大家好，我是資訊112的莊上緣，接下來由我來為各位介紹我們這組final project後半部的部分。首先，我們會遇到的技術層面上的挑戰可能有以下幾點:首先，選擇哪檔股票作為data set的來源依據很重要，像是知名的大公司如台積電、大立光、聯發科等等的公司股價波動，通常會和股市大盤有著密不可分的關係，選擇影響力較大的公司股票來train，所得到的機器學習預測模型會較為準確。接著是關於我們要使用何種機器學習演算法來做為訓練機器學習模型，這也是值得花心思去投入的地方。不同的機器學習演算法都有不同的特點，我們必須根據這次我們的final project主題來多方嘗試並選取相對適合的演算法。最後則是特徵值的提取，要對於data set內提取哪些數據與特徵並丟給機器學習演算法是很重要的，在座各位有修過我們系開的基於遊戲設計的機器學習，應該都有相關的體驗，提取的特徵值為何，絕對是能否訓練出準確度較高的機器學習模型的關鍵。

接下來，我們的data set來源主要是以下這兩個網站，yahoo finance與Taiwan stock exchange，我們會利用網路爬蟲技術去即時抓取使用者感興趣且想預測自己投資損益的那檔股票的許多資訊，並將他們規格統一化之後拿去訓練。

再來，我要為各位介紹的是幾個針對方才提出的問題的幾個可能可行的辦法，第一點就是我說的，要找出正確的特徵值，如股價變動斜率、機器學習演算法函式內的參數進行細部微調等等，比如KNN演算法裡面的test size這種參數，代表可以選取讀入的資料中百分之多少的資料下去訓練。第二點就是資料規格統一化，我們都知道，從不同網站上面用網路爬蟲爬下來的資料格式參差不齊，我們必須使用這學期學過的pandas dataframe等技術將其整理成格式相同的資料型態，才可以讓程式去讀入並運算，最後這個則是我認為最花時間的，必須找到合適的機器學習演算法，在這學期我修的系上的機器學習課程中，有許多同學為了又訓練出相對正確的模型，嘗試了不下四到五種演算法，比如KNN decision tree random forest等等，我認為這部分會花掉我們大量的時間。

接下來，我們的預估計畫流程其實很簡單，取得資料、將其標準化、測試演算法並訓練、debug，若是成功了則輸出資料，若是沒有成功則繼續上述的循環。

最後，我想跟各位報告一下我們的各個步驟的預計天數，我們會花3天在資料處理，包含網路爬蟲以及資料標準化，接著花一周在機器學習演算法測試與訓練，最後 debug與修正的時間可能會花三天，我們每個步驟都會預留一到兩天的緩衝時間，因此，整個final project總共會花我們大概兩周到2.5周的時間來完成。

以上是我們這組的報告，謝謝各位。