NTU FinTech Fall 2019 Final Project Report

NTU CSIE R08922024 胡安鳳

1. 前言: 使用LSTM，或同作業二採技術指標即可?

答案: 以CP值而論

很明顯本期末專題是時序性資料預測，一定想到用回饋式神經網路RNN或長短期記憶網路LSTM訓練。神經網路固然強大，卻可能存在一個致命缺陷: 過擬合 (Overfitting) : 在initail submission很好的分數，有可能在最後十四天的評測中，大大落敗 (而14天也較短，無法看出以多年資料作訓練集從長計議的優勢，例如模型以半年至一年的scale有賺頭，但在未來十四天的評測，因時間長度不夠，短線而言可能無法給出好結果)。

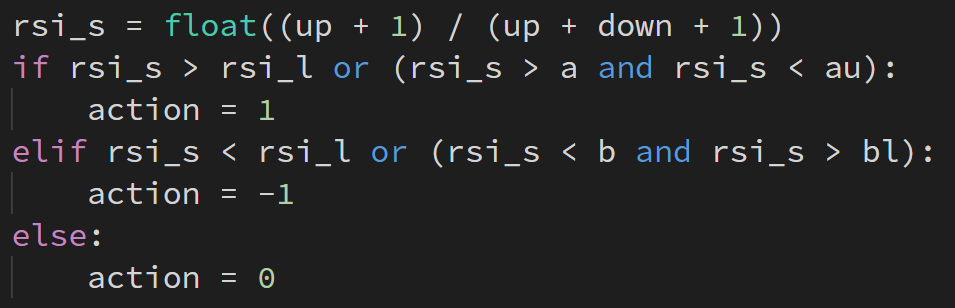
或許只要layer by layer fine-tuning，搭配Drop, pruning即可避免。但會耗不少時間，效果亦未知。

期末繁忙下，以類似作業二方式實作，會是較高投資報酬率 (指所付出的時間心力和所獲得的成果) 的方法。

1. 何種技術指標? 如何優化?



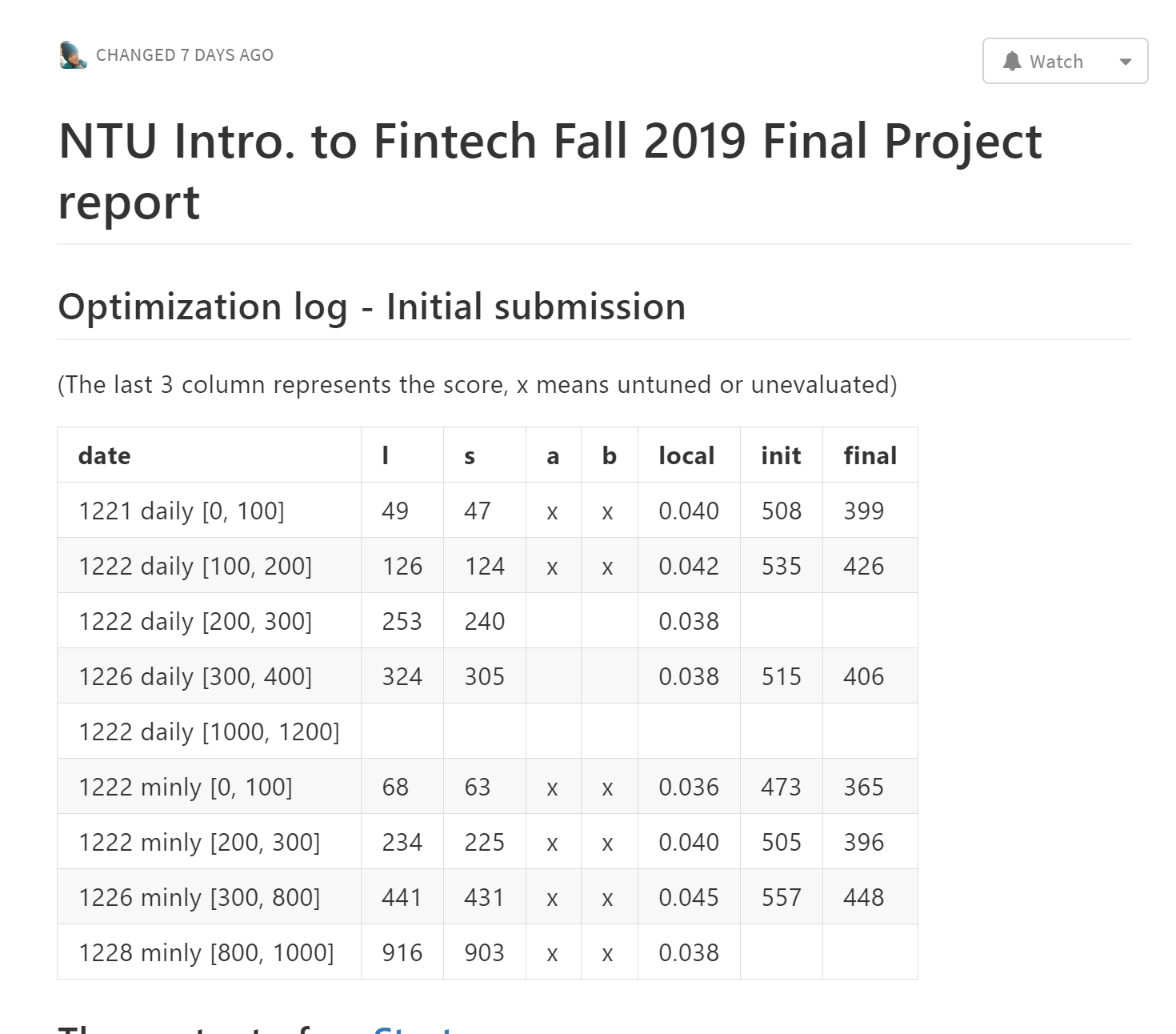
由上台指期走勢圖得知，指數雖沒有如亞馬遜公司在近幾年大幅竄升，卻是穩穩成長。選擇最基本技術指標MA、RSI即可。第二次作業，我用 RSI作指標，搭配短期RSU數值限制在某範圍配合各支股票fine tuning，而獲得了高達5.5萬分的高分成績位居全班第八。



