Question #: 2

Topic #: 1

[All Certified Machine Learning Professional Questions]

A machine learning engineer is monitoring categorical input variables for a production machine learning application. The engineer believes that missing values are becoming more prevalent in more recent data for a particular value in one of the categorical input variables.

FORUM

Q

Which of the following tools can the machine learning engineer use to assess their theory?

- A. Kolmogorov-Smirnov (KS) test
- B. One-way Chi-squared Test
- C. Two-way Chi-squared Test
- D. Jenson-Shannon distance
- E. None of these

HOME EXAMTOPICS PRO POPULAR EXAMS VIEW ALL EXAMS DOWNLOAD FREE COURSES CONTACT FORUM

IN E W

Actual exam question from Databricks's Certified Machine Learning Professional

Question #: 3

Topic #: 1

[All Certified Machine Learning Professional Questions]

A data scientist is using MLflow to track their machine learning experiment. As a part of each MLflow run, they are performing hyperparameter tuning. The data scientist would like to have one parent run for the tuning process with a child run for each unique combination of hyperparameter values.

They are using the following code block:

```
with mlflow.start_run(run_name="Parent run") as run:
    print("Start parent run")
with mlflow.start_run(run_name="Child 1", nested=True):
    mlflow.log_param("run_name", "child_1")
with mlflow.start_run(run_name="Child 2", nested=True):
    mlflow.log_param("run_name", "child_2")
```

The code block is not nesting the runs in MLflow as they expected.

Which of the following changes does the data scientist need to make to the above code block so that it successfully nests the child runs under the parent run in MLflow?

- A. Indent the child run blocks within the parent run block
- B. Add the nested=True argument to the parent run
- C. Remove the nested=True argument from the child runs
- D. Provide the same name to the run\_name parameter for all three run blocks
- E. Add the nested=True argument to the parent run and remove the nested=True arguments from the child runs

IAC AA

**FORUM** 

Actual exam question from Databricks's Certified Machine Learning Professional

Question #: 4

Topic #: 1

[All Certified Machine Learning Professional Questions]

mlflow.log model and data(

A machine learning engineer wants to log feature importance data from a CSV file at path importance\_path with an MLflow run for model model. Which of the following code blocks will accomplish this task inside of an existing MLflow run block?

```
model,
A.
       importance path,
       "feature-importance.csv"
  mlflow.log model(
       model,
B.
       importance path,
       "feature-importance.csv"
C. mlflow.log_data(importance_path, "feature-importance.csv")
```

- D. mlflow.log\_artifact(importance\_path, "feature-importance.csv")
- E. None of these code blocks tan accomplish the task.

Q

Question #: 6

Topic #: 1

[All Certified Machine Learning Professional Questions]

A data scientist has developed a model to predict ice cream sales using the expected temperature and expected number of hours of sun in the day. However, the expected temperature is dropping beneath the range of the input variable on which the model was trained.

Which of the following types of drift is present in the above scenario?

- A. Label drift
- B. None of these
- C. Concept drift
- D. Prediction drift
- E. Feature drift

- A. spark.read.format("delta").load(path).drop("star\_rating")
- B. spark.read.format("delta").table(path).drop("star\_rating")
- C. Delta tables cannot be modified
- D. spark.read.table(path).drop("star\_rating")
- E. spark.sql("SELECT \* EXCEPT star\_rating FROM path")

Q

FORUM

Question #: 9

Topic #: 1

[All Certified Machine Learning Professional Questions]

A machine learning engineer is in the process of implementing a concept drift monitoring solution. They are planning to use the following steps:

- 1. Deploy a model to production and compute predicted values
- 2. Obtain the observed (actual) label values
- 3. \_\_\_\_\_
- 4. Run a statistical test to determine if there are changes over time

Which of the following should be completed as Step #3?

- A. Obtain the observed values (actual) feature values
- B. Measure the latency of the prediction time
- C. Retrain the model
- D. None of these should be completed as Step #3
- E. Compute the evaluation metric using the observed and predicted values

NEW

Actual exam question from Databricks's Certified Machine Learning Professional

Question #: 11

Topic #: 1

[All Certified Machine Learning Professional Questions]

A data scientist is utilizing MLflow to track their machine learning experiments. After completing a series of runs for the experiment with experiment ID exp\_id, the data scientist wants to programmatically work with the experiment run data in a Spark DataFrame. They have an active MLflow Client client and an active Spark session spark.

