網路多媒體實驗：利用model output 偷取 Machine Learning-as-a-Service 平台 model

組員：林承德、趙冠豪、李羚毓　　指導助教：曾煒傑　　指導老師：林宗男教授

1.實驗目標：

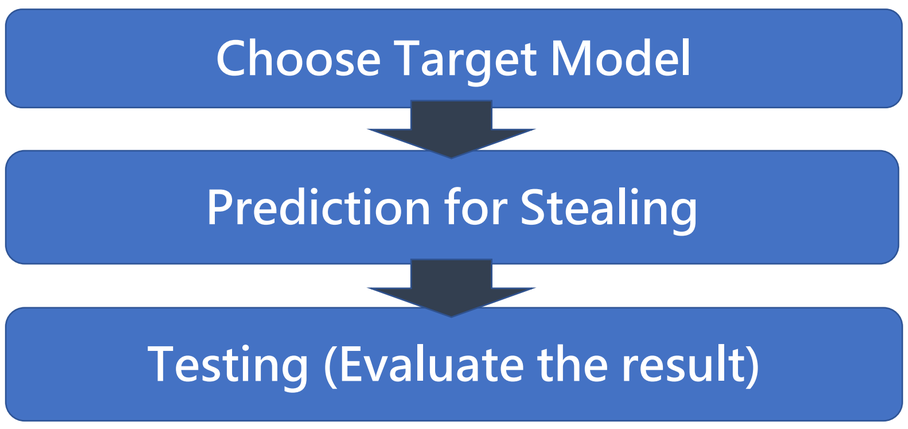
1) 實作論文 ”Stealing Machine Learning Models via Prediction APIs” 的方法，偷取ML model。

