

國立交通大學管理學院 資訊管理研究所

第二組

資料探勘研究與實務期末書面報告

-以RNN、LSTM、GRU預測IBM股價

指導教授：劉敦仁 教授

組員：0753426李柏均 0753427黃晨彧

0753428高碩興 0753437劉時嘉

一、專題題目

利用三種不同的深度學習方法(RNN、LSTM、GRU)來預測IBM股價走向並比較。

二、概念發想、動機

一般投資者在投資各種金融商品時，往往都是大賠小賺，因投資者都是聽信消息面及一時的衝動，或未作詳細的評估分析就輕易投資，若要單獨分析各種經濟資訊不僅需要蒐集繁多的相關資訊，不但非常費時且需要豐富的專業財經知識，而其預估未來的發展趨勢亦不一定準確，且通常經濟指標是落後市場反應的指標。

即投資者必須具備豐富的專業財經知識，同時要蒐集相關的資訊，才能做出分析，但這對一般投資者而言有其困難性存在。

近年來，科技不斷的發展，其中尤以資訊科技的演進最令人目不暇給，不但硬體、晶片、電腦運算的速度越來越快，人工智慧的理論在過去二三十年的基礎搭著硬體進步的浪潮慢慢被實做出來，甚至也不斷推陳出新，資訊科技運用在現實生活中，已是越來越普遍。而金融、財經的數據也越來越能搭配資訊科技的輔助，就電腦運用在股票預測方面，近來利用類神經網路來作預測可說是相當的熱門。除此之外，也由於電腦的輔助，許多統計上的分析也可在一瞬間完成，不必利用人工方式來加減乘除。因此，利用資訊科技來作投資理財的輔助工具，應是未來的趨勢。

而IBM對我們資管的同學來說可是息息相關，我們畢業也大有可能接觸此企業，所以我們決定拿IBM的股市來做此次的專題分析。

三、資料集敘述

1.資料名稱：Kaggle Data Set 中 Stock Time Series of IBM

2.說明：

所有文件都包含以下：日期 - 格式為(yy-mm-dd)

市場開盤時的開盤價及收盤價（這是紐約證券交易所的數據，因此全部為美元）

最高點 - 當天達到的最高價格

最低點 - 當天收到的最低價格

成交量 - 交易的股票數量

名稱 - 股票的股票名稱

3.資料來源：

<https://www.kaggle.com/szrlee/stock-time-series-20050101-to-20171231>

4.使用資料集檔案名稱：資料集 IBM_2006-01-01_to_2018-01-01.csv

Date	Open	High	Low	Close	Volume	Name
2006/1/3	82.45	82.55	80.81	82.06	11715200	IBM
2006/1/4	82.2	82.5	81.33	81.95	9840600	IBM
2006/1/5	81.4	82.9	81	82.5	7213500	IBM
2006/1/6	83.95	85.03	83.41	84.95	8197400	IBM
2006/1/9	84.1	84.25	83.38	83.73	6858200	IBM
2006/1/10	83.15	84.12	83.12	84.07	5701000	IBM
2006/1/11	84.37	84.81	83.4	84.17	5776500	IBM
2006/1/12	83.82	83.96	83.4	83.57	4926500	IBM
2006/1/13	83	83.45	82.5	83.17	6921700	IBM
2006/1/17	82.8	83.16	82.54	83	8761700	IBM
2006/1/18	84	84.7	83.52	84.46	11032800	IBM
2006/1/19	84.14	84.39	83.02	83.09	6484000	IBM
2006/1/20	83.04	83.05	81.25	81.36	8614500	IBM
2006/1/23	81.33	81.92	80.92	81.41	6114100	IBM
2006/1/24	81.39	82.15	80.8	80.85	6069000	IBM
2006/1/25	81.05	81.62	80.61	80.91	6374300	IBM
2006/1/26	81.5	81.65	80.59	80.72	7810200	IBM
2006/1/27	80.75	81.77	80.75	81.02	6103400	IBM

