which two modules can you users? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

A. Execute Python Script
B. Export Count Table
C. Convert to Indicator Values
D. Summarize Data
E. Compute linear Correlation

Answer: BE

Explanation:
The Export Count Table module is provided for backward compatibility with experiments that use the Build Count Table (deprecated) and Count Featurizer (deprecated) modules.
E: Summarize Data statistics are useful when you want to understand the characteristics of the complete dataset. For example, you might need to know:
How many missing values are there in each column? How many unique values are there in a feature column?
What is the mean and standard deviation for each column?
The module calculates the important scores for each column, and returns a row of summary statistics for each variable (data column) provided as input.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/export-count-table https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/summarize-data

NEW QUESTION 199
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
An IT department creates the following Azure resource groups and resources:

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

The IT department creates an Azure Kubernetes Service (AKS)-based inference compute target named
aks-cluster in the Azure Machine Learning workspace.
You have a Microsoft Surface Book computer with a GPU. Python 3.6 and Visual Studio Code are installed. You need to run a script that trains a deep neural network (DNN) model and logs the loss and accuracy
metrics.
Solution: Attach the mlvm virtual machine as a compute target in the Azure Machine Learning workspace. Install the Azure ML SDK on the Surface Book and run Python code to connect to the workspace. Run the training script as an experiment on the mlvm remote compute resource.

A. Yes
B. No

Answer: A

Explanation:
Use the VM as a compute target.
Note: A compute target is a designated compute resource/environment where you run your training script or host your service deployment. This location may be your local machine or a cloud-based compute resource.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target

NEW QUESTION 202
- (Exam Topic 3)
You must store data in Azure Blob Storage to support Azure Machine Learning. You need to transfer the data into Azure Blob Storage.
What are three possible ways to achieve the goal? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

A. Bulk Insert SQL Query
B. AzCopy
C. Python script
D. Azure Storage Explorer
E. Bulk Copy Program (BCP)

Answer: BCD

Explanation:
You can move data to and from Azure Blob storage using different technologies: Azure Storage-Explorer
AzCopy Python SSIS
References:
https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/move-azure-blob

NEW QUESTION 205
- (Exam Topic 3)
You are analyzing the asymmetry in a statistical distribution.
The following image contains two density curves that show the probability distribution of two datasets.

Graph 1   Graph 2
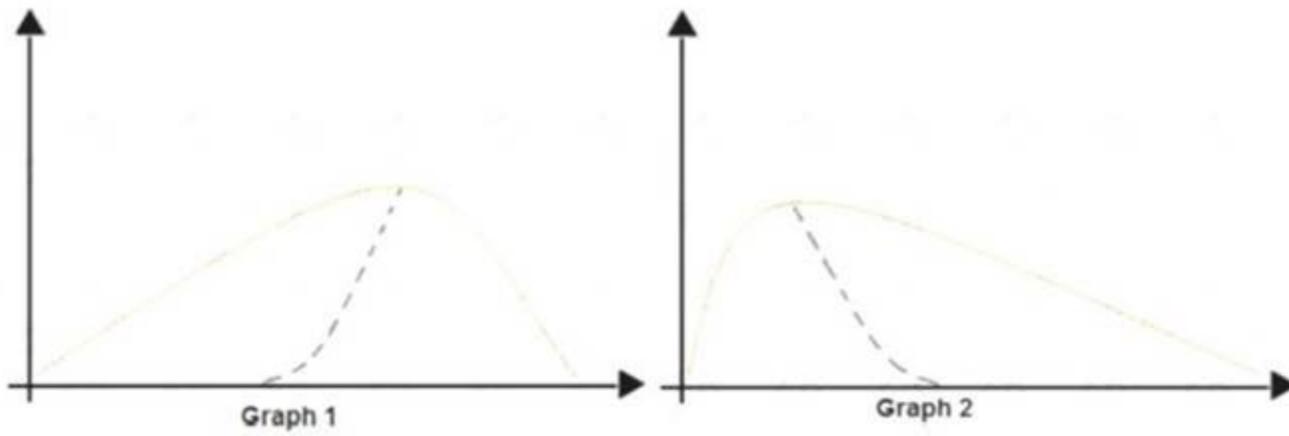
Use the drop-down menus to select the answer choice that answers each question based on the information presented in the graphic.
NOTE: Each correct selection is worth one point.