Which of the following lines of code can be used to obtain run-level results for exp\_id in a Spark DataFrame?

- A. client.list\_run\_infos(exp\_id)
- B. spark.read.format("delta").load(exp\_id)
- C. There is no way to programmatically return row-level results from an MLflow Experiment.
- D. mlflow.search\_runs(exp\_id)
- E. spark.read.format("mlflow-experiment").load(exp\_id)

**Show Suggested Answer** 

 $\sim$ 

CONTACT

Actual exam question from Databricks's Certified Machine Learning Professional

Question #: 12

Topic #: 1

[All Certified Machine Learning Professional Questions]

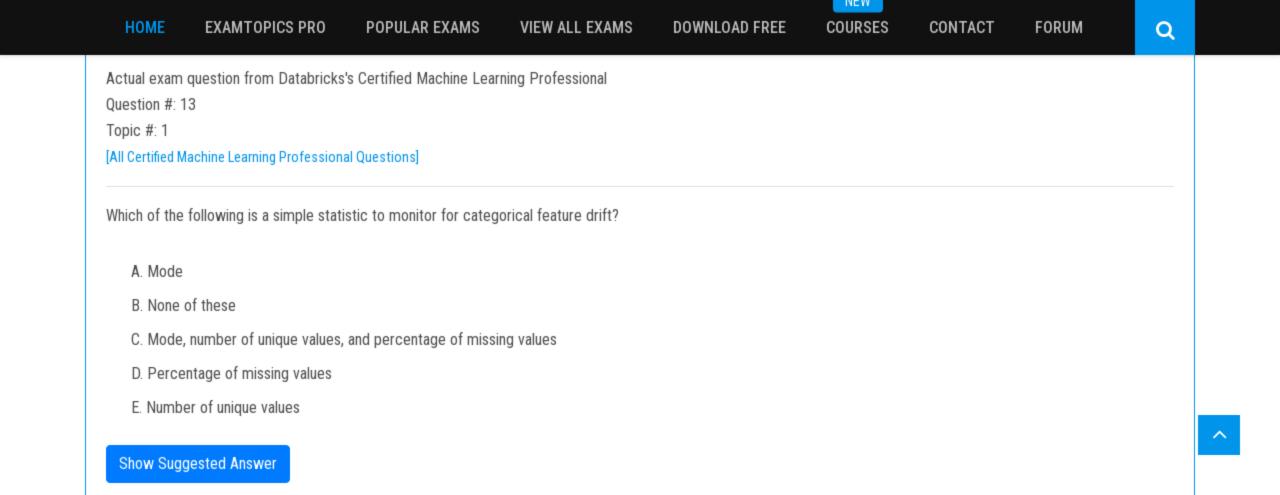
A data scientist has developed and logged a scikit-learn random forest model model, and then they ended their Spark session and terminated their cluster. After starting a new cluster, they want to review the feature\_importances\_ of the original model object.

Which of the following lines of code can be used to restore the model object so that feature\_importances\_ is available?

- A. mlflow.load\_model(model\_uri)
- B. client.list\_artifacts(run\_id)["feature-importances.csv"]
- C. mlflow.sklearn.load\_model(model\_uri)
- D. This can only be viewed in the MLflow Experiments UI
- E. client.pyfunc.load\_model(model\_uri)

