

指導教授：李家岩
指導助教：陳彥彰、方鈺學

Data Quality能大幅影響模型結果的好壞，資料缺失、無用資訊會造成「垃圾進，垃圾出」的情形。我們著眼於各情境資料補值的方法論，及資訊量評估指標，透過客觀的量測指標，評估資料處理的變化，維持資料完整性及資訊量。

若SHAP不足以提供適當決策輔助方案，可使用反事實解釋(Counterfactual Explanation)。其實例呈現機器學習模型輸入值之變化如何影響預測結果，直觀、直接地告知使用者如何調整輸入以達到理想結果。

參數會再經過基因演算法最佳化後，交給 FCM 再預測，流程依據使用者需求持續或終止，如設定最小接受域或遞迴次數。

	Age	Sex	Educ	Mar	Relig	Occup	Income	Health	LifeExp	DeathRate	InfantMort	Gas	Class				
91819	1.81571	20.2	25.734	36.077	-14.7	-20.34	-104.8	1280422	7.9056	26.782	34.512	47.654	-20.4	-34.69	-173.3	13734	4
91819	1.81571	20.2	25.734	36.077	-14.7	-20.34	-104.8	1280422	7.9056	26.782	34.512	47.654	-20.4	-34.69	-173.3	40562	1
91819	1.81571	20.2	25.734	36.077	-14.7	-20.34	-104.8	1280422	7.9056	26.782	34.512	47.654	-20.4	-34.69	-173.3	35993	1

最後Counterfactual Explanation 協助決策。表中第一行為輸入資料，後兩行展示建議修改方案。

▲ 上圖為Active Learning 的使用者標註標籤頁面，標註、加入訓練集並訓練模型後，會顯示出該迭代結果與累積數據，如過去所有迭代的訓練準確度之線圖等。

結論而言，Imputers 表現優於常見方法，缺失25%時，對比完整資料僅不到1%準確性下降。其次，Active Learning 提供以使用者意見更新資料 Label 的介面，有效降低訓練成本。知識蒸餾縮小數量之20%，準確性僅下降2%，能節省運算及部署成本。最終，SHAP 解釋資料點因子不同分佈於預測結果(Y)之影響，配合反事實解釋，進行反向推論，協助決策輔助。此平臺強調模型訓練中易被忽視但同等重要的部分，期望能作為資料科學的應用更具解釋性與效率。