Free Questions for DSA-C02 by dumpssheet

Shared by Donaldson on 04-10-2023

For More Free Questions and Preparation Resources

Check the Links on Last Page

Question 1

Question Type: MultipleChoice

Mark the incorrect statement regarding usage of Snowflake Stream & Tasks?

Options:

A- Snowflake automatically resizes and scales the compute resources for serverless tasks.

B- Snowflake ensures only one instance of a task with a schedule (i.e. a standalone task or the root task in a DAG) is executed at a given time. If a task is still running when the next scheduled execution time occurs, then that scheduled time is skipped.

C- Streams support repeatable read isolation.

D- An standard-only stream tracks row inserts only.

Answer:

D

Explanation:

All are correct except a standard-only stream tracks row inserts only.

A standard (i.e. delta) stream tracks all DML changes to the source object, including inserts, up-dates, and deletes (including table truncates).

Question 2

Question Type: **MultipleChoice**

Which tools helps data scientist to manage ML lifecycle & Model versioning?

Options:

A- MLFlow

B- Pachyderm

C- Albert

D- CRUX

Answer:

A, B

Explanation:

Model versioning in a way involves tracking the changes made to an ML model that has been previously built. Put differently, it is the process of making changes to the configurations of an ML Model. From another perspective, we can see model versioning as a feature that helps Machine Learning Engineers, Data Scientists, and related personnel create and keep multiple versions of the same model.

Think of it as a way of taking notes of the changes you make to the model through tweaking hyperparameters, retraining the model with more data, and so on.

In model versioning, a number of things need to be versioned, to help us keep track of important changes. I'll list and explain them below:

Implementation code: From the early days of model building to optimization stages, code or in this case source code of the model plays an important role. This code experiences significant changes during optimization stages which can easily be lost if not tracked properly. Because of this, code is one of the things that are taken into consideration during the model versioning process.

Data: In some cases, training data does improve significantly from its initial state during model op-timization phases. This can be as a result of engineering new features from existing ones to train our model on. Also there is metadata (data about your training data and model) to consider versioning. Metadata can change different times over without the training data actually changing. We need to be able to track these changes through versioning

Model: The model is a product of the two previous entities and as stated in their explanations, an ML model changes at different points of the optimization phases through hyperparameter setting, model artifacts and learning coefficients. Versioning helps take record of the different versions of a Machine Learning model.

MLFlow & Pachyderm are the tools used to manage ML lifecycle & Model versioning.

Question 3

Question Type: MultipleChoice

You are training a binary classification model to support admission approval decisions for a college degree program.

How can you evaluate if the model is fair, and doesn't discriminate based on ethnicity?

Options:

A- Evaluate each trained model with a validation dataset and use the model with the highest accuracy score.

B- Remove the ethnicity feature from the training dataset.

C- Compare disparity between selection rates and performance metrics across ethnicities.

D- None of the above.

Answer:

C

Explanation:

By using ethnicity as a sensitive field, and comparing disparity between selection rates and performance metrics for each ethnicity value, you can evaluate the fairness of the model.

Question 4

Question Type: MultipleChoice

You previously trained a model using a training dataset. You want to detect any data drift in the new data collected since the model was trained.

What should you do?

Options:

A- Create a new dataset using the new data and a timestamp column and create a data drift monitor that uses the training dataset as a baseline and the new dataset as a target.

B- Create a new version of the dataset using only the new data and retrain the model.

C- Add the new data to the existing dataset and enable Application Insights for the service where the model is deployed.

D- Retrained your training dataset after correcting data outliers & no need to introduce new data.

Answer:

A

Explanation:

To track changing data trends, create a data drift monitor that uses the training data as a baseline and the new data as a target.

Model drift and decay are concepts that describe the process during which the performance of a model deployed to production degrades on new, unseen data or the underlying assumptions about the data change.

These are important metrics to track once models are deployed to production. Models must be regularly re-trained on new data. This is referred to as refitting the model. This can be done either on a periodic basis, or, in an ideal scenario, retraining can be triggered when the performance of the model degrades below a certain pre-defined threshold.

Question 5

Question Type: **MultipleChoice**

Which metric is not used for evaluating classification models?

Options:

A- Recall

B- Accuracy

C- Mean absolute error

D- Precision

Answer:

C

Explanation:

The four commonly used metrics for evaluating classifier performance are:

1. Accuracy: The proportion of correct predictions out of the total predictions.

2. Precision: The proportion of true positive predictions out of the total positive predictions (precision = true positives / (true positives + false positives)).

3. Recall (Sensitivity or True Positive Rate): The proportion of true positive predictions out of the total actual positive instances (recall = true positives / (true positives + false negatives)).

4. F1 Score: The harmonic mean of precision and recall, providing a balance between the two metrics (F1 score = 2 * ((precision * recall) / (precision + recall))).

Root Mean Squared Error (RMSE)and Mean Absolute Error (MAE) are metrics used to evaluate a Regression Model. These metrics tell us how accurate our predictions are and, what is the amount of deviation from the actual values.

Question 6

Question Type: **MultipleChoice**

How do you handle missing or corrupted data in a dataset?

Options:

A- Drop missing rows or columns

B- Replace missing values with mean/median/mode

C- Assign a unique category to missing values

D- All of the above

Answer:

D

To Get Premium Files for DSA-C02 Visit

https://www.p2pexams.com/products/dsa-c02

For More Free Questions Visit

https://www.p2pexams.com/snowflake/pdf/dsa-c02