**Show Suggested Answer** 

 $\sim$ 



HOME EXAMTOPICS PRO POPULAR EXAMS VIEW ALL EXAMS DOWNLOAD FREE COURSES CONTACT FORUM

Actual exam question from Databricks's Certified Machine Learning Professional

Question #: 15

Topic #: 1

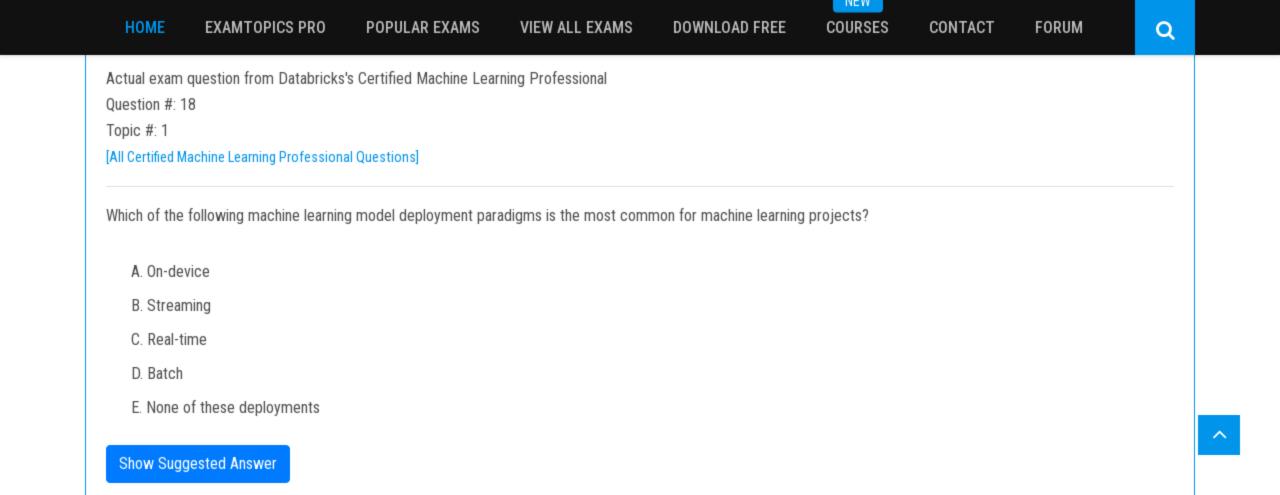
[All Certified Machine Learning Professional Questions]

A data scientist has computed updated feature values for all primary key values stored in the Feature Store table features. In addition, feature values for some new primary key values have also been computed. The updated feature values are stored in the DataFrame features\_df. They want to replace all data in features with the newly computed data.

Which of the following code blocks can they use to perform this task using the Feature Store Client fs?

```
fs.create table (
      name="features",
Α.
      df=features df,
      mode="overwrite"
  fs.write_table(
      name="features",
В.
      df=features df,
  fs.write table(
      name="features",
C.
      df=features df,
      mode="merge"
  fs.write table(
      name="features",
D.
      df=features df,
      mode="overwrite"
  fs.create table (
      name="features",
E.
      df=features df,
      mode="merge"
```

- A. VERSION
- B. DESCRIBE
- C. HISTORY
- D. DESCRIBE HISTORY
- E. TIMESTAMP



Q

Actual exam question from Databricks's Certified Machine Learning Professional

Question #: 19

Topic #: 1

[All Certified Machine Learning Professional Questions]

A data scientist would like to enable MLflow Autologging for all machine learning libraries used in a notebook. They want to ensure that MLflow Autologging is used no matter what version of the Databricks Runtime for Machine Learning is used to run the notebook and no matter what workspace-wide configurations are selected in the Admin Console.

Which of the following lines of code can they use to accomplish this task?

- A. mlflow.sklearn.autolog()
- B. mlflow.spark.autolog()
- C. spark.conf.set("autologging", True)
- D. It is not possible to automatically log MLflow runs.
- E. mlflow.autolog()

HOME EXAMTOPICS PRO

**POPULAR EXAMS** 

VIEW ALL EXAMS

DOWNLOAD FREE

COURSES

CONTACT

Q

**FORUM** 

Actual exam question from Databricks's Certified Machine Learning Professional

Question #: 20

Topic #: 1

[All Certified Machine Learning Professional Questions]

A data scientist has developed a model model and computed the RMSE of the model on the test set. They have assigned this value to the variable rmse. They now want to manually store the RMSE value with the MLflow run.

They write the following incomplete code block:

```
with mlflow.start_run(experiment_id=exp_id, run_name=run_name) as run:
    # Log rmse
    mlflow.____("rmse", rmse)
```

Which of the following lines of code can be used to fill in the blank so the code block can successfully complete the task?

