Machine Learning

① Supervised Learning [2 sessions] ✓

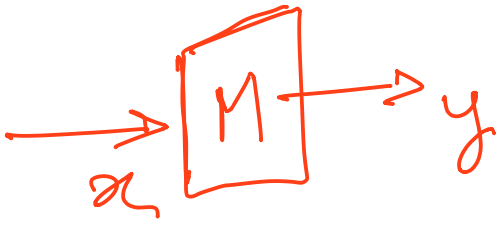
② Unsupervised Learning [2 sessions] ✓ (2)

(2) Regression + (2) Classification

(6)

Unit Test

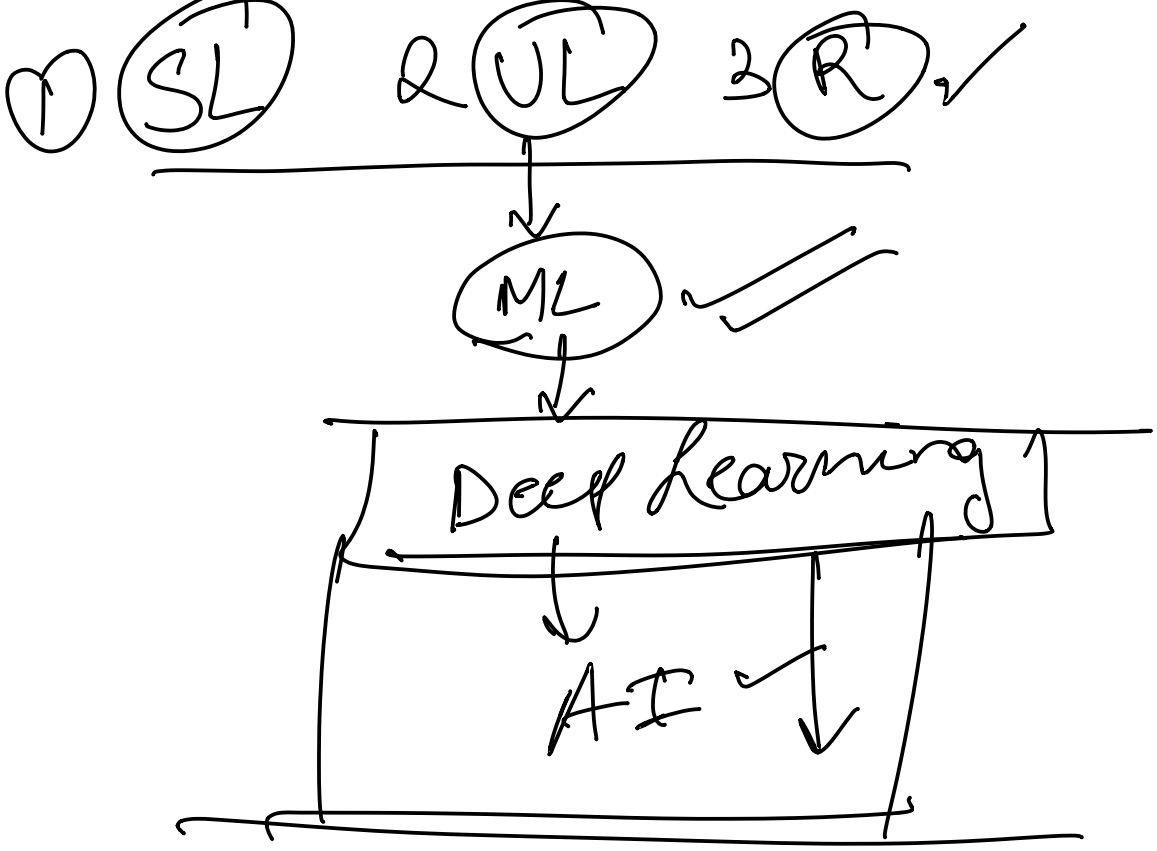
Machine Learning [ML].



