2022中軟約農盃

智慧中軟×創新約農競賽

機器學習應用於心臟超音波資料特徵

陳永承 陳姿尹 黃佩琪

專題簡介

我們以醫院方所提供的病患過去三年追蹤診療結果與其各項數值為基礎下去做模型設計，測試不同模型與調整參數，待精準度達預期後便可交付院方使用。
希望製作出來的模型所預測之結果，能幫助醫生快速判斷該病患的合適治療方法，並提升醫療準確率。

資料處理

將院方提供的資料分成以下九組並標記

三年追蹤	1.死亡	2. 持續在心臟內外科追蹤	3. 沒有繼續追蹤，未在中榮看診
1. 心導管電燒術	11	12	13
2. 心室去顫器	21	22	23
3. 藥物控制	31	32	33

將標記完的資料清除空值後，把重複或非必要的特徵刪除(ex:id、index)，再把一些特徵做數值轉換(ex:性別男/女用0/1代替)，即可帶入模型訓練。

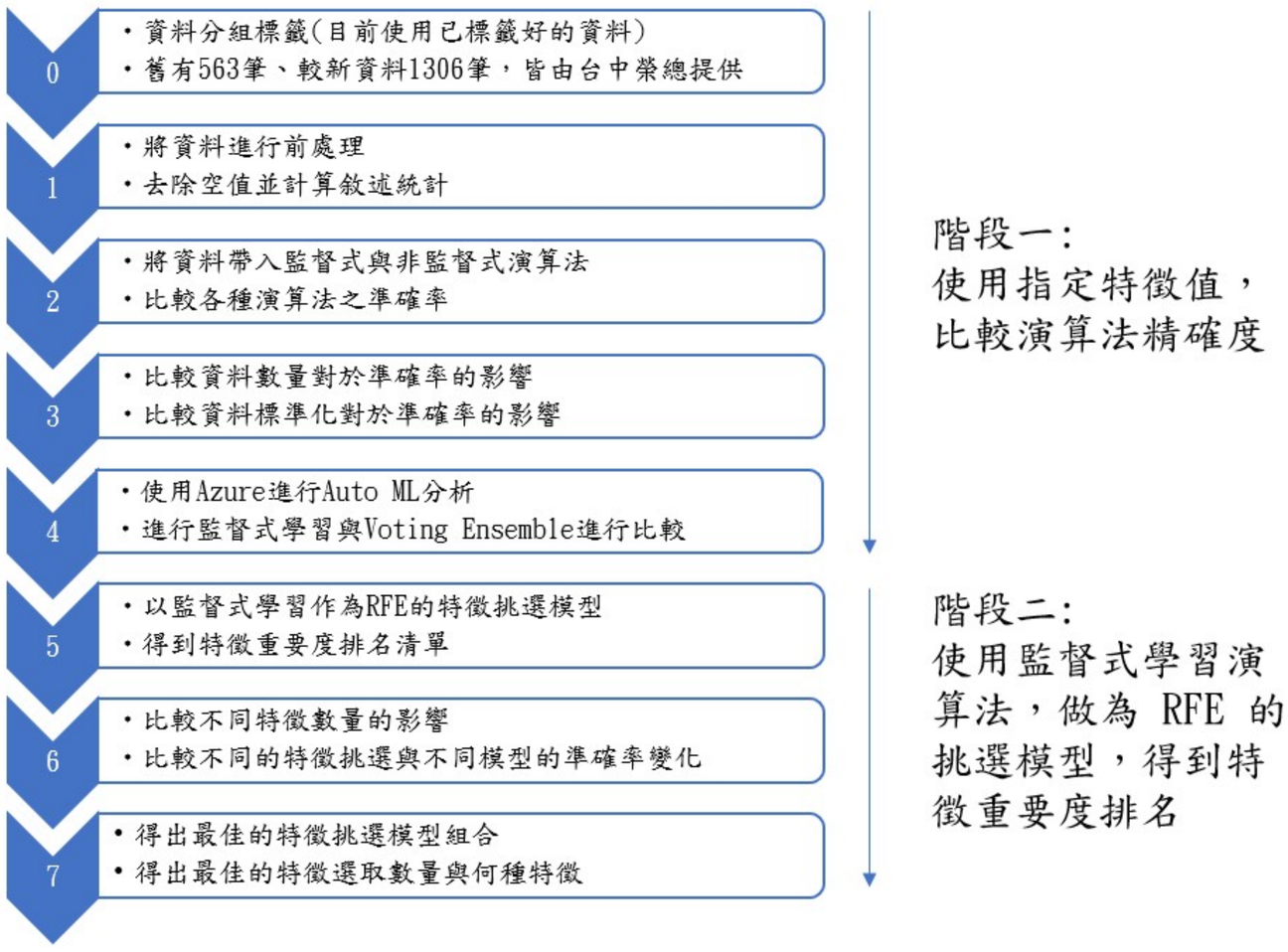
演算法

監督式學習:
Random Forest、XGBoost、LightGBM

非監督式學習:
AffinityPropagation 、Hierarchical Clustering、Gaussian Mixture Clustering、K-Means、Mini Batch K-Means、Mean Shift、Spectral Clustering

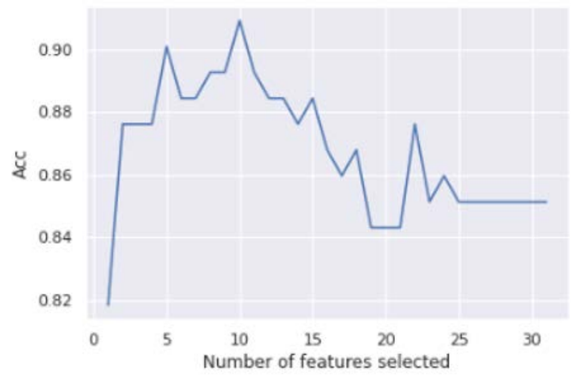
另外新增:
Voting Ensemble 、 Lasso 、Ridge、ElasticNet、Catboost 、 HDBSCAN

流程結構



▲流程圖

成果展示



▲將特徵數量與準確率變化圖形化

	XGBoost 挑出特徵	RandomForest 挑出特徵	LightGBM挑出特徵
XGBoost 作為預測模型	0.8678(10)	0.9174(11)	0.9008(5)
RandomForest 作為預測模型	0.8843(2)	0.9008(10)	0.9091(5)
LGBM 作為預測模型	0.9008(16)	0.9091(11)	0.9091(10)
Voting Ensemble 作為預測模型	0.8926(17)	0.9174(11)	0.9174(10)

▲準確率和特徵數的比較表格

▲供使用者使用的網頁介面

▲網頁顯示結果