- A. log\_artifact
- B. log\_model
- C. log\_metric
- D. log\_param
- E. There is no way to store values like this.

Question #: 22

Topic #: 1

[All Certified Machine Learning Professional Questions]

A data scientist has developed a scikit-learn random forest model model, but they have not yet logged model with MLflow. They want to obtain the input schema and the output schema of the model so they can document what type of data is expected as input.

FORUM

Q

Which of the following MLflow operations can be used to perform this task?

- A. mlflow.models.schema.infer\_schema
- B. mlflow.models.signature.infer\_signature
- C. mlflow.models.Model.get\_input\_schema
- D. mlflow.models.Model.signature
- E. There is no way to obtain the input schema and the output schema of an unlogged model.

Question #: 23

Topic #: 1

[All Certified Machine Learning Professional Questions]

A machine learning engineer and data scientist are working together to convert a batch deployment to an always-on streaming deployment. The machine learning engineer has expressed that rigorous data tests must be put in place as a part of their conversion to account for potential changes in data formats.

Which of the following describes why these types of data type tests and checks are particularly important for streaming deployments?

- A. Because the streaming deployment is always on, all types of data must be handled without producing an error
- B. All of these statements
- C. Because the streaming deployment is always on, there is no practitioner to debug poor model performance
- D. Because the streaming deployment is always on, there is a need to confirm that the deployment can autoscale
- E. None of these statements

**Show Suggested Answer** 

^

**FORUM** 

Question #: 25

Topic #: 1

[All Certified Machine Learning Professional Questions]

A machine learning engineering team wants to build a continuous pipeline for data preparation of a machine learning application. The team would like the data to be fully processed and made ready for inference in a series of equal-sized batches.

Which of the following tools can be used to provide this type of continuous processing?

- A. Spark UDFs
- B. Structured Streaming
- C. MLflow
- D. Delta Lake
- E. AutoML

COURSES

CONTACT

Actual exam question from Databricks's Certified Machine Learning Professional

Question #: 26

Topic #: 1

[All Certified Machine Learning Professional Questions]

A machine learning engineer wants to deploy a model for real-time serving using MLflow Model Serving. For the model, the machine learning engineer currently has one model version in each of the stages in the MLflow Model Registry. The engineer wants to know which model versions can be queried once Model Serving is enabled for the model.

Which of the following lists all of the MLflow Model Registry stages whose model versions are automatically deployed with Model Serving?

- A. Staging, Production, Archived
- B. Production
- C. None, Staging, Production, Archived
- D. Staging, Production
- E. None, Staging, Production

FORUM

Q

Actual exam question from Databricks's Certified Machine Learning Professional

Question #: 28

Topic #: 1

[All Certified Machine Learning Professional Questions]

A machine learning engineer is converting a Hyperopt-based hyperparameter tuning process from manual MLflow logging to MLflow Autologging. They are trying to determine how to manage nested Hyperopt runs with MLflow Autologging.

Which of the following approaches will create a single parent run for the process and a child run for each unique combination of hyperparameter values when using Hyperopt and MLflow Autologging?

- A. Starting a manual parent run before calling fmin
- B. Ensuring that a built-in model flavor is used for the model logging
- C. Starting a manual child run within the objective\_function
- D. There is no way to accomplish nested runs with MLflow Autologging and Hyperopt
- E. MLflow Autologging will automatically accomplish this task with Hyperopt

**Show Suggested Answer** 

 $\sim$ 

Question #: 29

Topic #: 1

[All Certified Machine Learning Professional Questions]

A data scientist has created a Python function compute\_features that returns a Spark DataFrame with the following schema:

```
customer_id STRING,
spend DOUBLE,
units INT,
loyal INT,
region STRING
```

The resulting DataFrame is assigned to the features\_df variable. The data scientist wants to create a Feature Store table using features\_df.