* Machine Learns

from data given as input

you receive output



Supervised Learning

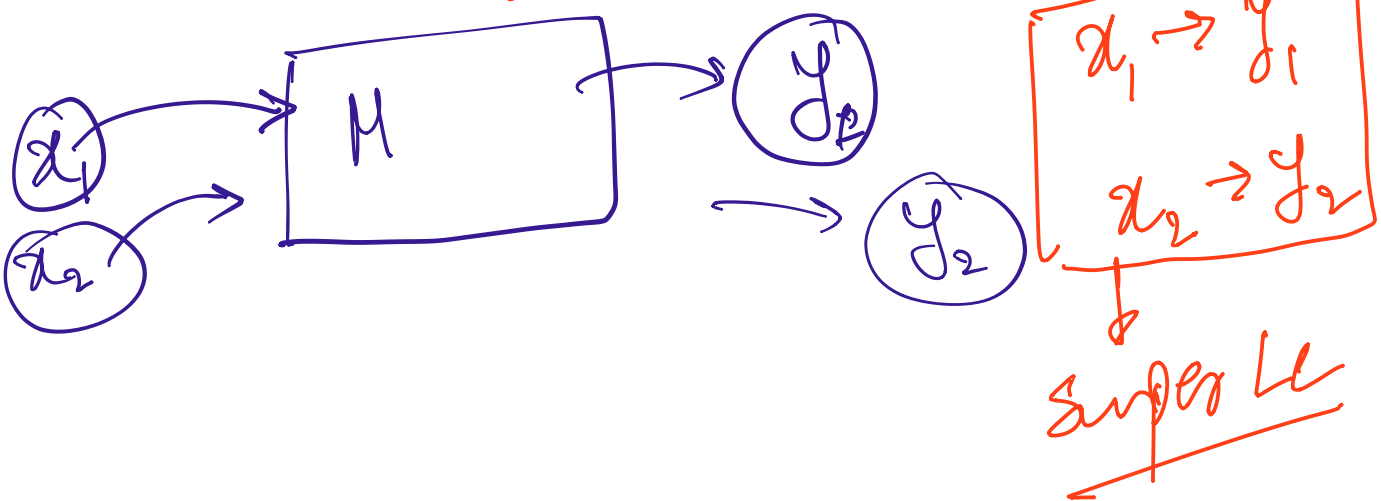
* Labelled data is used for training

input

output

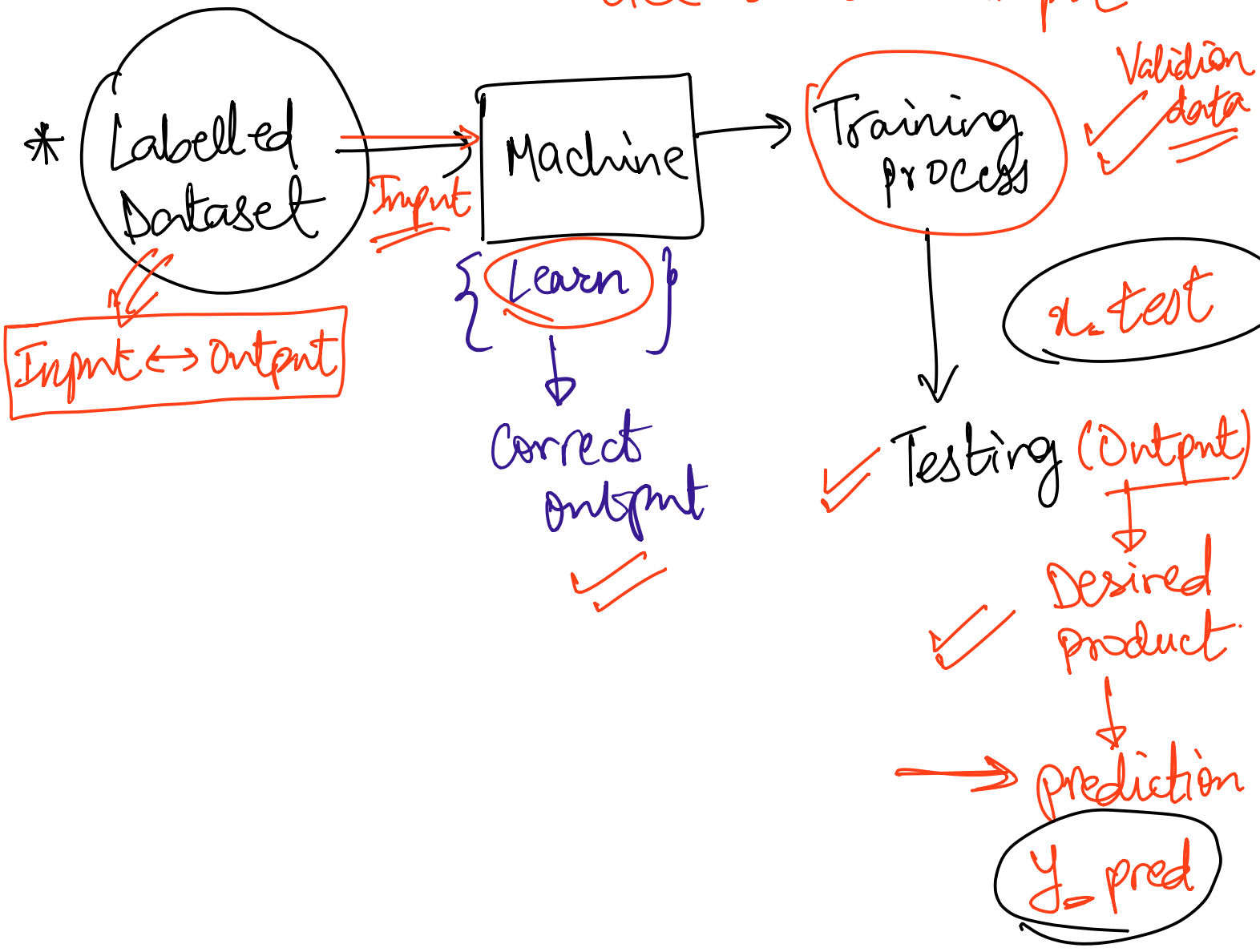
* Supervisor → training data.

* teaches the machine to give correct output.



* Training Data → Input [with output]

Output is predecided acc to an input



Steps Involved in Supervised Learning:

1. Determine the type of training data set
2. EDA [Cleaning / Manipulating data]
3. Visualization → [Corr, Outliers]

Heatmap Boxplot

④ Split your data → xtrain, ytrain, xtest, ytest

⑤ Create a model on a specific algo

Eg) linear, logistic, Random, Decision, KNN, xgboost

⑥ Fit data → xtrain, ytrain

⑦ Prediction → ytest

⑧ Errors → RMSE, R2S

⑨ Confusion Matrix → Accuracy Score → 85.73921... %

↓

Classification Report.

Conclusion:

Model → ① Accuracy on Algo

② Error " "

③ CM → TP, TF, FN, FP

④ Model → % Prob → correct output

Supervised Learning

① Regression

* R/p input & output
* Prediction of output from continous variables

* Eg.) Weather forecast

19°C → 20°C

19.1 19.↑
19.2
19.3

[Continous variable] → Output

Eg.) ① Linear Regression
② Decision Tree Regression
③ Random Forest Regression
etc....