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

A. Mastered
B. Not Mastered

Answer: A

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. References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/compute-elementary-statistic

NEW QUESTION 210
- (Exam Topic 3)
You arc creating a new experiment in Azure Machine Learning Studio. You have a small dataset that has missing values in many columns. The data does not require the application of predictors for each column. You plan to use the Clean Missing Data module to handle the missing data.
You need to select a data cleaning method. Which method should you use?

A. Synthetic Minority
B. Replace using Probabilistic PAC
C. Replace using MICE
D. Normalization

Answer: B

NEW QUESTION 214
- (Exam Topic 3)
You are evaluating a Python NumPy array that contains six data points defined as follows: data = [10, 20, 30, 40, 50, 60]
You must generate the following output by using the k-fold algorithm implantation in the Python Scikit-learn machine learning library:
train: [10 40 50 60], test: [20 30]
train: [20 30 40 60], test: [10 50]
train: [10 20 30 50], test: [40 60]
You need to implement a cross-validation to generate the output.
How should you complete the code segment? To answer, select the appropriate code segment in the dialog box in the answer area.
NOTE: Each correct selection is worth one point.

```
from numpy import array
from sklearn.model_selection import    [▼]
                                        ┌──────────────────┐
                                        │ K-Means          │
                                        │ k-fold           │
                                        │ CrossValidation  │
                                        │ ModelSelection   │
                                        └──────────────────┘
data - array([10, 20, 30, 40, 50, 60])
kfold - Kfold(n_splits-   [▼]        ,shuffle - True, random_state-1)
                          ┌──────┐
                          │ 1    │
                          │ 2    │
                          │ 3    │
                          │ 6    │
                          └──────┘
for train, test in kFold, split(   [▼]         ):
                                    ┌──────────────┐
                                    │ data         │
                                    │ k-fold       │
                                    │ array        │
                                    │ train, test  │
                                    └──────────────┘
print('train: %s, test: %5' % (data[train], data[test])
```

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: k-fold
Box 2: 3
K-F olds cross-validator provides train/test indices to split data in train/test sets. Split dataset into k consecutive folds (without shuffling by default).
The parameter n_splits (int, default=3) is the number of folds. Must be at least 2. Box 3: data
Example: Example:
>>>
>>> from sklearn.model_selection import KFold
>>> X = np.array([[1, 2], [3, 4], [1, 2], [3, 4]])
>>> y = np.array([1, 2, 3, 4])
>>> kf = KFold(n_splits=2)
>>> kf.get_n_splits(X) 2
>>> print(kf)
KFold(n_splits=2, random_state=None, shuffle=False)
>>> for train_index, test_index in kf.split(X): print("TRAIN:", train_index, "TEST:", test_index) X_train, X_test = X[train_index], X[test_index] y_train, y_test = y[train_index], y[test_index] TRAIN: [2 3] TEST: [0 1]
TRAIN: [0 1] TEST: [2 3]
References:
https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html

NEW QUESTION 218
- (Exam Topic 3)
You are performing sentiment analysis using a CSV file that includes 12,000 customer reviews written in a short sentence format. You add the CSV file to Azure Machine Learning Studio and configure it as the starting point dataset of an experiment. You add the Extract N-Gram Features from Text module to the experiment to extract key phrases from the customer review column in the dataset.
You must create a new n-gram dictionary from the customer review text and set the maximum n-gram size to trigrams.
What should you select? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Properties   Project

Extract N-Gram Features from Text

Text column

Selected columns
Column type: String Feature

Launch column selector

Vocabulary mode

Create
ReadOnly
Update
Merge

N-Grams size

3
4
4,000
12,000

0

Weighting function

Minimum word length

3

Maximum word length

25

Minimum n-gram document absolu...

