

# AMLFv1EN-6778

## Module 3 Knowledge Check

Which resources help define a machine learning (ML) problem? (Select TWO.)

- ☒ Access to labeled data
- ☒ A domain expert
- ☐ A traditional model
- ☐ Sufficient data
- ☐ A neural network

Correct

Access to labeled data is often key to training a successful ML model. A domain expert can help you understand your data and answer questions about it.

Continue



Number correct: 1 out of 1

When preparing data for supervised classification machine learning, which attributes should the data have? (Select TWO.)

- ☒ Data should be labeled
- ☐ Data should be generated
- ☐ Anyone can use the data
- ☐ Data should be secured
- ☒ Data should be representative of production data

Correct

For supervised machine learning, a target variable is required. Data that contains a target is said to be labeled. Data must be representative of production data, and should not be generated. Data should be secured to only those with a business need.

Continue



Number correct: 2 out of 2

What can you learn by examining the statistics of your data?

- ☒ Identifying anomalies
- ☐ Verifying data quality
- ☐ Removing outliers
- ☐ Filling missing values

Correct

Statistics give you insights into data that help you find anomalies in your data.

Continue



Number correct: 3 out of 3

You have a preprocessed dataset that's ready for use in training a model. How should you divide your training data?

- ☐ Use all the data to train the model
- ☐ Split the data into two equal halves for training and testing
- ☒ Split the data into 70-80% for training, 10% for validation, and 10% for testing
- ☐ Split the data into 70-80% for training and 20% for validation

Correct

You should use 70-80% of your data for training, and split the rest between testing and validation.

Continue



Number correct: 4 out of 4

True or False: You can select between single model and multi-model hosting with Amazon SageMaker.

- ☒ True
- ☐ False

Correct

You can use Amazon SageMaker to host your single model or with multiple models.

Continue



Number correct: 5 out of 5

What is the purpose of a confusion matrix?

- ☐ To plot
- ☒ To show true or false positives
- ☐ To show
- ☐ To strat

Correct

The confusion matrix shows the true or false positives, along with the true or false negatives.

Continue



Number correct: 6 out of 6

What does a correlation heatmap show?

- ☒ The level of correlation between features in a dataset.
- ☐ The level of correlation between features in a dataset.
- ☐ The level of correlation between features in a dataset.
- ☐ The level of correlation between features in a dataset.

Correct

A correlation heatmap shows any positive or negative correlation between features in a dataset.

Continue



Number correct: 7 out of 7

Which of the following file formats does pandas support for data importing? (Select TWO.)

- ☒ JSON
- ☐ MS Word
- ☒ CSV
- ☐ Binary
- ☐ PDF

Correct

pandas can import data stored in a JSON document or a CSV file.

Continue



Number correct: 8 out of 8

Which Amazon service can you use to deploy machine learning instances and run Jupyter Notebooks?

- ☐ Amazon SageMaker
- ☒ Amazon SageMaker
- ☐ Amazon SageMaker
- ☐ Amazon SageMaker

Correct

Amazon SageMaker is a service that you can use to deploy machine learning solutions, host Jupyter Notebooks, and also host JupyterLab.

Continue



Number correct: 9 out of 9

What is the goal of an Amazon SageMaker hyperparameter tuning job?

- ☐ To optimize the model's performance
- ☒ To optimize the model's performance
- ☐ To optimize the model's performance
- ☐ To optimize the model's performance

Correct

A hyperparameter tuning job tunes the model parameters to find the best model.

Continue



Number correct: 10 out of 10