② Classification

* R/p input & output
* Prediction of output from categorical variables

* Eg.) Genders

Male → Y → T
Female → N → F

L.E: Y → 0, N → 1

Eg) ① Logistic Regression
② Decision Tree
③ Random forest
④ SVM (Support Vector Machine)

Example for Supervised Learning.

fruit basket → C, S, T of fruit (output)

C → color, S → shape, T → texture

F.E

Inputs: apple, banana, orange

Training → Data

Test → Predict

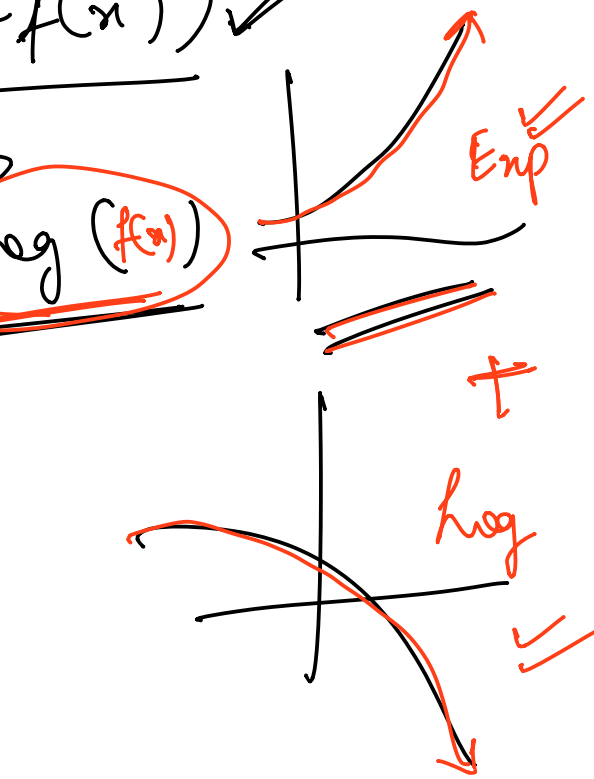
Logistic Regression → (Logit f(x))

Linear Combⁿ → log (f(x))

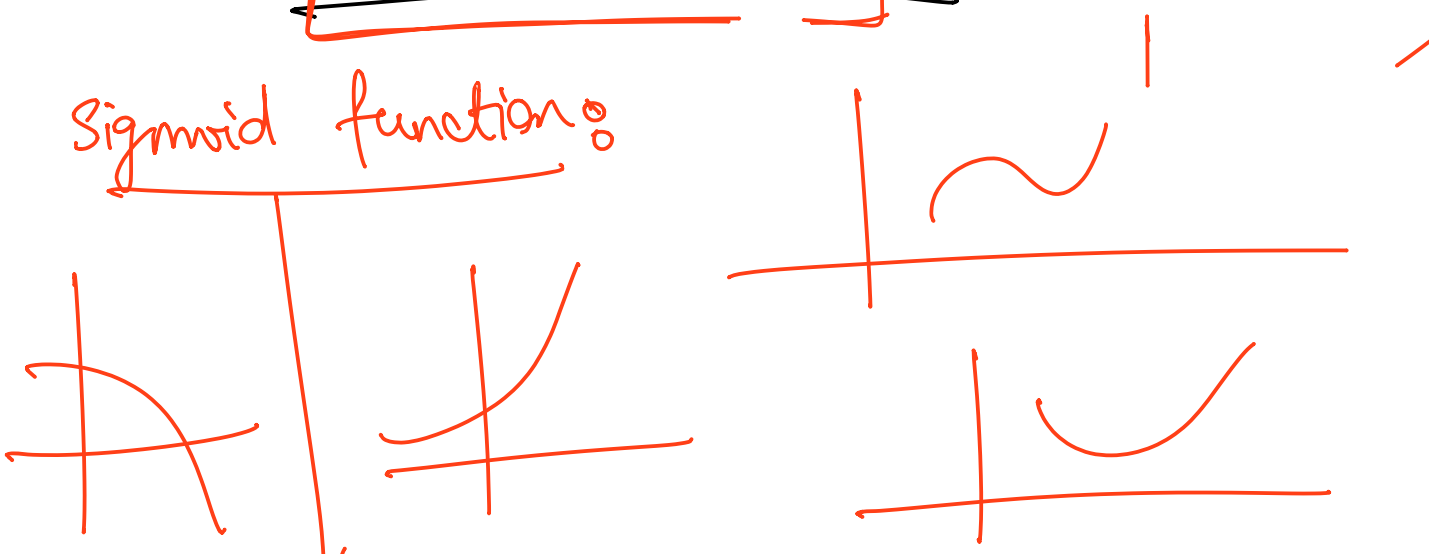
predicted Variables

training data.

