10994015 資管碩二 李承諺

本次心得我挑選許政堯同學的論文計畫來撰寫。政堯研究大致流程是先進行資料蒐集，並用一些資料不平衡的處理方式去處理資料不平衡的問題，最後丟進深度學習或者機器學習模型，來判斷虛假評論的檢測。而看完本次報告我有幾個問題想要請教。

**一、**

在生成資料時，真實資料與生成資料的比例該為多少，多少時會是最佳狀態?或是這也可以當成實驗的一部分?`

因為我自己的論文也有遇到類似的問題，在資料不足時，也需要一些資料增強的方式，但某些資料增強的方式看起來不太適合將1:1的去做增強，因此該使用何種比例才是最佳狀態也成為了一種問題。

**二、**

以往有沒有先對資料做資料不平衡處理再進行虛假評論分析等相關工作，或是他們是用哪種方法成功，而政堯跟他們做得有什麼不一樣。而虛假評論的檢測是需要用那麼大量的演算法或者模型才可以處理的有效嗎?會不會造成演算過多導致不準確，還是越多模型混合去訓練會越有效，這也是我在看報告時會好奇的點，雖然有提到最後會進行交叉驗證，但如果不需要那麼多準確率就夠高，相比之下會不會更輕鬆方便，也不必要耗費那麼大的工程來做訓練。

**三、**

訓練的模型有非常多，但市面上模型其實滿山滿海，為什麼政堯會挑選出那些特定的模型來做為訓練?不知道是我沒認真聽還是政堯沒有解釋，而特別選那些模型的原因又是什麼?是因為親民好用，還是效率高，或是準確率高，希望之後此論文能夠給多些解釋。

**個人疑問:**

處理資料不平衡是否也有其他方式，因為政堯在報期中的時候與期末的時候內容有些出入，不曉得是因為不適合或是怎麼樣先刪除，像是期中時政堯有提到資料不平衡的處理可以用生成對抗網路的方式去生成資料，而我認為這也是一種資料不平衡處理的一種非常好且可以嘗試的方式，但期末時似乎被刪掉了我認為有點可惜。