



Quiz: Dataset Principles



Why is it important to split your model into a training and testing dataset?

In order to train your model and test it on never before seen data. This will prevent overfitting and result in a more realistic evaluation metric.

- ☐ Because training takes time and resources, so splitting your data will decrease the time required for training.
- ☐ You don't need to split your data, it's best to use as much data as you can for your training.

SUBMIT

QUESTION 2 OF 2

Select all of the key takeaways for starting a project with a new dataset below.

Split your data into a train / validation / test set.

- ☐ Data is not that important for model training, so you can usually skip this step.

Use a `random_state` so you can reproduce the data splits if needed.

The split should usually contain 50% or more for the training dataset, while validation and testing can split the other half for 25%/25%.

SUBMIT