2) 利用 neural network 的方式嘗試偷取各種不同模型的 Model

2.實驗介紹：

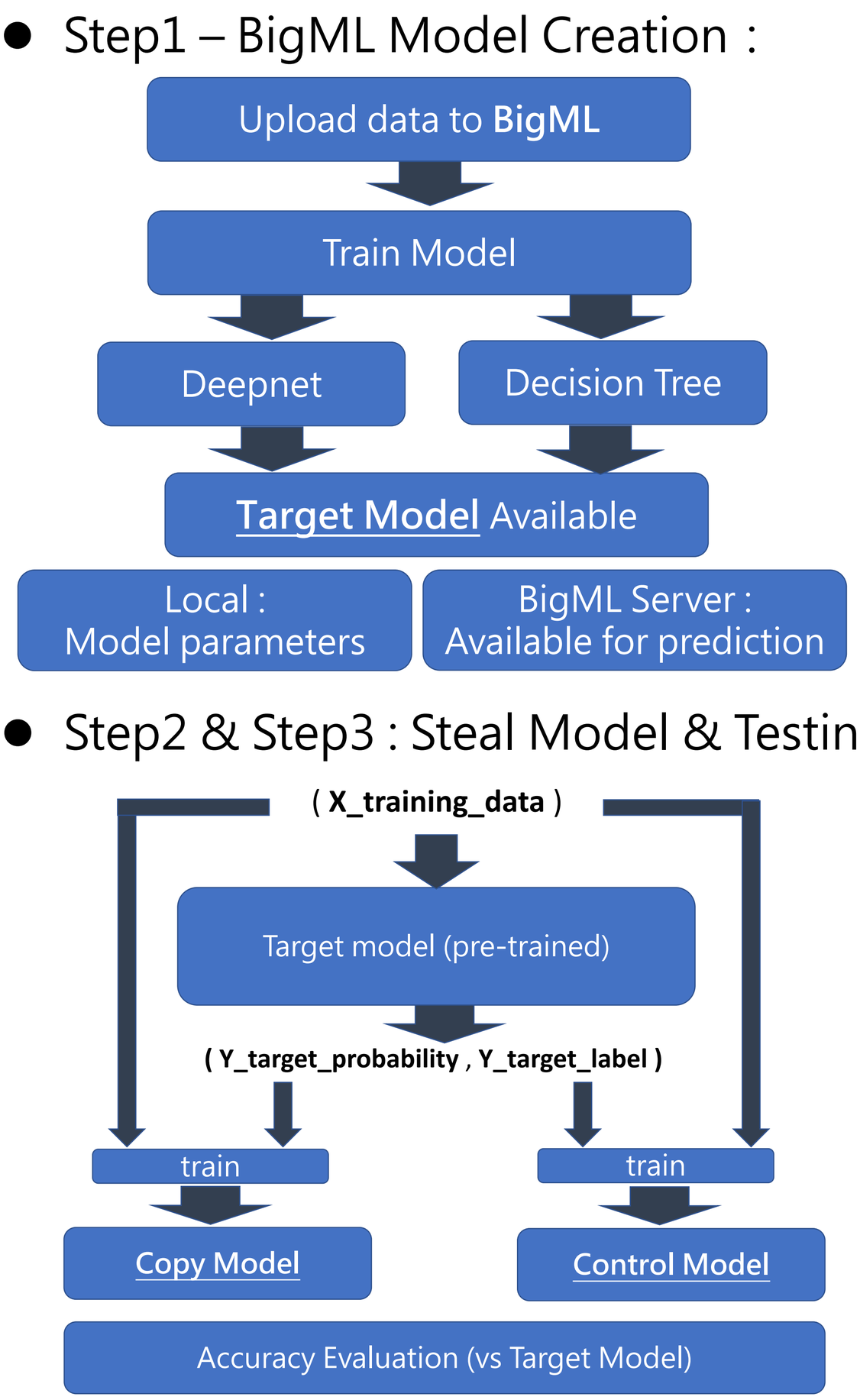
- 資料介紹 ：MNIST 資料集

- 流程圖：



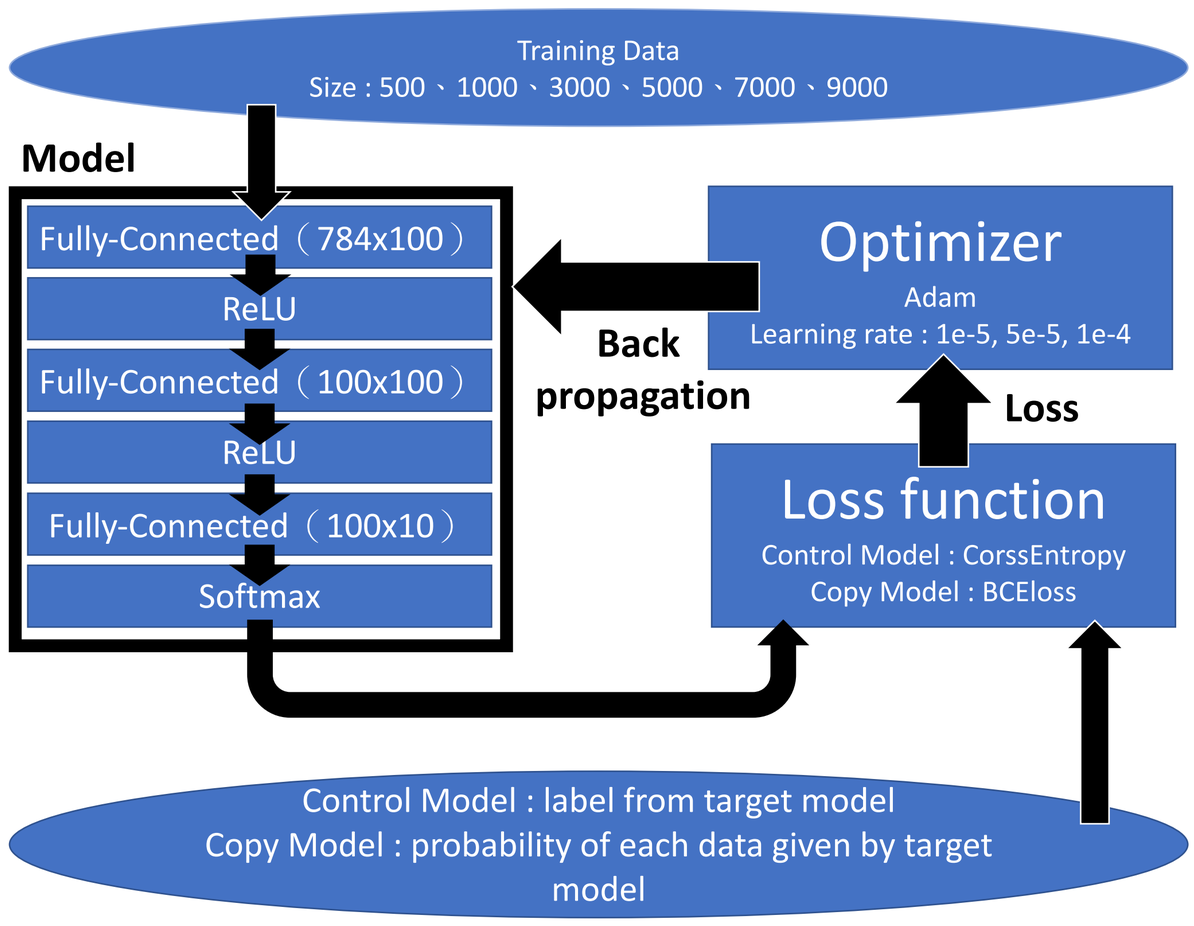
3.實作過程：

- Step1 – BigML Model Creation :



