

## **Model Validation:**

### **What is Model Validation ?**

Model validation is the set of processes and activities intended to verify that models are performing as expected.

### **Why Model Validation Matters?**

The goal of a model is to make predictions about data. Model validation determines whether the trained model is trustworthy. Also Model validation benefits in reducing the costs, discovering more errors, scalability and flexibility and enhancing the model quality.

### **The techniques of Model Validation**

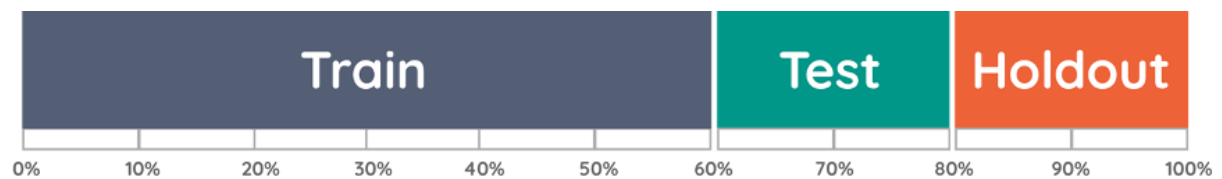
There are many techniques of Model validation:

- Train/test split
- k-Fold Cross-Validation
- Leave-one-out Cross-Validation
- Leave-one-group-out Cross-Validation
- Nested Cross-Validation
- Time-series Cross-Validation
- Wilcoxon signed-rank test
- McNemar's test
- 5x2CV paired t-test
- 5x2CV combined F test

**Here are three techniques we use more often:**

## 1. Train/Test Split

The most basic technique of Model Validation is to perform a train/validate/test split on the data. A typical ratio for this might be 80/10/10 to make sure we still have enough training data. After training the model with the training set, we will move onto validating the results and tuning the hyperparameters with the validation set till we reach a satisfactory performance metric. Once this stage is completed, we would move on to testing the model with the test set to predict and evaluate the performance.



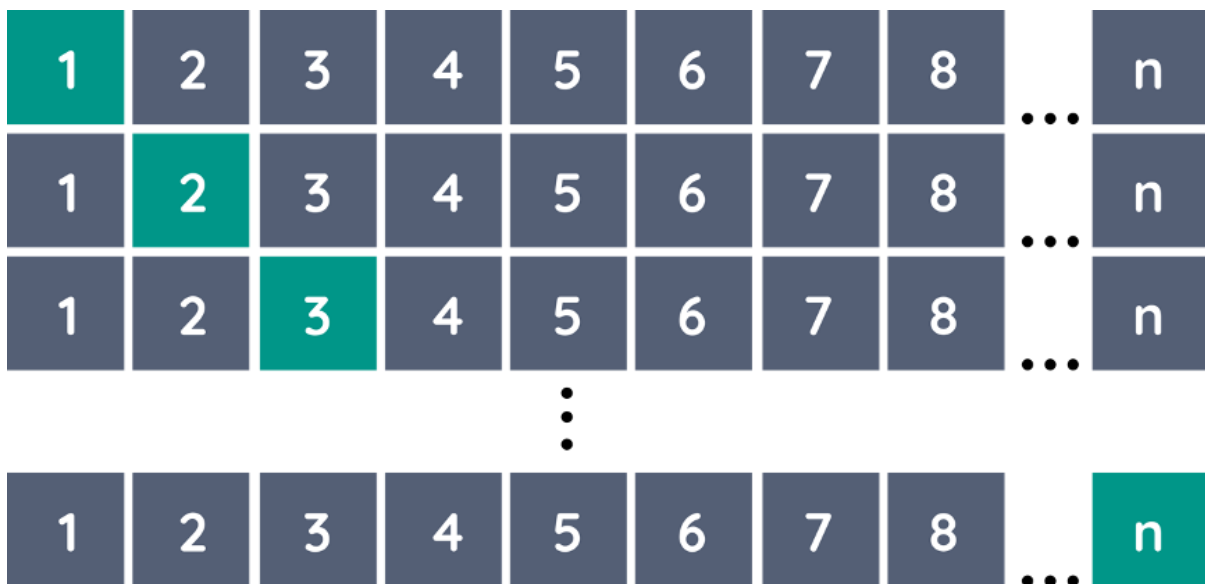
## 2. K-fold cross-validation with independent test data set.

In the situation that we would like to preserve as much data as possible for the training stage and not risk losing valuable data to the validation set. This technique will not require the training data to give up any portion for a validation set. In this instance, the dataset is broken into k number of folds wherein one fold will be used as the test set and the rest will be used as the training dataset and this will be repeated n number of times as specified by the user. In a regression the average of the results will be used as the final result. In a classification setting, the average of the results (i.e. Accuracy, True Positive Rate, F1, etc.) will be taken as the final result.



### 3. Leave-one-out cross-validation with independent test data set.

Leave-One-Out Validation is similar to the k-fold cross validation. The iteration is carried out and specified times and the dataset will be split into n-1 data sets and the one that was removed will be the test data. performance is measured the same way as k-fold cross validation.



This technique only use in small data validation.