MA不採用乃因測試數次後效果皆不如 RSI (MA單純採用**移動平均**，易受到**極端值**影響；然RSI是**量化的結果**，可清楚看出股票"**潛在勢能**")。

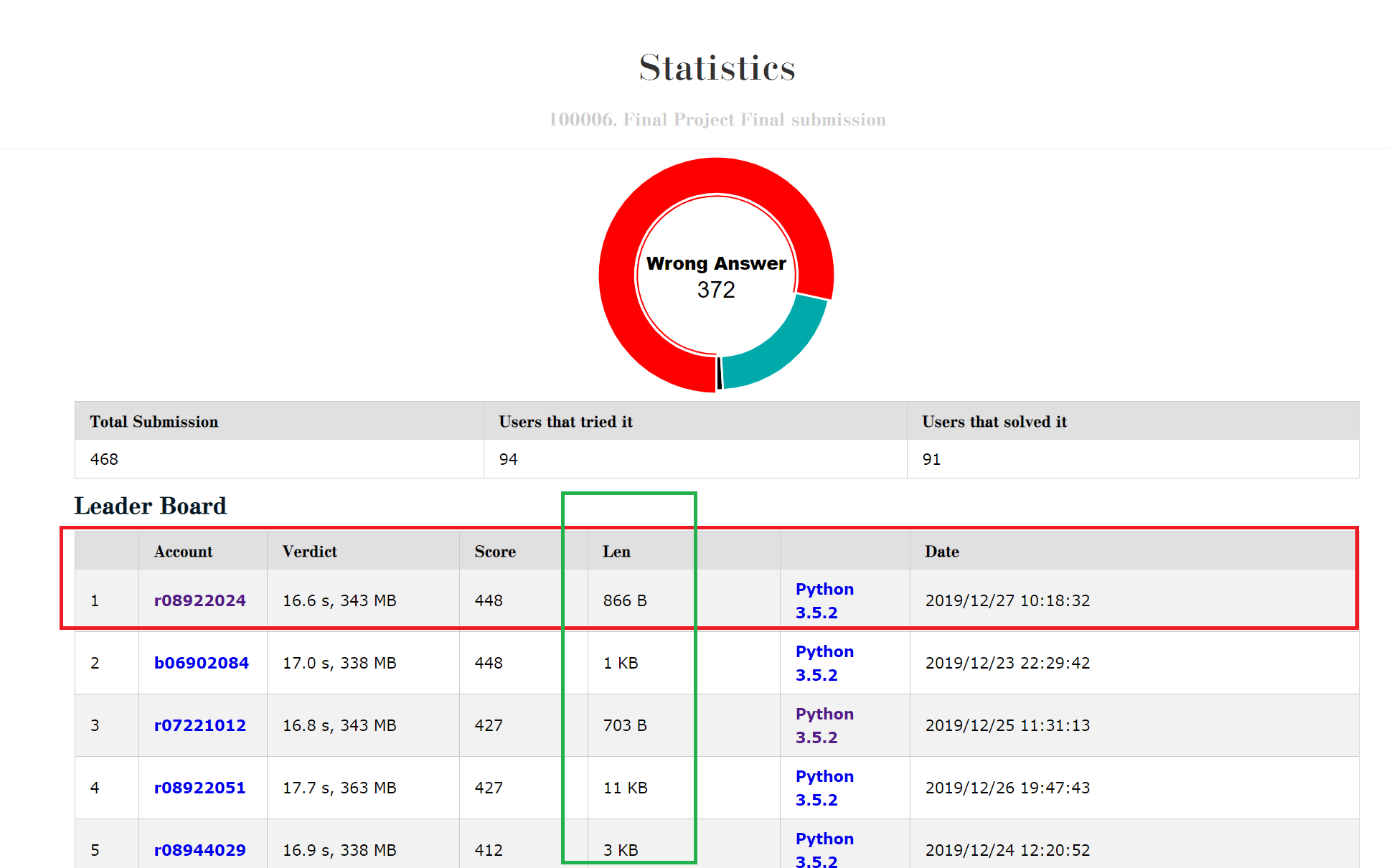
本期末專題捨棄短期RSI額外設上下界 (上圖的 rsi\_s > a and rsi\_s < au)，以免過擬合，純粹採用 rsi\_s > rsi\_l 作為買的指標，再暴力搜尋s、l的參數(天數)即可。