5

Maximum n-gram document ratio

1

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Vocabulary mode: Create
For Vocabulary mode, select Create to indicate that you are creating a new list of n-gram features. N-Grams size: 3
For N-Grams size, type a number that indicates the maximum size of the n-grams to extract and store. For example, if you type 3, unigrams, bigrams, and trigrams will be created.
Weighting function: Leave blank
The option, Weighting function, is required only if you merge or update vocabularies. It specifies how terms in the two vocabularies and their scores should be weighted against each other.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/extract-n-gram-features-from

NEW QUESTION 223
- (Exam Topic 3)
You are building recurrent neural network to perform a binary classification.
The training loss, validation loss, training accuracy, and validation accuracy of each training epoch has been provided. You need to identify whether the classification model is over fitted.
Which of the following is correct?

A. The training loss increases while the validation loss decreases when training the model.
B. The training loss decreases while the validation loss increases when training the model.
C. The training loss stays constant and the validation loss decreases when training the model.
D. The training loss .stays constant and the validation loss stays on a constant value and close to the training loss value when training the model.

Answer: B

Explanation:
An overfit model is one where performance on the train set is good and continues to improve, whereas performance on the validation set improves to a point and then begins to degrade.
References:
https://machinelearningmastery.com/diagnose-overfitting-underfitting-lstm-models/

NEW QUESTION 226
- (Exam Topic 3)
HOTSPOT
You collect data from a nearby weather station. You have a pandas dataframe named weather_df that includes the following data:

Temperature	Observation_time	Humidity	Pressure	Visibility	Days_since_last observation
74	2019/10/2 00:00	0.62	29.87	3	0.5
89	2019/10/2 12:00	0.70	28.88	10	0.5
72	2019/10/3 00:00	0.64	30.00	8	0.5
80	2019/10/3 12:00	0.66	29.75	7	0.5

The data is collected every 12 hours: noon and midnight.
You plan to use automated machine learning to create a time-series model that predicts temperature over the next seven days. For the initial round of training, you want to train a maximum of 50 different models.
You must use the Azure Machine Learning SDK to run an automated machine learning experiment to train these models.
You need to configure the automated machine learning run.
How should you complete the AutoMLConfig definition? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

```
automl_config = AutoMLConfig(task="▼ ",
                                    regression
                                    forecasting
                                    classification
                                    deep learning

    training_data=weather_df,
    label_column_name="▼ ",
                        humidity
                        pressure
                        visibility
                        temperature
                        days_since_last
                        observation_time

    time_column_name="▼ ",
                        humidity
                        pressure
                        visibility
                        temperature
                        days_since_last
                        observation_time

    max_horizon=▼ ,
                2
                6
                7
                12
                14
                50

    iterations=▼ ,
                2
                6
                7
                12
                14
                50

    iteration_timeout_minutes=5,
    primary_metric="r2_score")
```

A. Mastered

B. Not Mastered

Answer: A

Explanation:
Box 1: forcasting
Task: The type of task to run. Values can be 'classification', 'regression', or 'forecasting' depending on the type of automated ML problem to solve.
Box 2: temperature
The training data to be used within the experiment. It should contain both training features and a label column (optionally a sample weights column).
Box 3: observation_time
time_column_name: The name of the time column. This parameter is required when forecasting to specify the datetime column in the input data used for building the time series and inferring its frequency. This setting is being deprecated. Please use forecasting_parameters instead.
Box 4: 7
"predicts temperature over the next seven days"
max_horizon: The desired maximum forecast horizon in units of time-series frequency. The default value is 1. Units are based on the time interval of your training data, e.g., monthly, weekly that the forecaster should
predict out. When task type is forecasting, this parameter is required.
Box 5: 50
"For the initial round of training, you want to train a maximum of 50 different models."
Iterations: The total number of different algorithm and parameter combinations to test during an automated
ML experiment. Reference:
https://docs.microsoft.com/en-us/python/api/azureml-train-automl-client/azureml.train.automl.automlconfig.auto

