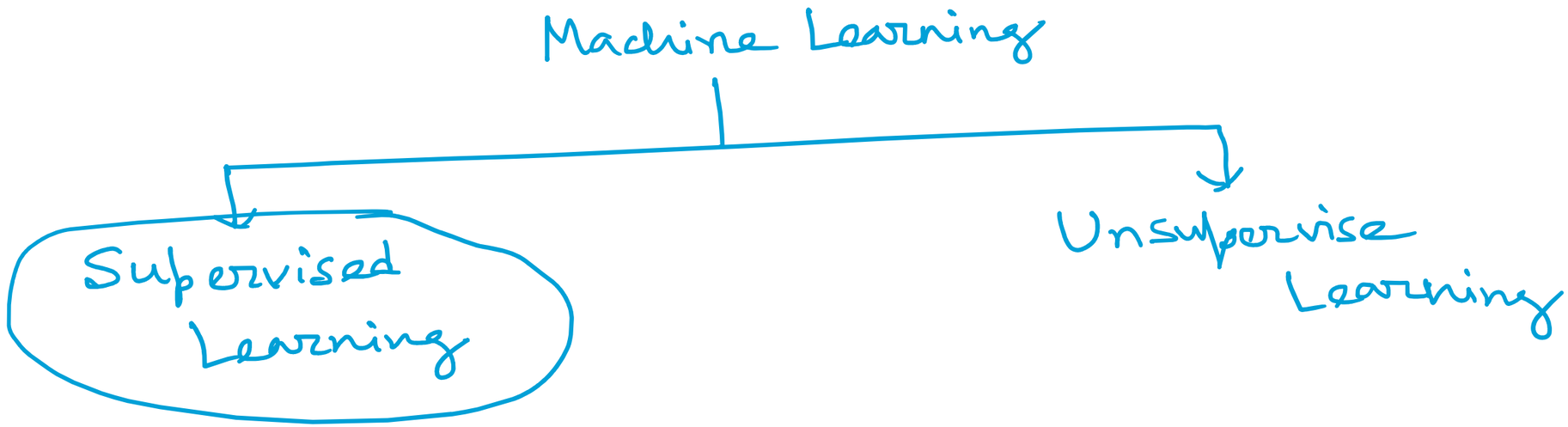
Recap

What topics we have learnt so far in machine learning



So far we have discussed ML Algorithms which are supervised.

→ $\vec{x}$ y

feature target

- Continuous (Regression).
- Categorical (classification)

Regression

	f_1	f_2	f_3	....	f_k	y
1	$x_1^{(1)}$	$x_2^{(1)}$	$x_3^{(1)}$	....	$x_k^{(1)}$	$y^{(1)}$
2	$x_1^{(2)}$	$x_2^{(2)}$	$x_3^{(2)}$	....	$x_k^{(2)}$	$y^{(2)}$
3	$x_1^{(3)}$	$x_2^{(3)}$	$x_3^{(3)}$	....	$x_k^{(3)}$	$y^{(3)}$
⋮	⋮					⋮
m	$x_1^{(m)}$	$x_2^{(m)}$	$x_3^{(m)}$	....	$x_k^{(m)}$	$y^{(m)}$

$y \rightarrow$ target

'y is continuous in nature

Ex:-

(1) Housing price prediction.

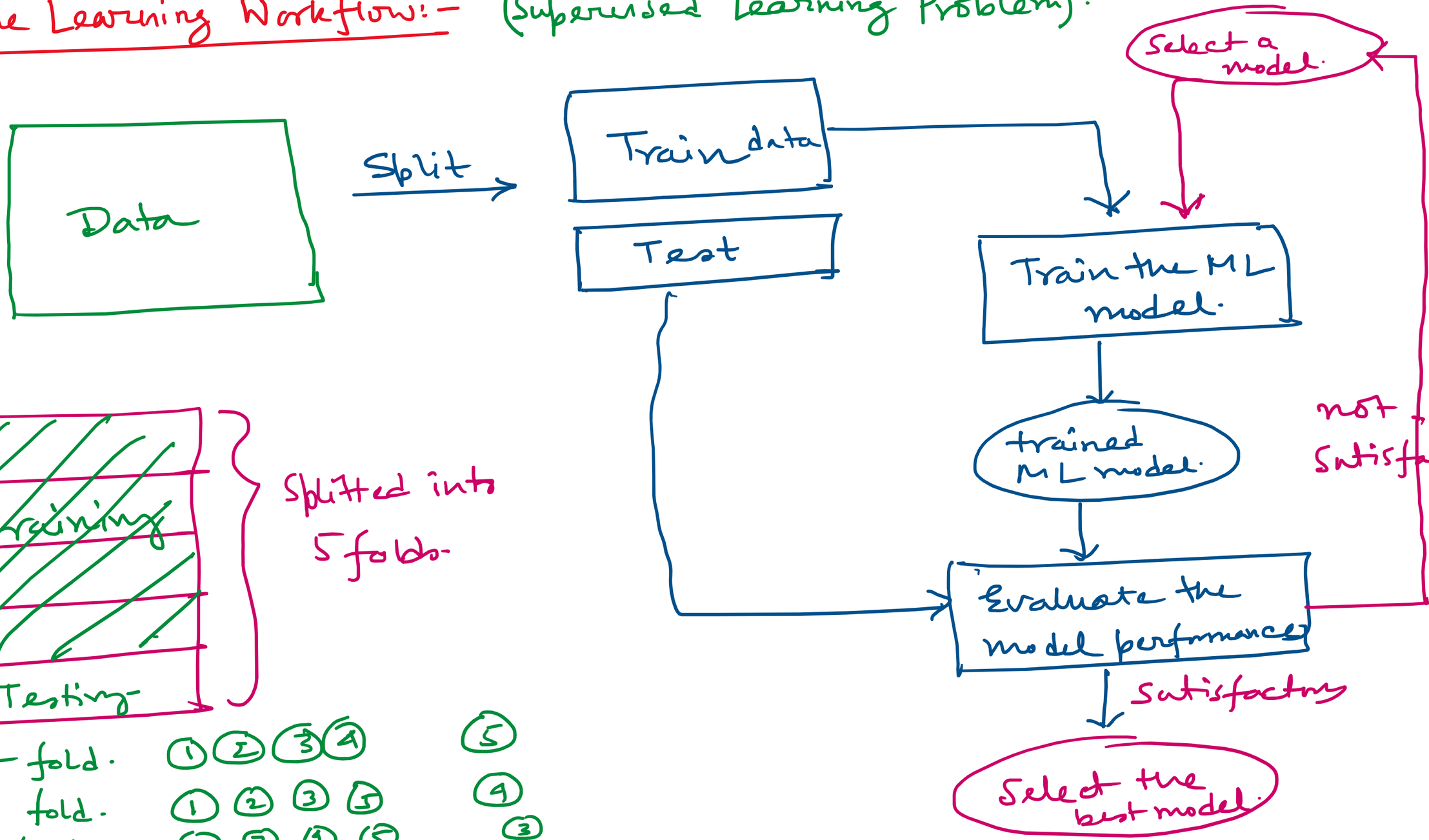
(2) Prediction of score in Oral test based on the score in written test.