又在HackMD作搜尋結果log，分別記錄以分鐘和以天為單位兩個csv檔參數搜尋結果，避免重複找參數區間而顯冗贅。



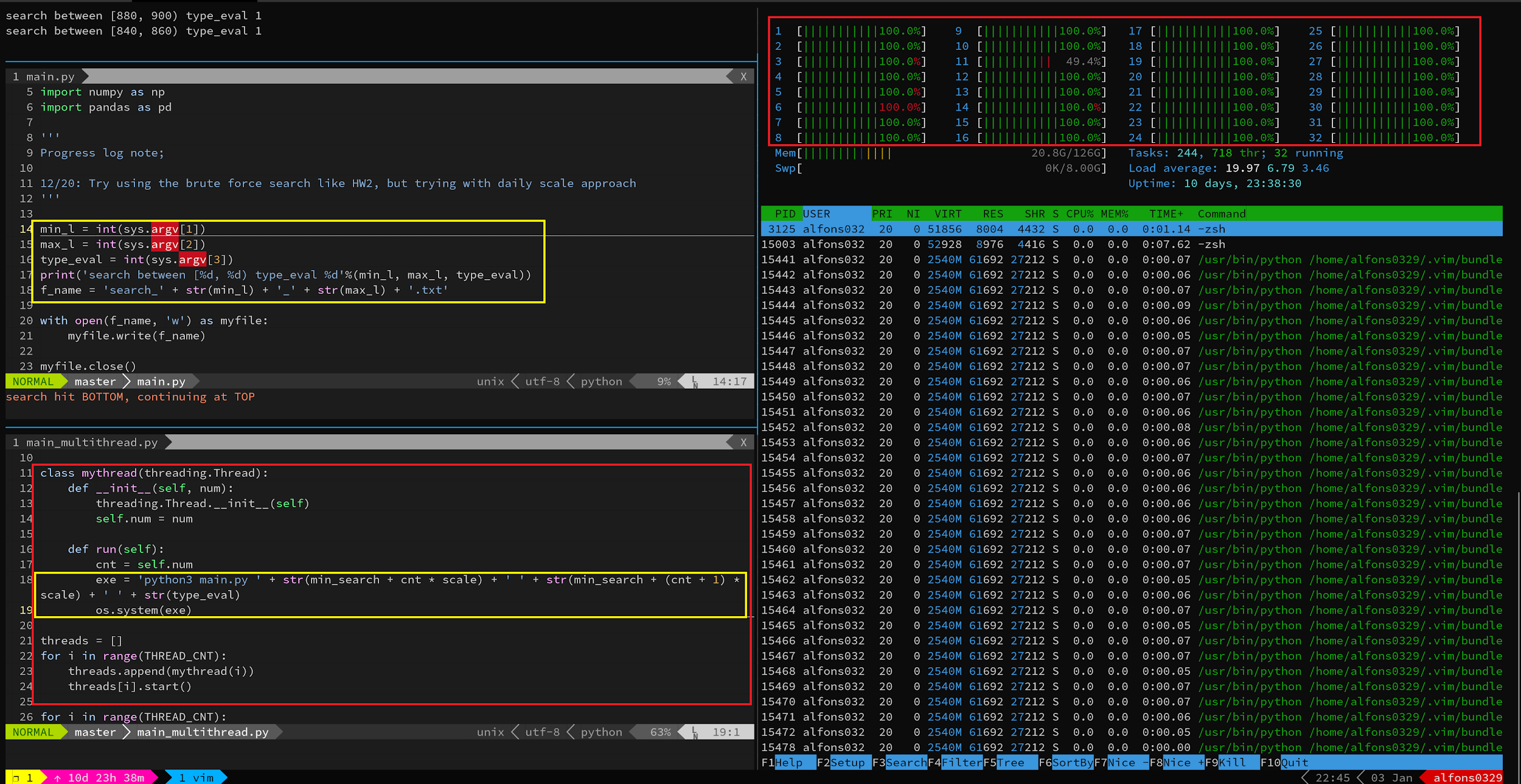
最後，不論**initial** 或 **final submission**，我都獲得了**第一名**，可見該參數模型具有一定的股價評判能力 (final submission 到正式開始評判前，已有超過 2019/12/16的資料，**完全不存在助教給的csv**中，可視為**testing set**概念，而得以客觀評價)

程式碼**既短、快速又高分。**



1. 如何加速參數搜尋? 以達最少時間最大效果

暴搜參數相當曠日廢時，用python multithread，各自搭配 os.system(‘python main.py a b’)，搜尋 [a, b)區間傳入 argv，既不用煩惱critical section問題，又能自動化，一舉數得。 (詳細程式碼請見 main.py, main\_multithread.py)



1. 評估(如果有 1/10的結果再補寫這裡)
2. 未來展望
3. 心得