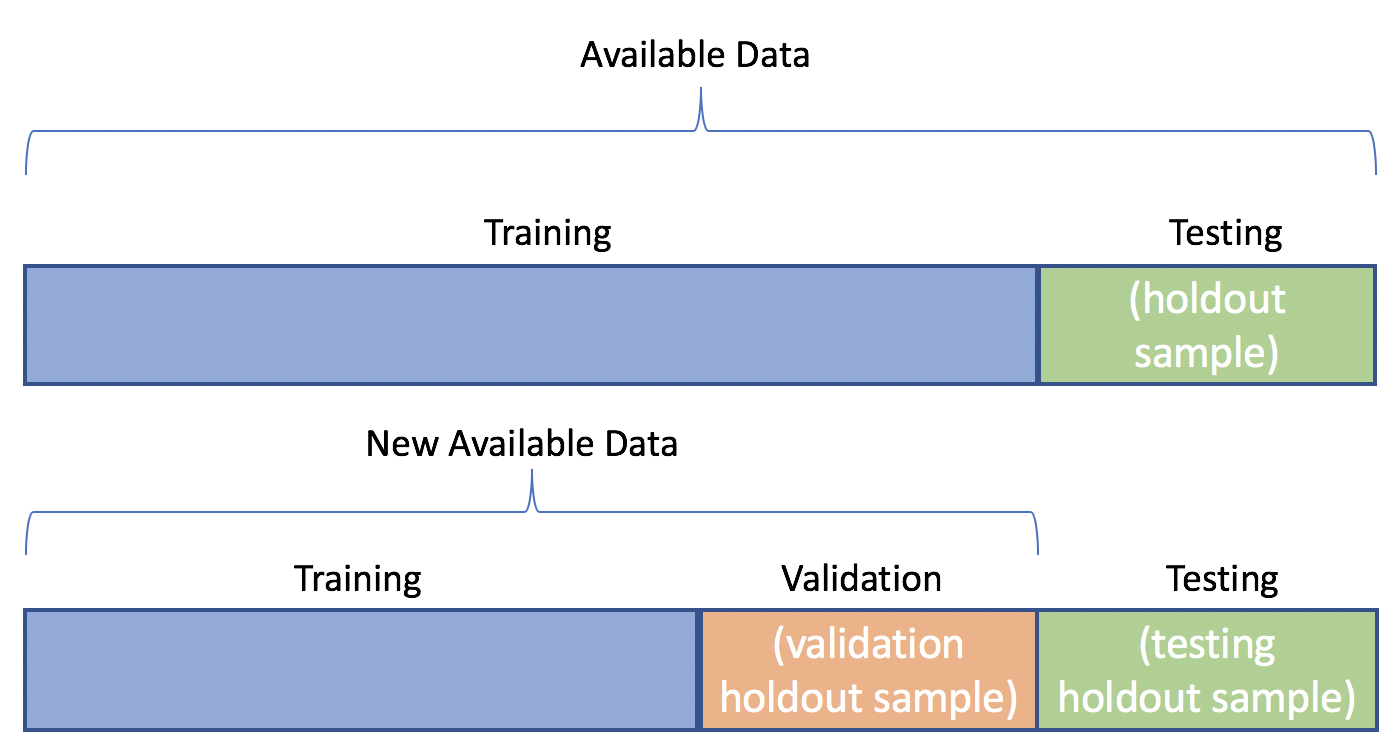
金融大數據 2023/3/24 筆記

* 決策→先想假設我是誰？
* Import Data  
  →將資料先放到Data後，暫存在軟體記憶體內，後再讀取。  
  →讀取速度會較快  
  →但要分享給別人，雙方皆須要確認路徑。
* Train-Test-Split



Split Data  
<https://docs.rapidminer.com/latest/studio/operators/blending/examples/sampling/split_data.html>

1. Linear→Ex.前面800筆訓練，後面200筆測試。適合時間訓練分析，例如預測台積電股價。
2. Shuffled→Ex.抽8筆跳過2筆（80%, 20%），8筆（80%）訓練、2筆（20%）測試。
3. Stratified→分層抽樣，抽取的比例盡量與母體相近。
4. Automatic→若Lable是Nominal則用Stratified，否則就用Shuffled。

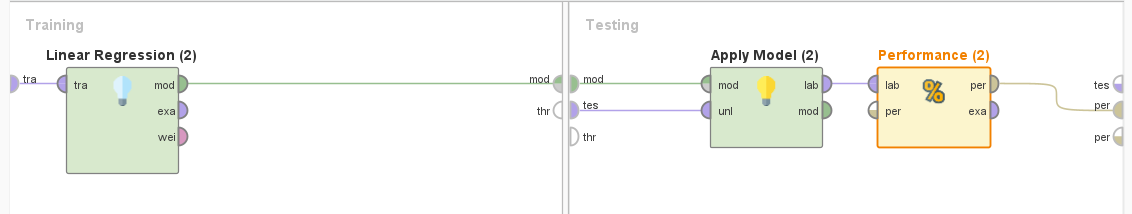
一張含有 文字, 信 的圖片

自動產生的描述

一張含有 圖表 的圖片

自動產生的描述

若為Perfect Match，則為45度角，y=x



* 紫色線為Optional不一定要接，權重調正。

exa→example