NEW QUESTION 227
- (Exam Topic 3)
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 create an Azure Machine Learning service datastore in a workspace. The datastore contains the following files:
• /data/2018/Q1.csv
• /data/2018/Q2.csv
• /data/2018/Q3.csv
• /data/2018/Q4.csv
• /data/2019/Q1.csv
All files store data in the following format: id,f1,f2i
1,1.2,0
2,1,1,
1 3,2.1,0
You run the following code:

```
data_store = Datastore.register_azure_blob_container(workspace=ws,
  datastore_name='data_store',
  container_name='quarterly_data',
  account_name='companydata',
  account_key='NRPxk8duxbM3…'
  create_if_not_exists=False)
```

You need to create a dataset named training_data and load the data from all files into a single data frame by using the following code:

```
data_frame = training_data.to_pandas_dataframe()
```

Solution: Run the following code:

```
from azureml.core import Dataset
paths = [(data_store,'data/2018/*.csv'), (data_store,'data/2019/*.csv')]
training_data = Dataset.File.from_files(paths)
```

Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:
 Use two file paths.
Use Dataset.Tabular_from_delimeted, instead of Dataset.File.from_files as the data isn't cleansed. Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-register-datasets

NEW QUESTION 230
- (Exam Topic 3)
You plan to use the Hyperdrive feature of Azure Machine Learning to determine the optimal hyperparameter values when training a model.
You must use Hyperdrive to try combinations of the following hyperparameter values:
• learning_rate: any value between 0.001 and 0.1
• batch_size: 16, 32, or 64
You need to configure the search space for the Hyperdrive experiment.
Which two parameter expressions should you use? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

A. a choice expression for learning_rate
B. a uniform expression for learning_rate
C. a normal expression for batch_size
D. a choice expression for batch_size

E. a uniform expression for batch_size

Answer: BD

Explanation:
B: Continuous hyperparameters are specified as a distribution over a continuous range of values. Supported distributions include:
> uniform(low, high) - Returns a value uniformly distributed between low and high
D: Discrete hyperparameters are specified as a choice among discrete values. choice can be:
> one or more comma-separated values
> a range object
> any arbitrary list object Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters

NEW QUESTION 231
- (Exam Topic 3)
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 plan to use a Python script to run an Azure Machine Learning experiment. The script creates a reference to the experiment run context, loads data from a file, identifies the set of unique values for the label column, and completes the experiment run:
from azureml.core import Run import pandas as pd
run = Run.get_context()
data = pd.read_csv('data.csv') label_vals = data['label'].unique()
# Add code to record metrics here run.complete()
The experiment must record the unique labels in the data as metrics for the run that can be reviewed later.
You must add code to the script to record the unique label values as run metrics at the point indicated by the comment.
Solution: Replace the comment with the following code: for label_val in label_vals:
run.log('Label Values', label_val) Does the solution meet the goal?

A. Yes
B. No

Answer: A

Explanation:
The run_log function is used to log the contents in label_vals: for label_val in label_vals:
run.log('Label Values', label_val) Reference:
https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai

NEW QUESTION 236
- (Exam Topic 3)
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 plan to use a Python script to run an Azure Machine Learning experiment. The script creates a reference to the experiment run context, loads data from a file, identifies the set of unique values for the label column, and completes the experiment run:
from azureml.core import Run import pandas as pd
run = Run.get_context()
data = pd.read_csv('data.csv') label_vals = data['label'].unique()
# Add code to record metrics here run.complete()
The experiment must record the unique labels in the data as metrics for the run that can be reviewed later.
You must add code to the script to record the unique label values as run metrics at the point indicated by the comment.
Solution: Replace the comment with the following code:
run.upload_file('outputs/labels.csv', './data.csv') Does the solution meet the goal?

