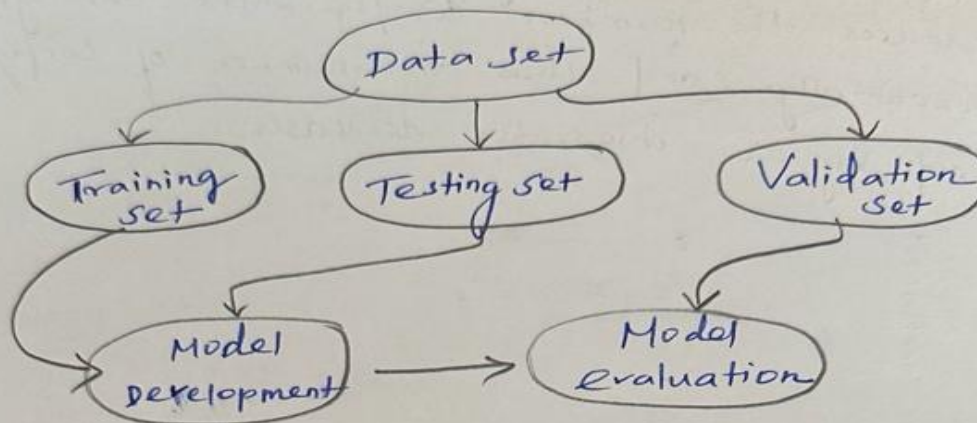




Testing of Machine Learning Algorithm



It's necessary if you care about the quality of the model. ML testing has a couple of peculiarities. It demands that you test the quality of the data, not just the model, and go through a couple of iteration adjusting the hyperparameters to get the best results!

Validation set

Having only a training set and a testing set is not enough if you do many rounds of hyperparameter tuning and can result in overfitting. To avoid that you can select a small validation data set to evaluate a model.



SNS COLLEGE OF TECHNOLOGY

AN AUTONOMOUS INSTITUTION

COIMBATORE 35

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING



Test Set

In order to assure that you select samples for testing set from training set - Examples that the machine hasn't seen before.

It's important to remain unbiased during selection and draw samples at random.

Your test set should be large enough to provide statistically meaningful results and be representative of the data set as a whole.

Cross Validation

Cross Validation is a model evaluation technique that can be performed even on a limited dataset. The training set is divided into small subsets, and the model is trained and validated on each of these samples.

Data Set

It's a set of rows and columns in an organized way

Table	Column 1	Column 2	Column 3
Row 1	Value 1	Value 2	Value 3
Row 2	Value 1	Value 2	Value 3
Row 3	Value 1	Value 2	Value 3



SNS COLLEGE OF TECHNOLOGY

AN AUTONOMOUS INSTITUTION

COIMBATORE 35

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING



- Now a days the dataset is highly unstructured (images, videos, text, etc..)
- We must transform them into an organized way that ML can understand