Classification.



Sigmoid function:



log (f(x)) → Curve → Classification.

Regression → No sigmoid f(x) Maximum → line.

Evaluating Supervised Learning Models:

(I) Regression

(1) Mean Squared Error (MSE):

$$\sum_{x,y} (\text{Predicted values} - \text{Actual values})^2$$

→ Avg of sqrd values of the diff of Predicted values & Actual values

(MSE) ↓ Performance of Model ↑

(2)

- $x \approx 0$, P.M ↑ ✓✓
- $0 < x < 1$, P.M ≃ ↓ → Unknown.
- $x \approx 1$, P.M ↓↓ ✓✓ → known

★ Limitation → $0 < x < 1$

→ No conclusion can be drawn on the basis of MSE.

② Root Mean Squared Error (RMSE):

* Soln for the limitation of MSE
* Sqr root of MSE

$$\left[\sum_{\text{all } y} \frac{(P.V - A.V)^2}{n}\right]^{1/2}$$

RMSE → (y)

Code → (%) ✓

$y \approx 0$ P.M ↑↑

$y \approx 1$ P.M ↓↓

Conclude
M P ↑
P ↓

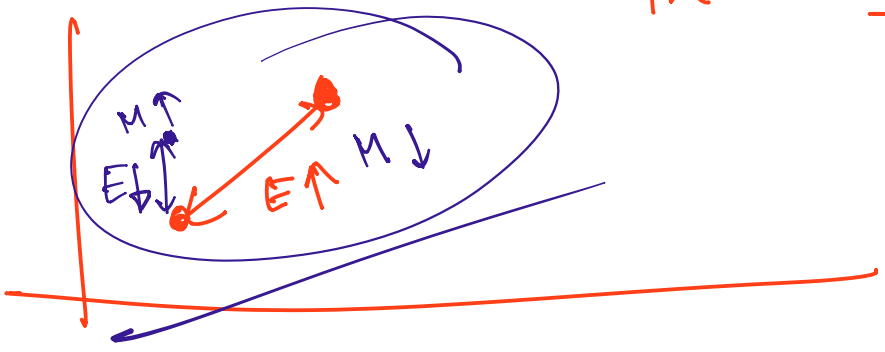
$0 < y < 1$ → $y \approx 1$ → P.M ↓

$y \approx 0$ → P.M ↑

0.23

0.53

Theoritically-



II Classification :

① Accuracy : → % of predictions

↓

$$\frac{\text{Total no. of correct predictions}}{\text{Total no. of predictions.}}$$

3 → CP

10 → TP

Acc → $\frac{3}{10} = 0.3 \rightarrow 30\%$

② Precision : → The percentage of positive predictions

↓

$$\frac{\text{Total no. of True Positives}}{\text{Total no. of Positive Predictions}}$$

