

[Get started](#)[Open in app](#)**Eijaz Allibhai**[Follow](#)

1.4K Followers

[About](#)

# Hold-out vs. Cross-validation in Machine Learning



Eijaz Allibhai · Oct 3, 2018 · 2 min read

I recently wrote about hold-out and cross-validation in my post about [building a k-Nearest Neighbors \(k-NN\) model to predict diabetes](#). Last week in my Machine Learning module, many students had questions about hold-out and cross-validation methods for testing, so I thought it deserved its own post.

## Hold-out

Hold-out is when you split up your dataset into a 'train' and 'test' set. The training set is what the model is trained on, and the test set is used to see how well that model performs on unseen data. A common split when using the hold-out method is using 80% of data for training and the remaining 20% of the data for testing.

## Cross-validation

Cross-validation or 'k-fold cross-validation' is when the dataset is randomly split up into 'k' groups. One of the groups is used as the test set and the rest are used as the training set. The model is trained on the training set and scored on the test set. Then the process is repeated until each unique group has been used as the test set.

For example, for 5-fold cross validation, the dataset would be split into 5 groups, and the model would be trained and tested 5 separate times so each group would get a chance to be the test set. This can be seen in the graph below.