Building A Machine Learning Validation Framework.

Before training a machine learning (ML) model, the dataset is divided into **training, validation, and testing sets**. This process of dividing the dataset is also known as ***building a validation framework***.

In an ideal situation, the dataset is normally divided into these sets and hence one can continue with the other data preprocessing steps and then proceed to training the mode.

In a situation where the dataset has not been divided into these sets, the best method to use a means of building a validation framework is **Cross Validation**. Cross Validation is the process of splitting the dataset into **Training and Testing** set in an unbiased format.

### **The Steps For Using Cross Validation In Building A Validation Framework is as follows:**

Step 1: Randomly divide the data into different groups **(n)**.

Step 2: Use **(n-1**) group for training the model and the last group for testing.

Step 3: The process is repeated for **n** number of iterations which is called **n-Fold Cross validation.**

#### Notes

*The number of iterations are called Folds.*

*The method of Cross Validation that is used in a situation where we have lots of data is the* ***10-Fold cross validation*** *which is the preferred algorithm method.*

*Another method for Cross Validation is the* ***Leave-One-Out Cross Validation****. Which is used in a situation where we are dealing with less amount of data.*

*Lastly, one of the important reasons for building a validation framework is to avoid* ***Data Leakage****. Which occurs when the same data used for training a model is also used for testing the model.*

**Moro A. Wahab**