Machine learning Process Flow

# 1. Data Pre-processing

## Import the data

Load the dataset into your project.

## Clean the data

Remove or fix missing, duplicate, or incorrect values to ensure the data is usable.

## Split into training and test sets

Divide the data into two parts—one to train the model (training set) and one to test it (test set).

# 2. Modelling

## Build the model

Choose the type of model (like decision tree, neural network) that will be used to learn from the data.

## Train the model

Feed the training data to the model so it can learn the patterns and relationships.

## Make Predcitions

After training, the model can now make predictions on new (test) data.

# 3. Evaluation

## Calculate performance metrics

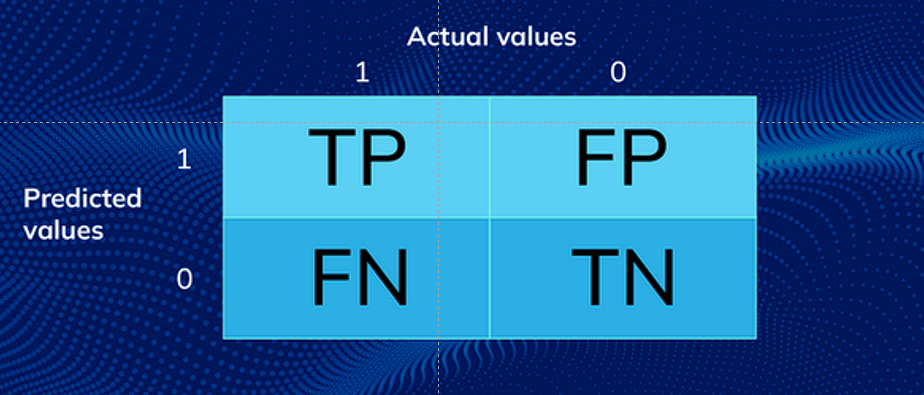
Use metrics like accuracy, precision, recall, etc., to measure how well the model performs on the test data.

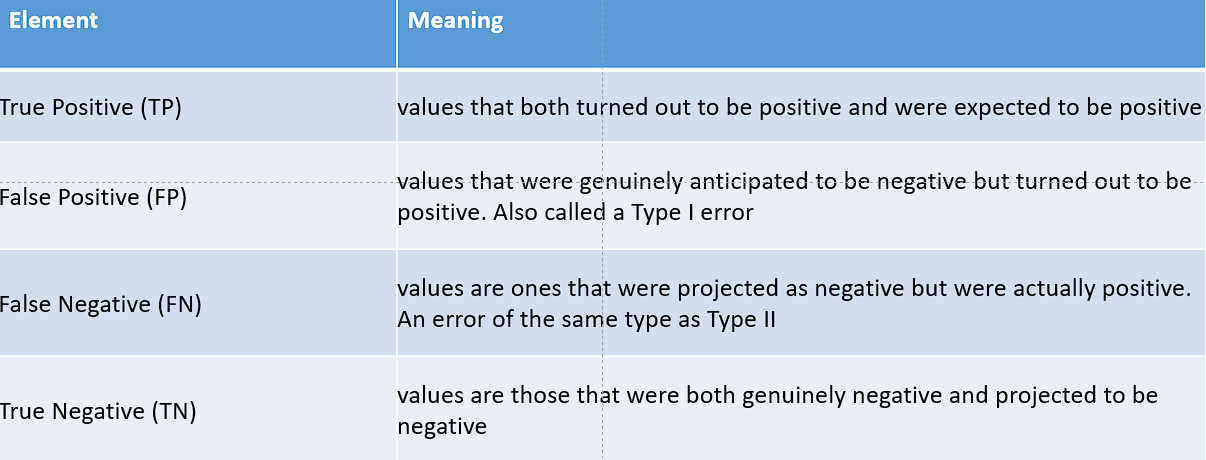
## Make a verdict

Based on the metrics, decide if the model is good enough or needs improvement.

ML model evaluation metrics

## Confusion Matrix





## Accuracy

Accuracy is a metric that measures how often a machine learning model correctly predicts the outcome.

