

Practical Data Science Specialization



Week 3 quiz

Practice Quiz • 20 min

1. Machine learning pipelines are created as a way to iterate over the ML workflow quickly with less handoffs and reduced manual efforts. This is done by automating the tasks and orchestrating the sequence of task execution. What is the on-going set of tasks that are performed after a model is deployed with a pipeline ?

1 / 1 point

- ☐ Data Tasks
- ☐ Model Building Tasks
- ☐ Model Deployment Tasks
- ☒ Model Monitoring Tasks

✓ **Correct**

Correct! The model monitoring tasks are the on-going set of tasks performed after the model is deployed by the pipeline.

2. One keypoint in building an effective pipeline is establishing traceability of artifacts. Which of the following is/are reasons why traceability is important. (choose all that apply)

1 / 1 point

- ☐ Traceability slows down model deployment.
- ☐ Traceability allows you to manage end-to-end automation of the pipeline.
- ☒ Traceability ensures easy recovery of any resource or artifact.

✓ **Correct**

Correct! Every step in the pipeline is recorded and detailed logging information is available for traceability. Hence, recovery of resources and artifacts are easy.

- ☒ Traceability provides better visibility and support.

✓ **Correct**

Correct! With traceability you have a managed view into how the pipeline artifacts were created.

3. Data lakes are enterprise storage solutions which can host virtually any amount of data files. What are some reason(s) for the establishment of a governed data lake recommended in the ML pipeline? (Select all that apply)

1 / 1 point

- ☒ Ensures traceability for the data

✓ **Correct**

Correct! There is improved traceability when using a data lake as we can directly track the data used by practitioners for their data science projects.

- ☐ Decreases model training time by 50%
- ☒ Decreases the wait time to begin model development

✓ **Correct**

Correct! Data Scientists don't have to wait for dataset collection, long approvals, security clearance etc. They can just go straight to feature engineering and model building.

- ☒ Improves retraining workflows

✓ **Correct**

Correct! Retraining workflows are easier to support when using data lakes.

Practical Data Science Specialization



Week 3 quiz

Practice Quiz • 20 min

4. True or False: In the ML pipeline, all inputs and artifacts that are produced as part of an automated task have versions associated with them.

1 / 1 point

- ☒ True
☐ False

✓ **Correct**

Correct! Versioning inputs and outputs of each step in the workflow ensures traceability. Some types of versioning include data, code, and runtime versioning.

5. In the data tasks, several validation tasks can be incorporated into the machine learning pipeline and automated to improve quality. What are some of these validation tasks? (Choose all that apply)

1 / 1 point

- ☒ Data quality checks

✓ **Correct**

Correct! Data quality issues can be found here by implementing automated steps that gather statistics about your input data and ensuring the data complies with specific data-quality guidelines.

- ☒ Detection of statistical data bias

✓ **Correct**

Correct! Automated checks to detect statistical data bias can be included as a step within the pipeline.

- ☐ Hyper-parameter Tuning

- ☒ Data schema

✓ **Correct**

Correct! This automated check ensures that the data structure and schema that's coming from the input is what's expected.

6. The Model Registry is a central store to manage model metadata and model artifacts. What other information can be retrieved from the model registry? (choose all that apply)

1 / 1 point

- ☐ Includes a collection of built-in CI/CD templates for popular pipelines

- ☒ Which model versions may or may not be deployed in production.

✓ **Correct**

Correct! It provides information about the model version that was deployed - as well as the environment the model was deployed into.

- ☐ Steps to build, train, and deploy a models in Amazon SageMaker Pipelines

- ☒ Provides traceability and auditability

✓ **Correct**

Correct! In the model registry you can easily view information about model lineage and traceability.

Practical Data Science Specialization

7. Someone inadvertently deletes a live endpoint without knowing exactly how it was built. What functionality in the ML pipeline makes it possible to retrieve the deployed model?

1 / 1 point

- ☐ ML workflow orchestration
- ☒ Artifact and lineage tracking
- ☐ Artifacts
- ☐ Feature store

✓ **Correct**

Correct! Artifact and lineage tracking provide the operational capabilities needed to reproduce a deleted model.