A. Yes
B. No

Answer: B

Explanation:
label_vals has the unique labels (from the statement label_vals = data['label'].unique()), and it has to be logged.
Note:
Instead use the run_log function to log the contents in label_vals: for label_val in label_vals:
run.log('Label Values', label_val) Reference:
https://www.element61.be/en/resource/azure-machine-learning-services-complete-toolbox-ai

NEW QUESTION 239
- (Exam Topic 3)
You have an Azure Machine Learning workspace named workspace1 that is accessible from a public endpoint. The workspace contains an Azure Blob storage datastore named store1 that represents a blob container in an Azure storage account named account1. You configure workspace1 and account1 to be accessible by using private endpoints in the same virtual network.
You must be able to access the contents of store1 by using the Azure Machine Learning SDK for Python. You must be able to preview the contents of store1 by using Azure Machine Learning studio.
You need to configure store1.
What should you do? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Requirement	Action
Access the contents of store1 by using the Azure Machine Learning SDK for Python.	Set store1 as the default datastore. Disable data validation for store1. Update authentication for store1. Regenerate the keys of account1.
Preview the contents of store1 by using Azure Machine Learning studio.	Set store1 as the default datastore. Disable data validation for store1. Update authentication for store1. Regenerate the keys of account1.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Text, table Description automatically generated
Box 1: Regenerate the keys of account1.
Azure Blob Storage support authentication through Account key or SAS token.
To authenticate your access to the underlying storage service, you can provide either your account key, shared access signatures (SAS) tokens, or service principal
Box 2: Update the authentication for store1.
For Azure Machine Learning studio users, several features rely on the ability to read data from a dataset; such as dataset previews, profiles and automated machine learning. For these features to work with storage behind virtual networks, use a workspace managed identity in the studio to allow Azure Machine Learning to access the storage account from outside the virtual network.
Note: Some of the studio's features are disabled by default in a virtual network. To re-enable these features, you must enable managed identity for storage accounts you intend to use in the studio.
The following operations are disabled by default in a virtual network:

> Preview data in the studio.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-access-data

NEW QUESTION 241
- (Exam Topic 3)
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. Yes
B. No

Answer: B

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

NEW QUESTION 246
- (Exam Topic 3)
You plan to use a Data Science Virtual Machine (DSVM) with the open source deep learning frameworks Caffe2 and Theano. You need to select a pre configured DSVM to support the framework.
What should you create?

A. Data Science Virtual Machine for Linux (CentOS)
B. Data Science Virtual Machine for Windows 2012
C. Data Science Virtual Machine for Windows 2016
D. Geo AI Data Science Virtual Machine with ArcGIS
E. Data Science Virtual Machine for Linux (Ubuntu)

Answer: E

NEW QUESTION 248

- (Exam Topic 3)
You are a data scientist building a deep convolutional neural network (CNN) for image classification. The CNN model you built 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. Reduce the amount of training data.
B. Add an additional dense layer with 64 input units
C. Add L1/L2 regularization.
D. Use training data augmentation
E. Add an additional dense layer with 512 input units.

Answer: AC

Explanation:
 References:
https://machinelearningmastery.com/how-to-reduce-overfitting-in-deep-learning-with-weight-regularization/
https://en.wikipedia.org/wiki/Convolutional_neural_network

