

How to present the Kaggle project you select for term project?

You should focus on these aspects:

1. [problem] What kind of the problem the project is? Briefly state the problem/goal it addressed. Why is it difficult and a challenge? What are the performances of previous methods on the problem (or similar problems) if any.

這個 project 的主題是用 LSTM 模型預測股票走勢，資料包含 S&P 500 中各股的盤後整理資料，收集期間為 2003 年到 2017 年，且預測的 RMSE 可以到達 0.1 以下。目前它可以針對資料中的股票預測未來走勢，但只有美股，對於常買台股的台灣人來說不太適用。

2. [data] What are the data types, format and size and available resources? Does the data need processing?

資料類型為 csv 檔，筆數為 3500*n(n 由我們所抓取的股票數決定)，由於資料為自行抓取，故全部為可用資源。

資料內容包含日期、代號(不同代號代表不同股票)、開盤價格、收盤價格、最低價、最高價、成交量

	date	symbol	open	close	low	high	volume
0	2016-01-05 00:00:00	WLTW	123.430000	125.839996	122.309998	126.250000	2163600.0
1	2016-01-06 00:00:00	WLTW	125.239998	119.980003	119.940002	125.540001	2386400.0
2	2016-01-07 00:00:00	WLTW	116.379997	114.949997	114.930000	119.739998	2489500.0
3	2016-01-08 00:00:00	WLTW	115.480003	116.620003	113.500000	117.440002	2006300.0
4	2016-01-11 00:00:00	WLTW	117.010002	114.970001	114.089996	117.330002	1408600.0

3. [method/model] What might be your proposal/ procedure to solve the problem? Why you choose the models/methods to deal with the problem? State your rationales.

由於這個 project 只有美股，所以我們的主要目標是讓這個模型符合台股。因此，我們計畫先用股票交易網站下載台股歷年資料（每日），並用證交所的 API 取得新的即時資料，持續更新訓練資料庫。

之所以選擇這個方法，是因為股票市場隨時在變化，需要有新資料匯入以維持預測準確度。且原 project 的 data 就沒有台股，所以我們一定要自己下載。