



HOW TO APPROACH ANY MACHINE LEARNING PROBLEM





STEP 1 - FRAME THE PROBLEM

- As a first step, you need to articulate your problem by identifying the type which depends on your business problem.
- Type can be anything like Binary classification, Unidimensional regression, Multi-class single-label classification, Multi-class multi-label classification, Multidimensional regression, Clustering (unsupervised), Other (translation, parsing, bounding box id, etc.)



STEP 2 - GET THE DATA

- Next step is to get the data and store it in the right format according to your problem statement.
- Analyze your data to check whether you have enough data or not and also check the quality of the data.
- The quality of the data fundamentally determines if you will be able to solve the problem at all or not.



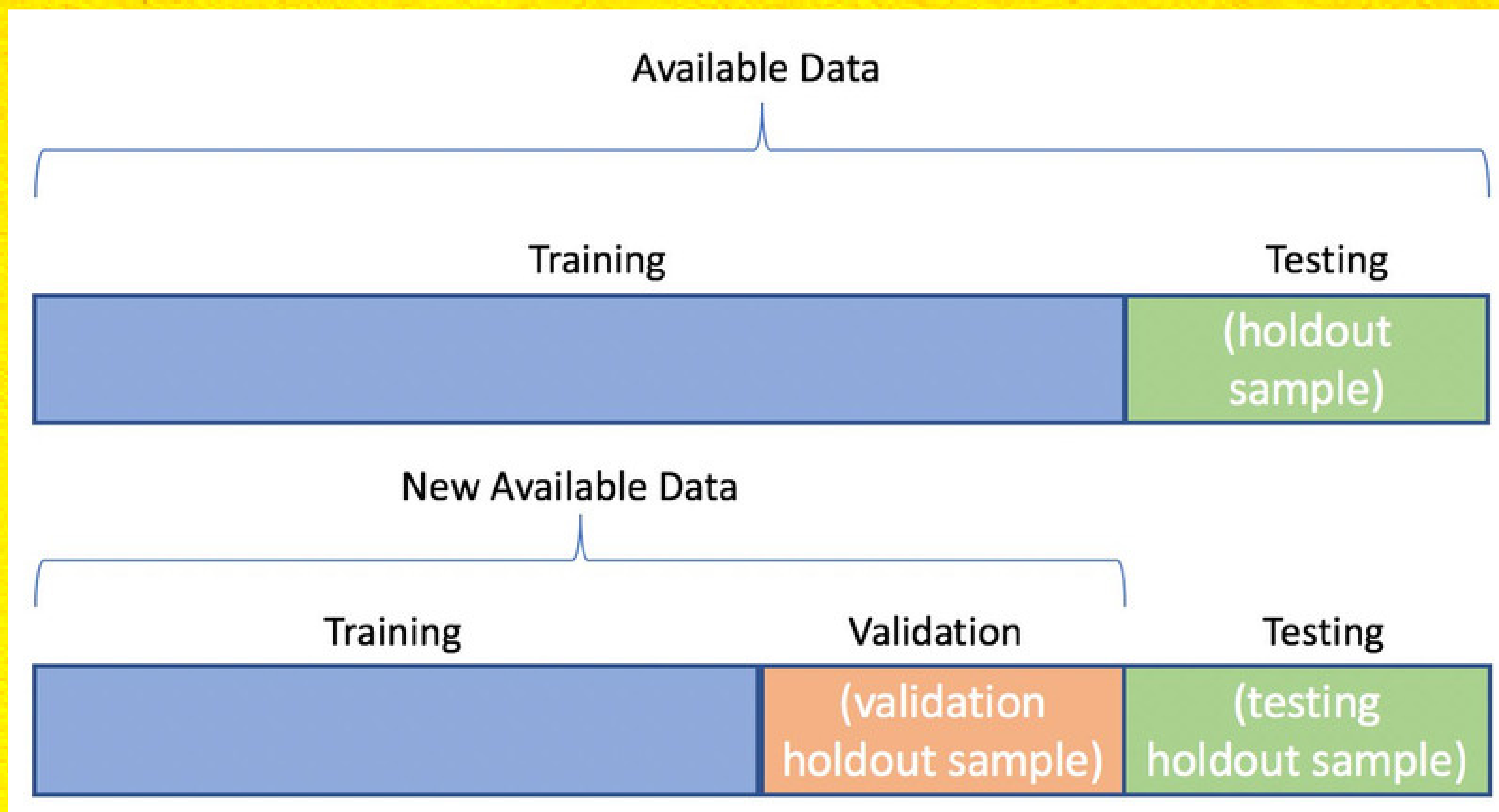
STEP 3 - DATA PREPROCESSING

- After having the data next step is to analyse it and extract insights to make business decisions.
- Also, apply basic data preprocessing operations to bring the data in a go to go format.
- Choose the right library.



