Vendor: Databricks

Exam Code:DATABRICKS-MACHINE-LEARNING-PROFESSIONAL

Exam Name: Databricks Certified Machine Learning Professional

Version: Demo

QUESTION 1

A data scientist has written a function to track the runs of their random forest model. The data scientist is changing the number of trees in the forest across each run. Which of the following MLflow operations is designed to log single values like the number of trees in a random forest?

A. mlflow.log_artifact

B. mlflow.log_model

C. mlflow.log_metric

D. mlflow.log_param

E. There is no way to store values like this.

Correct Answer: D

QUESTION 2

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

Correct Answer: A

QUESTION 3

A machine learning engineer wants to move their model version model_version for the MLflow Model Registry model model from the Staging stage to the Production stage using MLflow Client client. Which of the following code blocks can they use to accomplish the task?

```
A.      client.transition_model_version_stage(
            name=model,
            version=model_version,
            stage="Staging"
        )

B.      client.transition_model_stage(
            name=model,
            version=model_version,
            stage="Production"
        )

C.      client.transition_model_version_stage(
            name=model,
            version=model_version,
            stage="Production"
        )

        client.transition_model__stage(
            name=model,
            version=model_version,
D.          from="Staging",
            to="Production"
        )

        client.transition_model_version_stage(
            name=model,
            version=model_version,
E.          from="Staging",
            to="Production"
        )
```

A. Option A

B. Option B

C. Option C

D. Option D

E. Option E

Correct Answer: C

QUESTION 4

A machine learning engineering manager has asked all of the engineers on their team to add text descriptions to each of the model projects in the MLflow Model Registry. They are starting with the model project "model" and they\\'d like to add

the text in the model_description variable.

The team is using the following line of code:

```
client = MlflowClient()
client.update_registered_model(
    name="model",
    description=model_description
)
```

Which of the following changes does the team need to make to the above code block to accomplish the task?

A. Replace update_registered_model with update_model_version

B. There no changes necessary

C. Replace description with artifact

D. Replace client.update_registered_model with mlflow

E. Add a Python model as an argument to update_registered_model

Correct Answer: B

QUESTION 5

Which of the following describes concept drift?

A. Concept drift is when there is a change in the distribution of an input variable

B. Concept drift is when there is a change in the distribution of a target variable

C. Concept drift is when there is a change in the relationship between input variables and target variables

D. Concept drift is when there is a change in the distribution of the predicted target given by the model

E. None of these describe Concept drift

Correct Answer: C

QUESTION 6

Which of the following is a simple statistic to monitor for categorical feature drift?

A. Mode

B. None of these

C. Mode, number of unique values, and percentage of missing values

D. Percentage of missing values

E. Number of unique values

Correct Answer: C

QUESTION 7

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

Correct Answer: B

QUESTION 8

Which of the following is a reason for using Jensen-Shannon (JS) distance over a Kolmogorov-Smirnov (KS) test for numeric feature drift detection?

A. All of these reasons

B. JS is not normalized or smoothed

C. None of these reasons

D. JS is more robust when working with large datasets

E. JS does not require any manual threshold or cutoff determinations

Correct Answer: E

QUESTION 9

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:

image14

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")

Correct Answer: B

QUESTION 10

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",
A.       df=features_df,
         mode="overwrite"
     )

     fs.write_table(
         name="features",
B.       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. Option A

B. Option B

C. Option C

D. Option D

E. Option E

Correct Answer: D

QUESTION 11

A machine learning engineer wants to programmatically create a new Databricks Job whose schedule depends on the result of some automated tests in a machine learning pipeline. Which of the following Databricks tools can be used to programmatically create the Job?

A. MLflow APIs

B. AutoML APIs

C. MLflow Client

D. Jobs cannot be created programmatically

E. Databricks REST APIs

Correct Answer: E

QUESTION 12

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

Correct Answer: A