Formula:

tp + tn / tp + tn + fp + fn

## Precision

Within a given set of positively-labeled results, the fraction that were true positives

Formula:

tp/(tp + fp)

## Recall

Given a set of positively-labeled results, the fraction of all positives that were retrieved

Formula:

tp/(tp + fn)

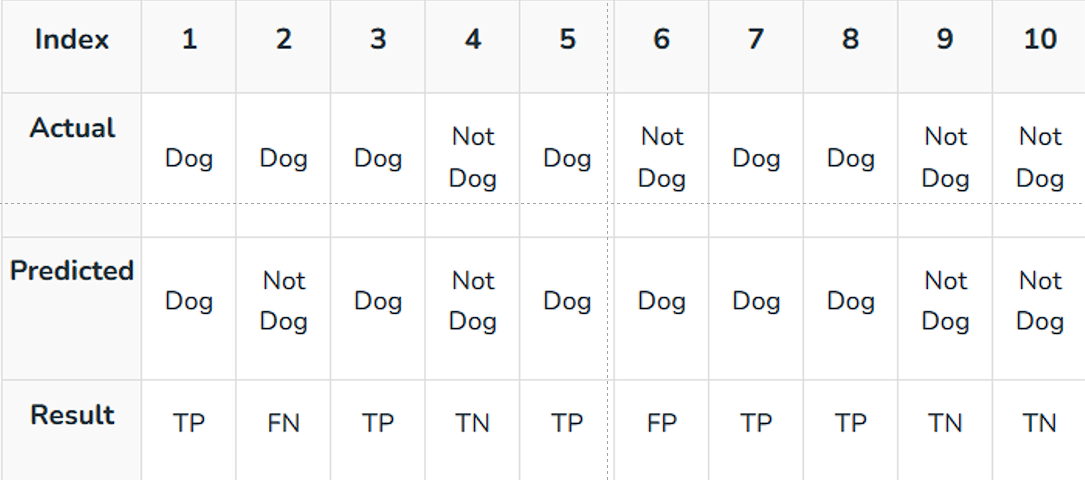
## F1-Score

F-1 score is one of the common measures to rate how successful a classifier is. It’s the harmonic mean of two other metrics, namely: precision and recall.

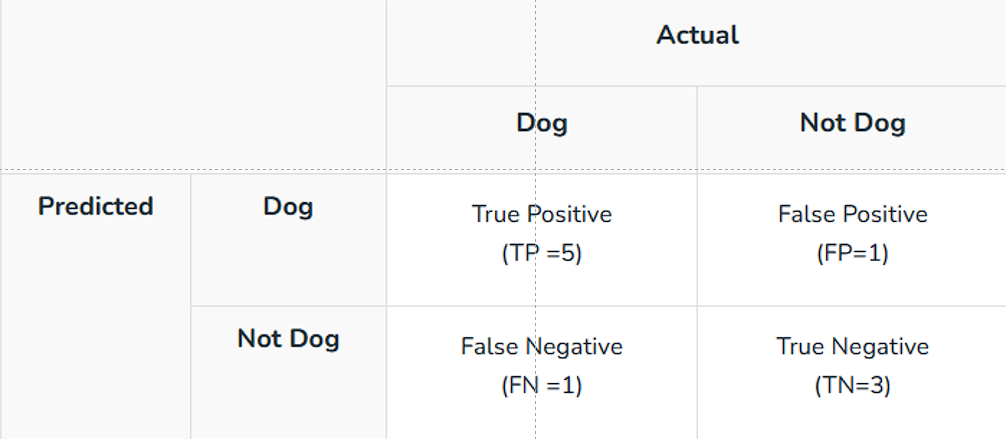
Formula: FOR BINARY CLASSIFICATION

2 \* Precision \* Recall / Precision + Recall

Question –



Confusion Matrix



Accuracy = 5 + 3 / (5 + 3 + 1 + 1) = 0.8

Precision = 5 / (5 + 1) = 0.833

Recall = 5/ (5 + 1) = 0.833

F1-Score = 2

## F1-Score Multi Classication

For Multiclass Classification Problem, we calculate the F-1 score per class in a one-vs-rest manner.

**For a multi-class classification problem, we don’t calculate an overall F-1 score.**

**Instead, we calculate the F-1 score per class in a one-vs-rest manner.**

Question –