STEP 4 - EVALUATION METRIC

- The most important step is to know how to evaluate our results.
- We need to choose the right evaluation metric according to the problem we are going to solve.
- Example - if we have any imbalance dataset then we usually choose ROC-AUC metric.



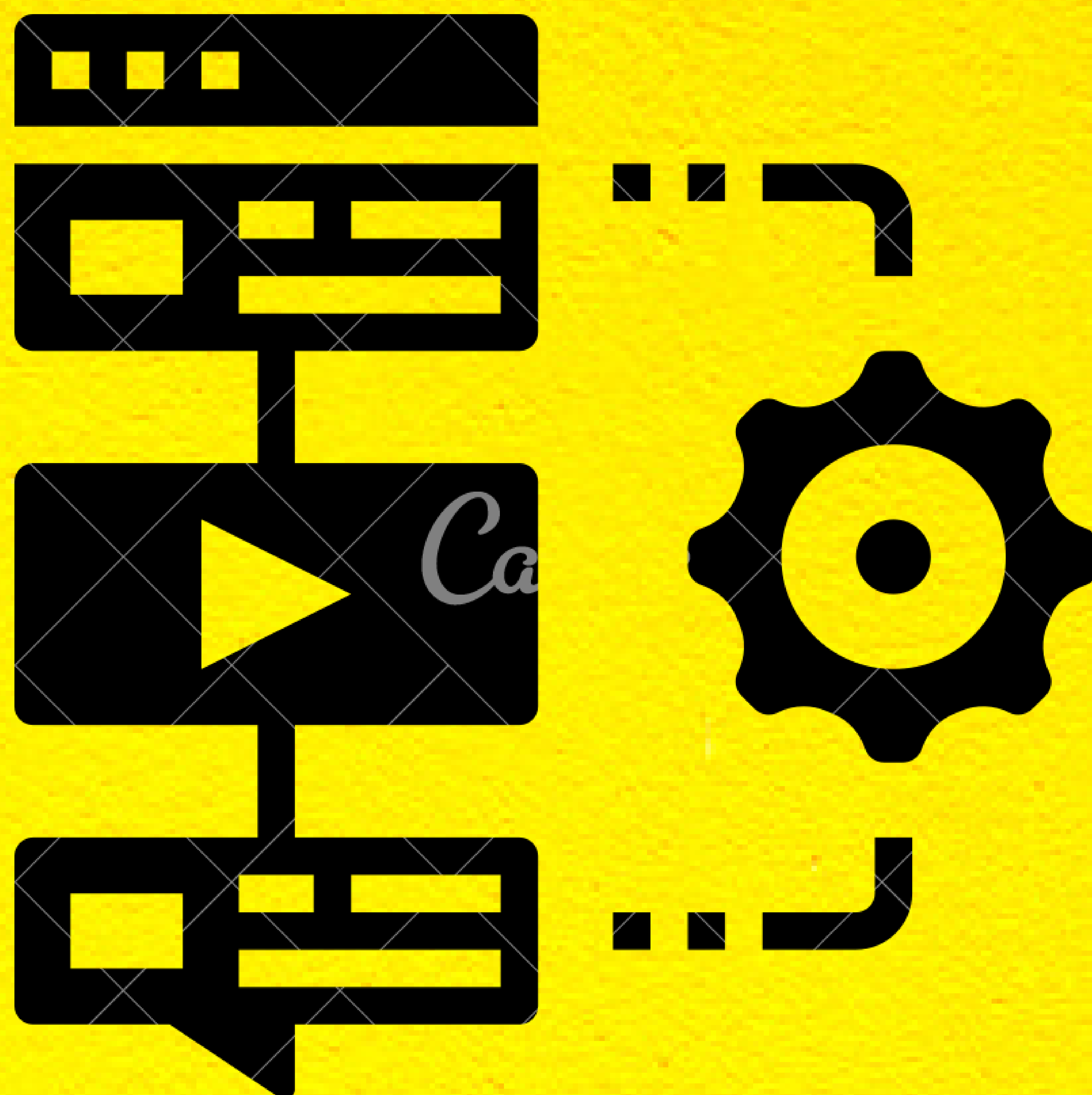
STEP 5 - SPLIT THE DATA

- In any machine learning problem, we split the data into multiple sets like training, validation and test.
- Stratified splitting is the most used for classification problems and K-Fold for regression problems.
- The most important thing to note is whatever the operations you apply on train set must be applied on the validation and test set.



STEP 6 - CONVERT THE DATA

- There can be three types of variables (numerical, categorical and text).
- We need to convert the categorical and text data to numbers and that can be done in many ways (depends on the context of the variable).



STEP 6 - APPLY ML ALGORITHMS

- And finally, we will apply ML models on the data. We can't say which models work best it's just hit and trial.
- Apply multiple algorithms do hyperparameter tuning, evaluate the results and choose the best model which gives satisfying results.
- Benchmark your solution based on your selected evaluation metric.