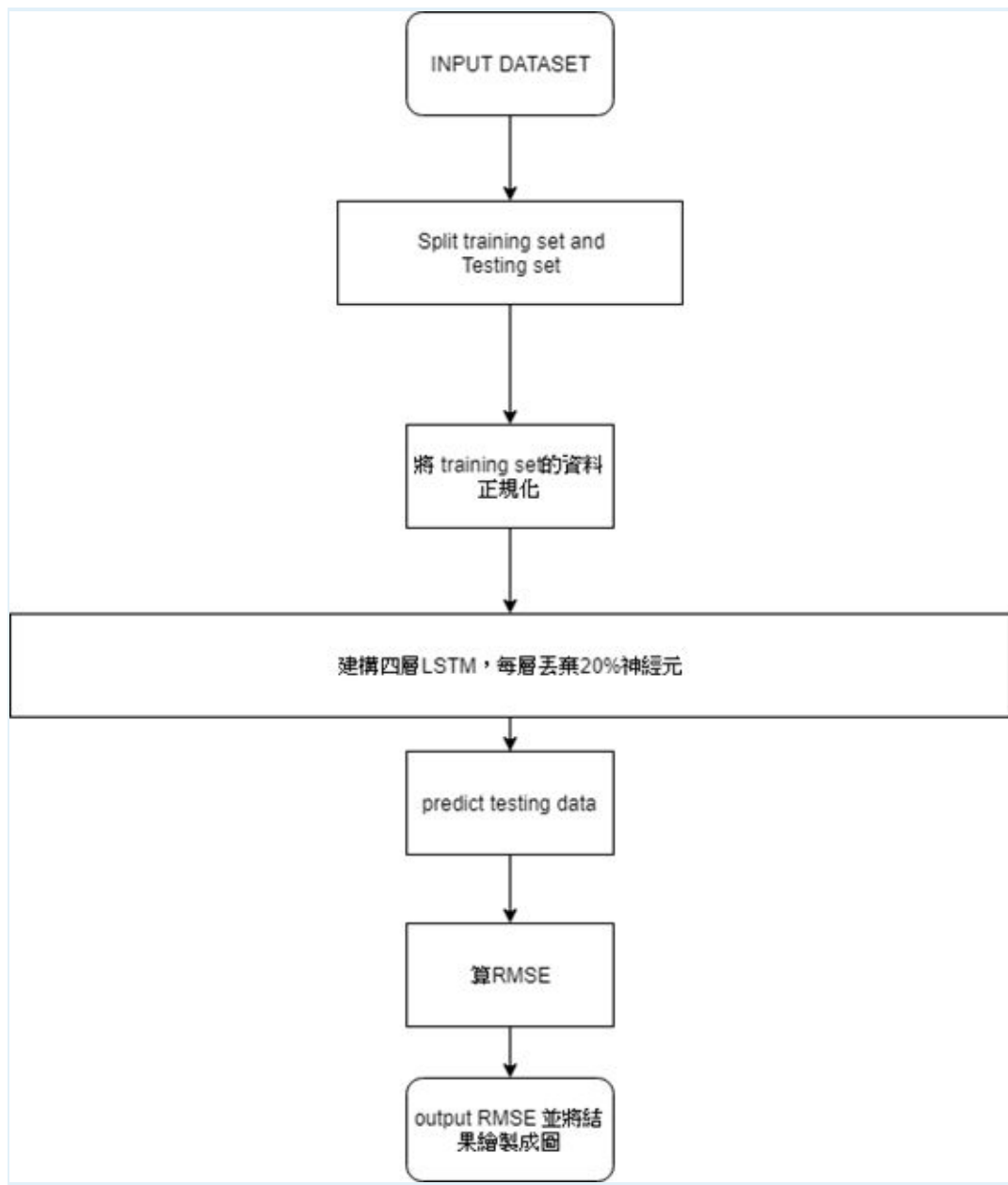
四、分析工具

1.基於python的深度學習庫：Keras

2.機器學習框架：sklearn

五、流程與實作結果

流程圖：(以LSTM為例)