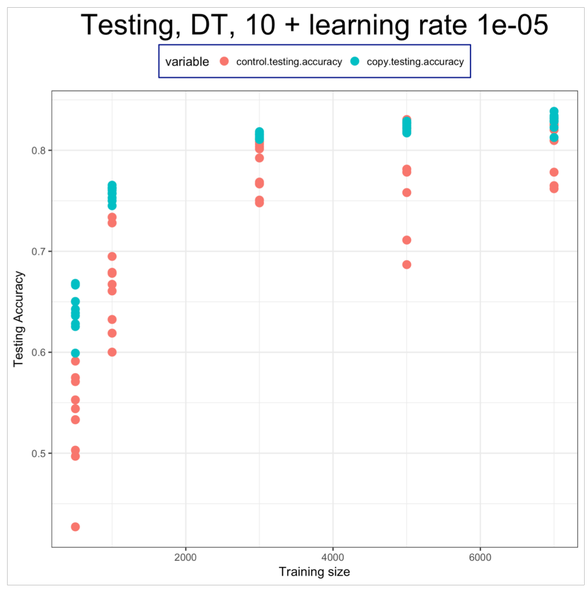
- Step2 & Step3 : Steal Model & Testing

- Step2 & Step3 : Steal Model & Testing

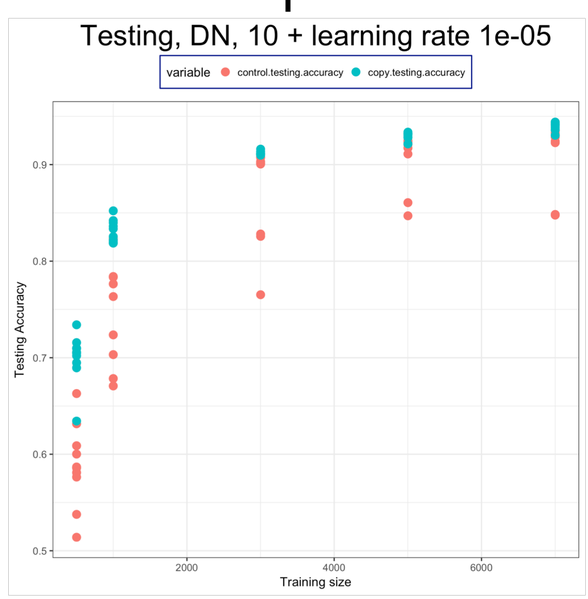


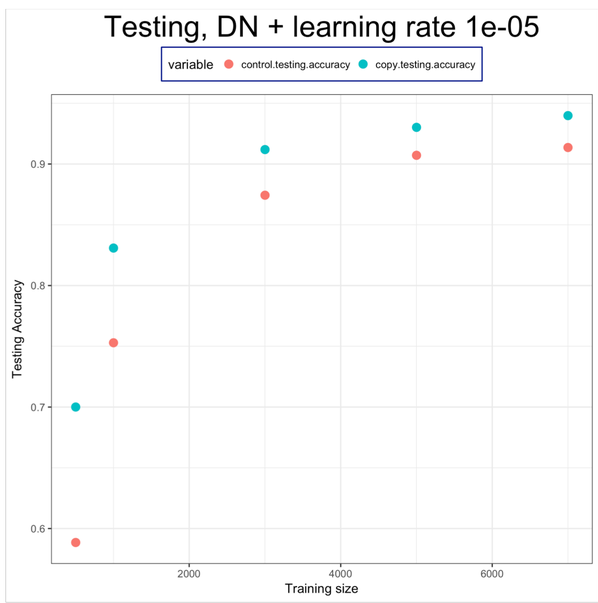
4.結果

- Decision Tree :



- Deepnet :





6.結論 & 討論

1. 在不同Training size, learning rate 狀態下，Loss function 為 BCEloss ，200 epoch，Copy model 的準確率皆比 Control Model 高。

2. 用Neural Network 偷取 Decision Tree Model 以及 Deepnet Model，都可以有效的Copy

3. 在有限 Training data 下，Copy Model 的準確率比Control Model 的準確率高(約 10 %)