classification

If the target variable y is categorical in nature then it is called classification.

Binary classification (only two possible classes). Multiclass classification (More than two classes)

Machine Learning Workflow:- (Supervised learning Problem):-



Data

First fold. (1)(2)(3)(4) (5)

Second fold. (1)(2)(3)(5) (4)

third fold (1)(2)(4)(5) (3)

What Algorithms we have learnt so far?

Regression

Linear Regression

KNN Regression.

Decision tree Regression

Randomforest Regression.

Support Vector Machine Regression.

Classification

Logistic Regression

K-NN classifier

Decision tree classifier.

Random forest classifier

Support Vector Machine classifier.

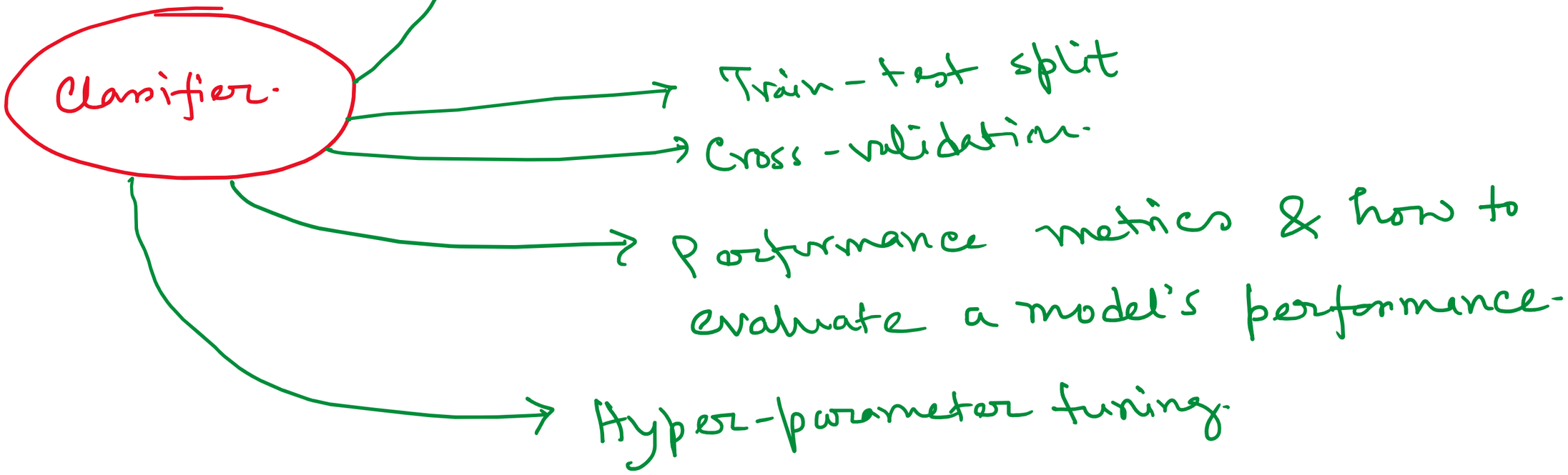
Naive Bayes classifier.

yet to learn.

AdaBoost

XGBoost

ANN


Classifier.
Binary classification
Multiclass classification.
Train-test split
Cross-validation.
Performance metrics & how to evaluate a model's performance.
Hyper-parameter tuning.

Imbalanced Classification Problem

Class	#instance
0	950
1	50

Suppose I train a classifier which produces '0' all the time.

Suppose I feed this dataset → 1000 to our classifier the prediction is 0 for all 1000 instance.

$\frac{950}{1000} \times 100\% = 95\%$ Accuracy.