Which of the following code blocks can they use to create and populate the Feature Store table using the Feature Store Client fs?

```
fs.create table(
       name="new table",
       primary_keys="customer_id",
       df=features df,
       description="Customer features"
  fs.create_table(
      name="new table",
В.
      primary keys="customer id",
      description="Customer features"
C. features_df.write.mode("fs").path("new_table")
  fs.create table(
      name="new table",
      primary keys="customer id",
D.
      function=compute features,
      description="Customer features"
```

E. features\_df.write.mode("feature").path("new\_table")

CONTACT FORUM

Q

Actual exam question from Databricks's Certified Machine Learning Professional

Question #: 32

Topic #: 1

[All Certified Machine Learning Professional Questions]

A machine learning engineer has deployed a model recommender using MLflow Model Serving. They now want to query the version of that model that is in the Production stage of the MLflow Model Registry.

Which of the following model URIs can be used to guery the described model version?

- A. https://<databricks-instance>/model-serving/recommender/Production/invocations
- B. The version number of the model version in Production is necessary to complete this task.
- C. https://<databricks-instance>/model/recommender/stage-production/invocations
- D. https://<databricks-instance>/model-serving/recommender/stage-production/invocations
- E. https://<databricks-instance>/model/recommender/Production/invocations

Question #: 35

Topic #: 1

[All Certified Machine Learning Professional Questions]

A data scientist set up a machine learning pipeline to automatically log a data visualization with each run. They now want to view the visualizations in Databricks. Which of the following locations in Databricks will show these data visualizations?

- A. The MLflow Model Registry Model page
- B. The Artifacts section of the MLflow Experiment page
- C. Logged data visualizations cannot be viewed in Databricks
- D. The Artifacts section of the MLflow Run page
- E. The Figures section of the MLflow Run page

**Show Suggested Answer** 

^

FORUM

**FORUM** 

Actual exam question from Databricks's Certified Machine Learning Professional

Question #: 36

Topic #: 1

[All Certified Machine Learning Professional Questions]

A data scientist has developed a scikit-learn model sklearn\_model and they want to log the model using MLflow.

They write the following incomplete code block:

```
with mlflow.start_run(experiment_id=exp_id, run_name=run_name) as run:
    # Log model
```

Which of the following lines of code can be used to fill in the blank so the code block can successfully complete the task?

- A. mlflow.spark.track\_model(sklearn\_model, "model")
- B. mlflow.sklearn.log\_model(sklearn\_model, "model")
- C. mlflow.spark.log\_model(sklearn\_model, "model")
- D. mlflow.sklearn.load\_model("model")
- E. mlflow.sklearn.track\_model(sklearn\_model, "model")

- A. The launch of a new cost-efficient SQL endpoint
- B. CI/CD pipelines are not needed for machine learning pipelines
- C. The arrival of a new feature table in the Feature Store
- D. The launch of a new cost-efficient job cluster
- E. The arrival of a new model version in the MLflow Model Registry

Question #: 39

Topic #: 1

[All Certified Machine Learning Professional Questions]

A machine learning engineering team has written predictions computed in a batch job to a Delta table for querying. However, the team has noticed that the querying is running slowly. The team has already tuned the size of the data files. Upon investigating, the team has concluded that the rows meeting the query condition are sparsely located throughout each of the data files.

Based on the scenario, which of the following optimization techniques could speed up the query by colocating similar records while considering values in multiple columns?

- A. Z-Ordering
- B. Bin-packing
- C. Write as a Parquet file
- D. Data skipping
- E. Tuning the file size

**Show Suggested Answer** 

^

Question #: 40

Topic #: 1

[All Certified Machine Learning Professional Questions]

A machine learning engineer needs to deliver predictions of a machine learning model in real-time. However, the feature values needed for computing the predictions are available one week before the query time.

Which of the following is a benefit of using a batch serving deployment in this scenario rather than a real-time serving deployment where predictions are computed at query time?

- A. Batch serving has built-in capabilities in Databricks Machine Learning
- B. There is no advantage to using batch serving deployments over real-time serving deployments
- C. Computing predictions in real-time provides more up-to-date results
- D. Testing is not possible in real-time serving deployments
- E. Querying stored predictions can be faster than computing predictions in real-time

**Show Suggested Answer** 

 $\sim$ 

FORUM