**透過期貨交易資訊預期股票價格**

*y：股票價格*

*X : 期貨價格、期貨交易量、未平倉量、新聞資訊(Option，需統一字數)*

• Abstract (w/ keywords) 摘要 (關鍵字詞)

• Introduction (簡介)

• Motivation (研究動機)

對金融領域而言，準確預測金融商品未來價格是很重要的事，若能準確預測價格就可以獲得極大的報酬，傳統而言會從基本面或技術面去分析價格走勢，現在利用機器學習的方式提高猜測的精準度

• Purpose (研究目的)

為從股市中獲利

• Problem Formulation (問題描述)

• Proposed Methods (解決方法)

機器學習(random forest、cat boost、xg boost、CNN)

• Expected Results & Impact (預期結果與效益)

期貨價格跟股票價格呈現正相關，且領先股票價格，利用期貨價格去預測股票未來的走勢

• References (參考資料)

### 2021/11/20 分工 11/26 資料處理

* 找資料:5個數據feature(開高低收、價格變化量、交易量....)

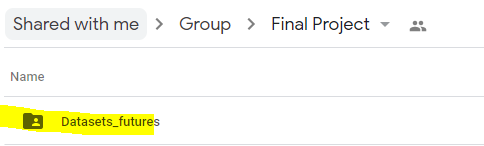
資料處理1周:含盤後 (**20年**)、空值(**補前一筆資料**)、

資料期間 : 2000/1/1 - 2021/9/30

資料範例 : 如下方截圖



資料放置位子 :



* Yahoo Finance for historical futures data (<https://finance.yahoo.com/>)
* x：期貨(找多個替換，VIX, 費半(莉), S&P500(Nick), S&P500-Mini(Nick), NASDAQ, Dow(達), 滬深300, 歐洲斯托克50(茂))

y：股價指數(S&P 500)(Nick)

預處理1周: 資料處裡(以美國s&p期貨時間為準)、數值比例縮放、權重/標準化(模型跑完才知道)

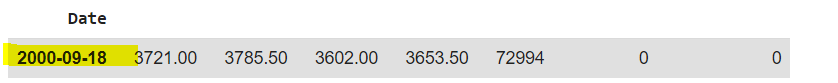
* X1:交易量：標準化 **volume\_std**
* X2:價格變動率 **ROC** = [ ( 當日收盤價 －日前收盤價 ) ／ 日前收盤價 ] \* 100%
* X3:收盤價 (先不考慮標準化期貨)：原本**close** / 標準化**close\_std** / 各自除以自己點數的倍數**close\_m**

<https://www.yuantafutures.com.tw/contract/%E8%8A%9D%E5%95%86%E6%89%80CME/%E6%9C%9F%E8%B2%A8>



=> 除20每分

時間格式：yyyy-mm-dd



VIX期貨忽略成交量

* 模型 (20年: train 15年/5年)

建模型和TRAIN約2周/、label可設1/0或機率

x由少加進去

1)Deep Learning(Lstm)

2)ML(XG boost)

Data Group1 (2002-2021) :

S&P500期貨+ 道瓊期貨 + NASDAQ期貨 + VIX指數

Data Group2 (2012-2021):

Eurostoxx50+ S&P500 futures+ 道瓊期貨+NASDAQ期貨+ VIX指數+滬深300期貨+恒生期貨

(5min.) 研究目的、預測

資料研究方式

(各5min.) 使用的模型(大概介紹模型特性)