Ex:) TP = 5

PP = 15

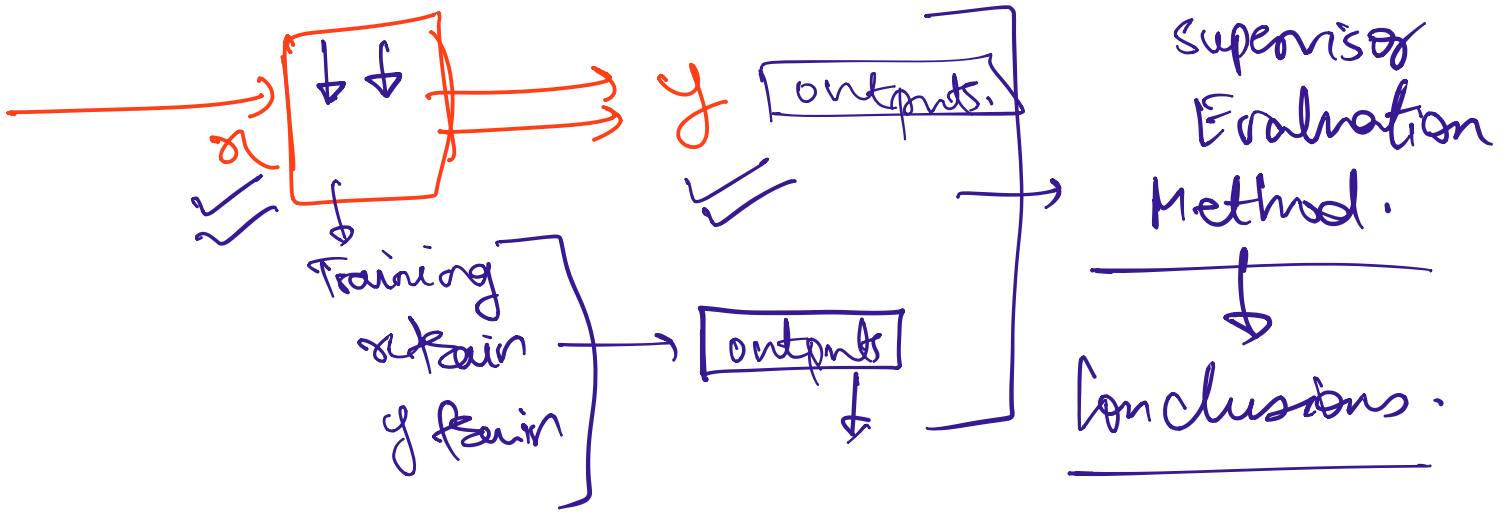
$\rightarrow \frac{5}{15} = \frac{1}{3} \rightarrow 0.33\ldots$

↓

33.33%

③ Recall : → % of all the positive examples a model can correctly identify

→ Dividing the no. of True positives by the total positive examples.



④ **F1-Score:** → weighted average of precision and recall

→ Harmonic Mean of Precision and recall.

F1-Score ↑ , Acc ↑ , Perf ↑

⑤ **Confusion Matrix:**

→ Table showing the no. of predictions for each class and its class labels

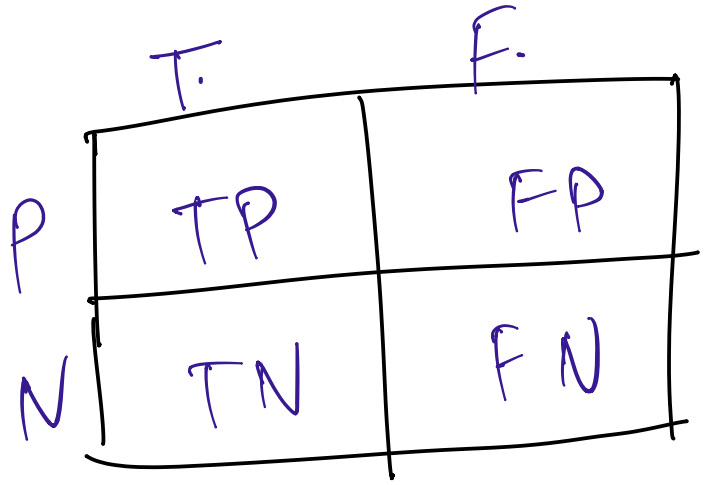
→ ① Visualizing the performance
② Identify areas with issues in model.

T – True

F – False

P – Positive

N – Negative



TP → True Positive

FP → False Positive

TN → True Negative

FN → False Negative

$$\text{Accuracy} \rightarrow \frac{TP + TN}{TP + FP + TN + FN}$$

$$\text{Precision} \rightarrow \frac{TP}{TP + FP}$$

$$\text{Recall} \rightarrow \frac{TP}{TP + FN}$$

F1-Score →

$$\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

↓

Harmonic Mean

$$\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}$$

Things mentioned will be used in Supervised Learning (Reg + Cla)