

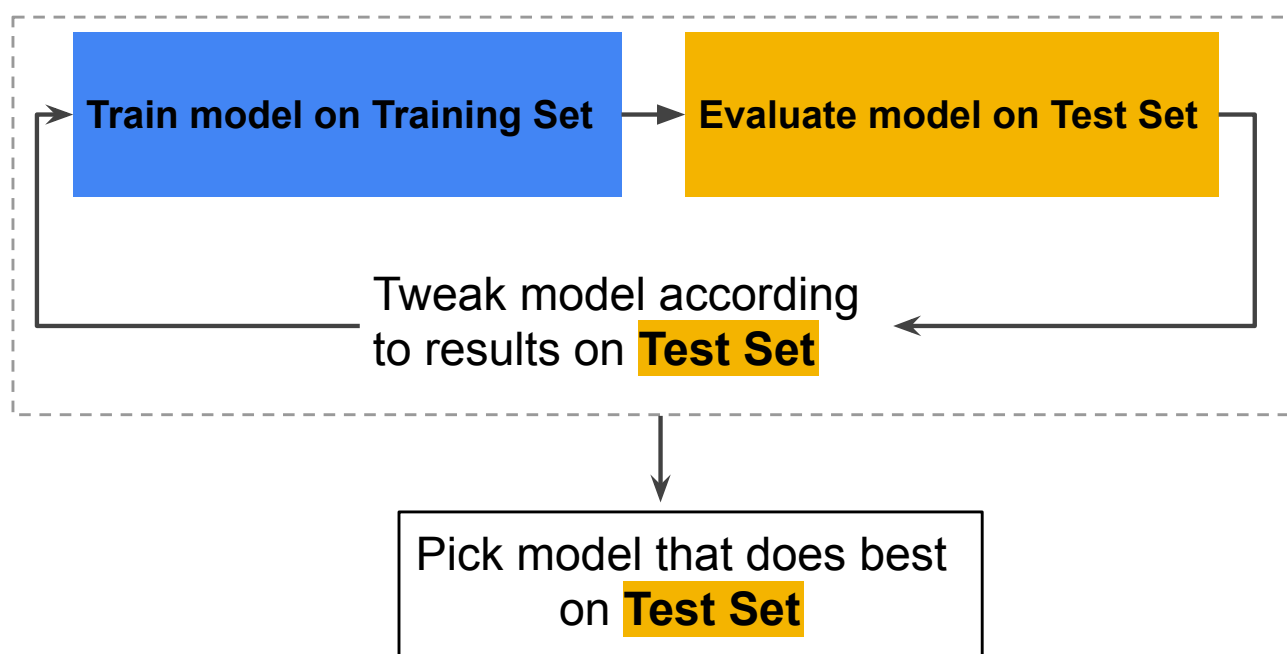
# Validation Set: Another Partition

**Estimated Time:** 8 minutes

The previous module

(<https://developers.google.com/machine-learning/crash-course/training-and-test-sets/video-lecture?authuser=0>)

introduced partitioning a data set into a training set and a test set. This partitioning enabled you to train on one set of examples and then to test the model against a different set of examples. With two partitions, the workflow could look as follows:



**Figure 1. A possible workflow?**

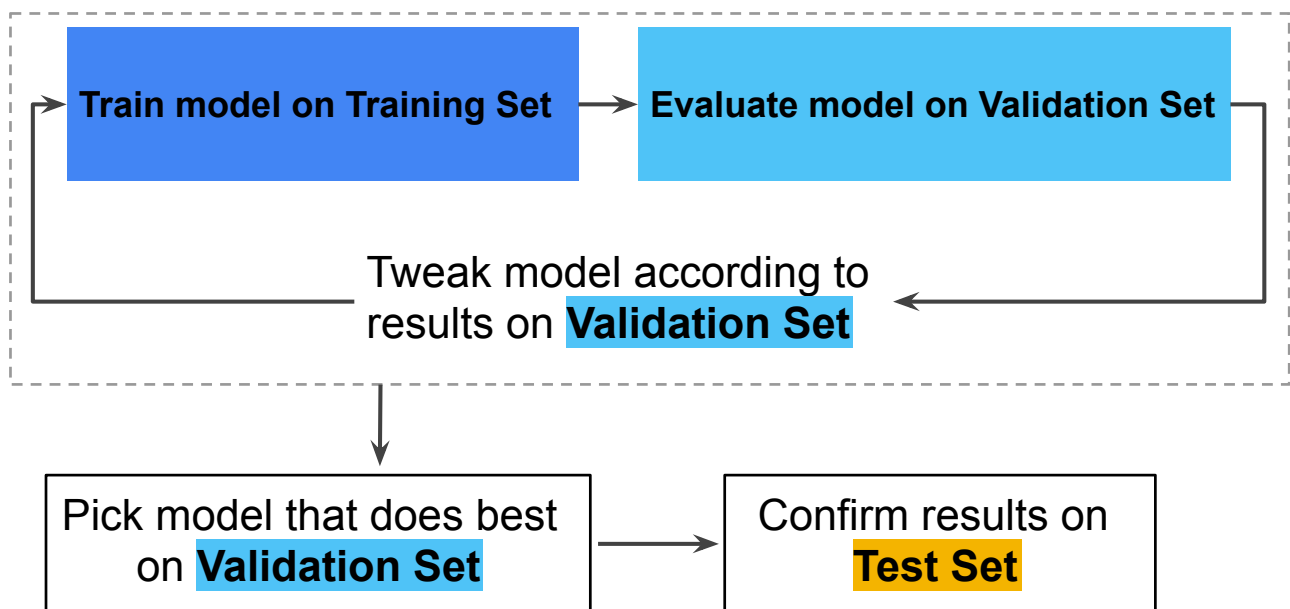
In the figure, "Tweak model" means adjusting anything about the model you can dream up—from changing the learning rate, to adding or removing features, to designing a completely new model from scratch. At the end of this workflow, you pick the model that does best on the *test set*.

Dividing the data set into two sets is a good idea, but not a panacea. You can greatly reduce your chances of overfitting by partitioning the data set into the three subsets shown in the following figure:



**Figure 2. Slicing a single data set into three subsets.**

Use the **validation set** to evaluate results from the training set. Then, use the test set to double-check your evaluation *after* the model has "passed" the validation set. The following figure shows this new workflow:



**Figure 3. A better workflow.**

In this improved workflow:

1. Pick the model that does best on the validation set.
2. Double-check that model against the test set.

This is a better workflow because it creates fewer exposures to the test set.

ets and validation sets "wear out" with repeated use. That is, the more you use the same data to make decisions, the more you overfit to that data. If you repeatedly adjust hyperparameter settings or other model improvements, the less confidence you'll have that these results generalize to new, unseen data.

ossible, it's a good idea to collect more data to "refresh" the test set and validation set. Starting anew is a gre