NEW QUESTION 252
- (Exam Topic 3)
You write five Python scripts that must be processed in the order specified in Exhibit A – which allows the same modules to run in parallel, but will wait for modules with dependencies.
You must create an Azure Machine Learning pipeline using the Python SDK, because you want to script to create the pipeline to be tracked in your version control system. You have created five PythonScriptSteps and have named the variables to match the module names.
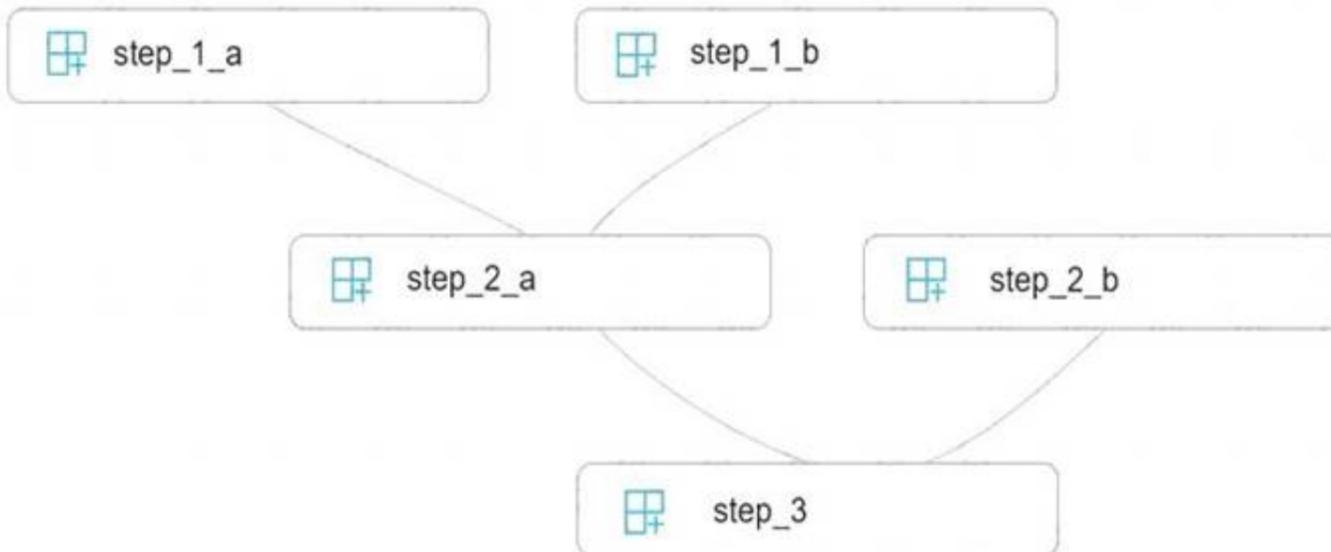
You need to create the pipeline shown. Assume all relevant imports have been done. Which Python code segment should you use?

```
A.  p = Pipeline(ws, steps=[[[[step_1_a, step_1_b], step_2_a],  step_2_b], step_3])

B.  pipeline_steps = {
        "Pipeline": {
            "run": step_3,
            "run_after": {[
                {"run": step_2_a,
                    "run_after":
                        [{"run": step_1_a},
                         {"run": step_1_b}]
                },
                {"run": step_2_b}]
            }
        }
    }
    p = Pipeline(ws, steps=pipeline_steps)

C.  step_2_a.run_after(step_1_b)
    step_2_a.run_after(step_1_a)
    step_3.run_after(step_2_b)
    step_3.run_after(step_2_a)
    p = Pipeline(ws, steps=[step_3])

D.  p = Pipeline(ws, steps=[step_1_a, step_1_b, step_2_a, step_2_b, step_3])
```

A. Option A
B. Option B
C. Option C
D. Option D

Answer: A

Explanation:
The steps parameter is an array of steps. To build pipelines that have multiple steps, place the steps in order in this array.
Reference:
https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-parallel-run-step

NEW QUESTION 256
- (Exam Topic 3)
You use Azure Machine Learning Studio to build a machine learning experiment. You need to divide data into two distinct datasets.
Which module should you use?

A. Partition and Sample
B. Assign Data to Clusters
C. Group Data into Bins
D. Test Hypothesis Using t-Test

Answer: A

Explanation:
Partition and Sample with the Stratified split option outputs multiple datasets, partitioned using the rules you specified.
References:
https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/partition-and-sample

NEW QUESTION 259
- (Exam Topic 3)
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

Answer: B

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

NEW QUESTION 262
......

Thank You for Trying Our Product

We offer two products:

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

2nd - Questons and Answers in PDF Format

DP-100 Practice Exam Features:

* DP-100 Questions and Answers Updated Frequently

* DP-100 Practice Questions Verified by Expert Senior Certified Staff

* DP-100 Most Realistic Questions that Guarantee you a Pass on Your FirstTry

* DP-100 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

100% Actual & Verified — Instant Download, Please Click
Order The DP-100 Practice Test Here