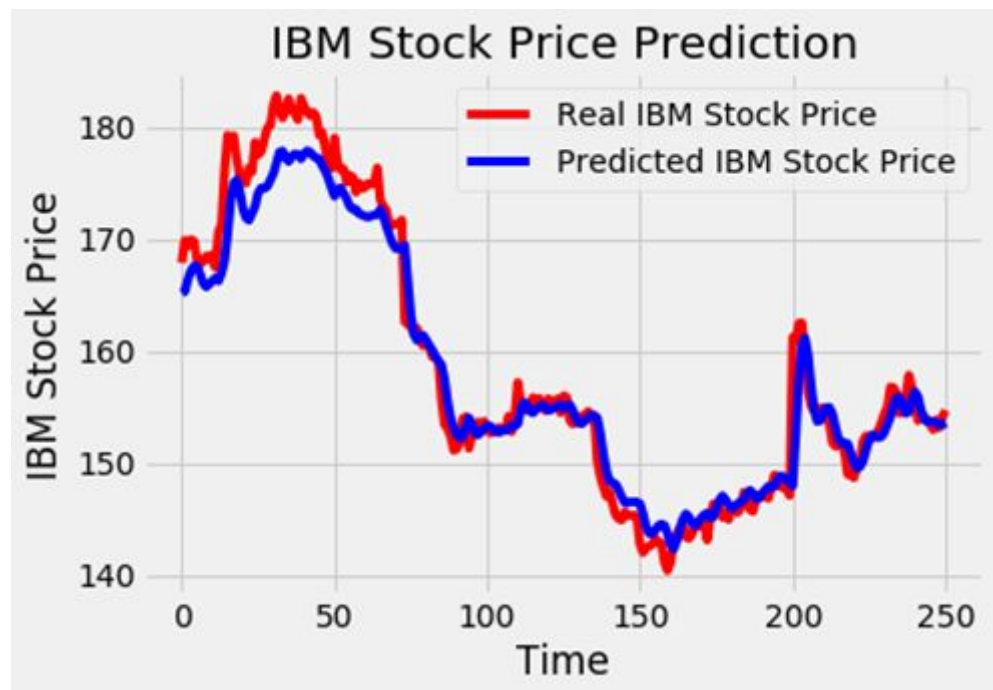
評估方法：

本次實驗皆採用Keras來搭建深度學習模型，分別利用RNN、LSTM、GRU等深度學習方法來預測。並以均方根誤差(RMSE)來當作評估標準。

實作結果：

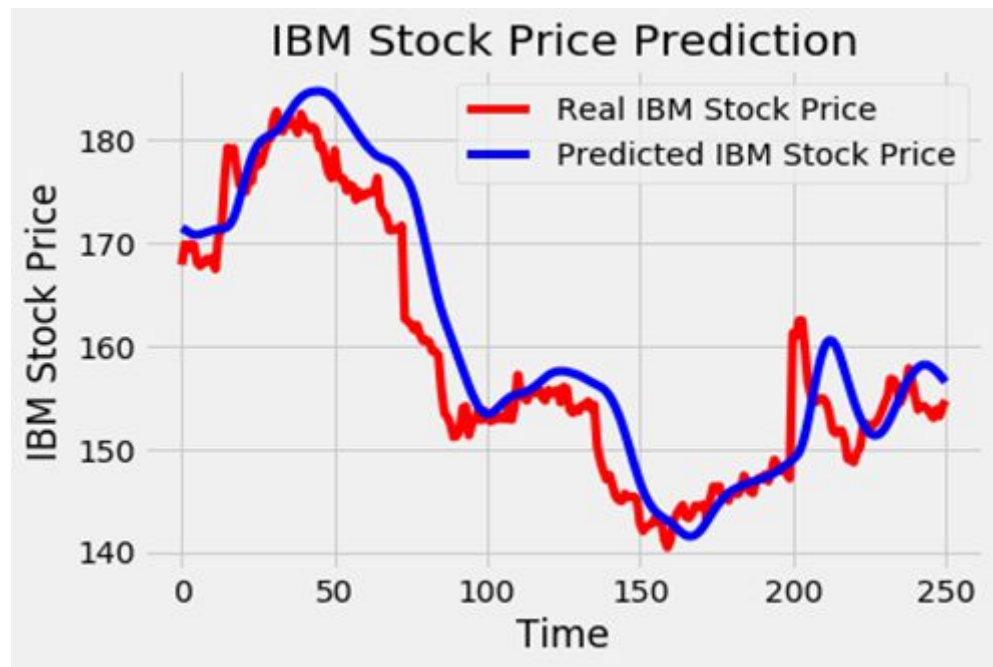
(1)LSTM:

