**Multivariate Time Series Prediction for Stock Market Data**

資科一 資科一 資科二

曹昱維 鄭詠儒 謝政彥

110753201 110753126 109753207

1. **Introduction**

　　時間序列資料是由按照時間發生先後順序進行排列的數據點序列，以股票市場資料為例，從2000年至2020年為止，股票A的每日收盤股價就是一種時間序列資料。我們希望透過單一變數的時間序列資料(如股票A的每日收盤價)以及多變數的時間序列資料(如所有能源類股的每日收盤價)來訓練各種模型，來比較單變數與多變數對於預測效果的影響，也同時比較相同類型的變數下各種模型的預測效果。

　　股票市場以波動性、動態性和非線性著稱。由於政治、全球經濟狀況、突發事件、公司財務業績等多重（宏觀和微觀）因素，準確預測股價極具挑戰性。但是，所有這一切也意味著有大量數據可供尋找模式。因此，金融分析師、研究人員和數據科學家不斷探索分析技術來檢測股市趨勢。股票分析基本上可利用基本面分析與技術分析。

ARIMA (p, d, q)模型被廣泛使用在時間序列分析，在了解ARIMA模型之前可以了解相關的模型，(1)自我迴歸模型(AR Model)，它用前期的資料來預測本期的資料，越接近本期的資料，對預測結果的影響力就越大，設定一筆資料會與他過去p期的資料相關，(2)移動平均模型(MA Model)，方法在於本期的隨機誤差會與過去產生的隨機誤差有關，接著設定要計算q期移動平均，這些模型的缺點是只能處理穩定的資料，時間序列資料是否為定態資料，會影響後續預測的結果，因此，ARIMA模型改進了以上缺點，使用ADF-test計算差分次數(q)將時間序列資料處理成定態資料。

LSTM模型是近期時間序列資料常見的深度學習模型，屬於RNN (Recurrent Neural Network) 的一種，適合在輸入特徵空間中提取模式，其中輸入數據跨越長序列，可以從多個輸入變量的問題進行建立many to many 或 many to one 模型，在建模問題方面提供了很大的靈活性，包括可以很好地控制時間序列的幾個參數。

本專案嘗試針對S&P 500中的一檔能源類股APA Corporation股價，利用深度學習RNN模型LSTM以及傳統的時間序列模型ARIMA進行預測與比較。

1. **Related Work**

在早期有發展預測模型用來預測與分析股價，像是ANN模型[1]，近年來更是有相對穩定的ARIMA模型[2, 3]用來預測金融相關序列資料，而深度學習模型更是盛行，深度學習模型的效能仰賴參數的設定，以LSTM模型為例重要的參數有activation function (sigmoid, tanh, softmax等等)、optimizer (Adam, Adadelta, RMSprop 等等)、batch size、epoch數量以及hidden layers數量等等[4]。

ARIMA模型處理數據的非平穩性收集和建模，以類似的方式，作為基於深度學習的算法的代表 – LSTM模型是儲存和訓練於較長時間內給定的數據特徵，因此[5]對ARIMA和LSTM做了比較，提供了我們比較依據與方法。

1. **Method**
   1. **Dataset**

數據集為S&P 500其中的APA股價數據，為一檔美國能源類股，APA Corporation是一家是從事油氣勘探的美國公司，我們從Kaggle[6]上抓取S&P 500股價，並將其APA的股價抓出來，時間序列從2010年一月至2021年十一月。

* 1. **ARIMA (Autoregressive Integrated Moving Average)**
  2. **LSTM** (**Long Short-Term Memory)**

1. **Results**
2. **Discussion**
3. **Reference**

**Related Work**

1. G.S. Atsalakis and P.V. Kimon, “Forecasting stock market short-term trends using a neuro-fuzzy methodology”, Expert Systems with Applications, vol. 36, no. 7, pp.10696–10707, 2009.
2. Debadrita Banerjee, “Forecasting of Indian Stock Market using Time-series ARIMA Model”, ICBIM 2014
3. Ayodele A. Adebiyi., Aderemi O. Adewumi and Charles K. Ayo, “Stock Price Prediction Using the ARIMA Model”, UKSim-AMSS 2014
4. Anita Yadava, C K Jhaa and Aditi Sharanb, “Optimizing LSTM for time series prediction in Indian stock market”, ICCIDS 2019
5. Sima Siami-Namini, Neda Tavakoli and Akbar Siami Namin, “A Comparison of ARIMA and LSTM in Forecasting Time Series”, IEEE 2018
6. S&P 500 stocks price with financial statement <https://www.kaggle.com/hanseopark/sp-500-stocks-value-with-financial-statement>
7. Predicting Stock Prices Using Machine Learning <https://neptune.ai/blog/predicting-stock-prices-using-machine-learning>
8. Prediction of price for ML with finance stats <https://www.kaggle.com/hanseopark/prediction-of-price-for-ml-with-finance-stats/data>
9. Time-Series Forecasting: Predicting Stock Prices Using An LSTM Model <https://towardsdatascience.com/lstm-time-series-forecasting-predicting-stock-prices-using-an-lstm-model-6223e9644a2f>
10. Berkshire Hathaway - Stock Time Series Analysis <https://www.kaggle.com/kalilurrahman/berkshire-hathaway-stock-time-series-analysis>
11. A Multivariate Time Series Modeling and Forecasting Guide with Python Machine Learning Client for SAP HANA <https://blogs.sap.com/2021/05/06/a-multivariate-time-series-modeling-and-forecasting-guide-with-python-machine-learning-client-for-sap-hana/>