

```
mlflow.log_model(  
    model,  
    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 can accomplish the task.

Answer: A

Question: 5

Which of the following is a simple, low-cost method of monitoring numeric feature drift?

- A. Jensen-Shannon test
- B. Summary statistics trends
- C. Chi-squared test
- D. None of these can be used to monitor feature drift
- E. Kolmogorov-Smirnov (KS) test

Answer: B

Question: 6

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

Answer: E

Question: 7

A data scientist wants to remove the star_rating column from the Delta table at the location path. To do this, they need to load in data and drop the star_rating column.

Which of the following code blocks accomplishes this task?

- 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")`

Answer: D

Question: 8

Which of the following operations in Feature Store Client `fs` can be used to return a Spark DataFrame of a data set associated with a Feature Store table?

- A. `fs.create_table`
- B. `fs.write_table`
- C. `fs.get_table`
- D. There is no way to accomplish this task with `fs`
- E. `fs.read_table`

Answer: A

Question: 9

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

Answer: D

Question: 10

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

Answer: D

Question: 11

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)`

Answer: B

Question: 12

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)`

Answer: A

Question: 13

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

Answer: C

Question: 14

Which of the following is a probable response to identifying drift in a machine learning application?

- A. None of these responses
- B. Retraining and deploying a model on more recent data
- C. All of these responses
- D. Rebuilding the machine learning application with a new label variable
- E. Sunsetting the machine learning application

Answer: A

Question: 15

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?

A)

```
fs.create_table(  
    name="features",  
    df=features_df,  
    mode="overwrite"  
)
```

B)

```
fs.write_table(  
    name="features",  
    df=features_df,  
)
```

C)

```
fs.write_table(  
  name="features",  
  df=features_df,  
  mode="merge"  
)
```

D)

```
fs.write_table(  
  name="features",  
  df=features_df,  
  mode="overwrite"  
)
```

E)

```
fs.create_table(  
  name="features",  
  df=features_df,  
  mode="merge"  
)
```

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

Answer: E

Question: 16

After a data scientist noticed that a column was missing from a production feature set stored as a Delta table, the machine learning engineering team has been tasked with determining when the column was dropped from the feature set.

Which of the following SQL commands can be used to accomplish this task?

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

Answer: D

Question: 17

Which of the following describes label drift?

- A. Label drift is when there is a change in the distribution of the predicted target given by the model
- B. None of these describe label drift
- C. Label drift is when there is a change in the distribution of an input variable
- D. Label drift is when there is a change in the relationship between input variables and target variables
- E. Label drift is when there is a change in the distribution of a target variable

Answer: C

Question: 18

Which of the following machine learning model deployment paradigms is the most common for machine learning projects?

- A. On-device
- B. Streaming
- C. Real-time
- D. Batch
- E. None of these deployments

Answer: B

Question: 19

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()`

Answer: C

Question: 20

A data scientist has developed a 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_n
    # 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.

Answer: A

Question: 21

Which of the following MLflow operations can be used to automatically calculate and log a Shapley feature importance plot?

- A. `mlflow.shap.log_explanation`
- B. None of these operations can accomplish the task.
- C. `mlflow.shap`
- D. `mlflow.log_figure`
- E. `client.log_artifact`

Answer: C

Question: 22

A data scientist has developed a scikit-learn random forest 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.

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

- A. `mlflow.models.schema.infer_schema`