RMSE:2.5459669921977524.



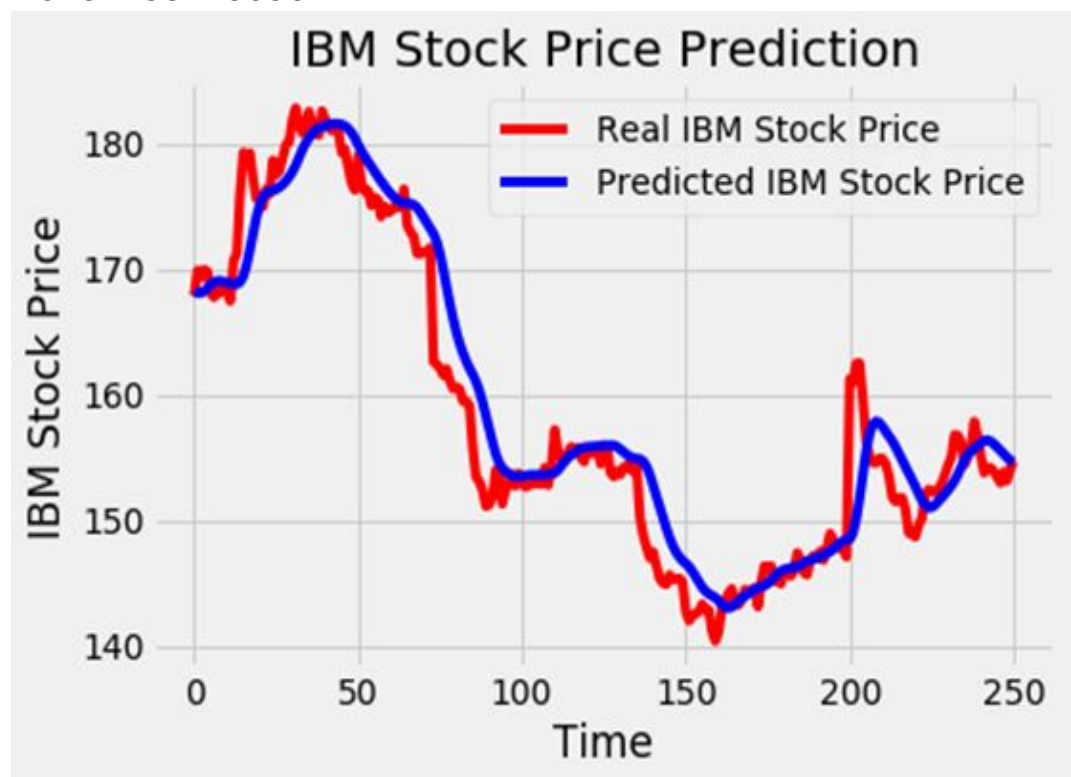
(2)RNN:

RMSE: 4.72949126626088.



(3)GRU:

RMSE: 3.3223451158225696.



七、分析結果與討論

從上面結果圖可以看出，我們所預測的走勢與實際情況相近，其中又以LSTM最佳，RNN最差，我們推測原因是因為RNN只考慮最近的狀態，而LSTM結構更為複雜，會將過去的輸出和當前的輸入concatenate到一起。在現實中的股價波動往往並不單純只與前一天的股價相關，而是會與歷史中的股價變化有所關連。