投稿類別：資訊類

篇名：

rAIn - 用AI預測水量

作者：

游子萱。新北市南山高中。普通科一年仁班

李建霖。新北市南山高中。普通科一年仁班

游閔皓。新北市南山高中。普通科一年仁班

指導老師：

洪慧美老師

張智皓老師

1. 前言
   1. 研究動機

颱風是台灣常見的天災之一，一旦颱風生成並登陸，依據不同強度將會對當地造成傷害，輕則狂風豪雨，重則山崩泥流，淹沒房舍，居民流離失所。對於莫測的大自然，人類既敬畏又想征服掌控，因此開始開發能夠預測颱風路徑的程式。出於對這樣一個預測並能使人民提前防範颱風，替人類社會提供助益的程式感到好奇，並學會利用程式能力貢獻社會的精神，故撰此文紀錄研究與實作此程式的過程與結果。

* 1. 研究目的

本篇研究論文，希望藉由蒐集資料，實作雨量預測的程式，進行短時間的雨量預測，以進行災害評估，並藉由其對於雨量規模的預測，也可以提早投注資源對易造成洪患的地區進行加強，以求災害所造成的影響最少。

1. 文獻探討
   1. AI作為天氣預測的應用現有實例（為何用AI）

日本「降雨雷達」APP，藉人工智慧綜合分析多元氣象資訊，預測暴雨發生機率，以提前向民眾預警水災來臨，降低來不及逃跑的風險。另外曾開發AlphaGo的Google子公司Deepmind近年開發出來的天氣預測也是媒體的焦點，其和英國氣象局合作得到資料後，更加強了天氣預報的準確度，Deepmind的氣象預測模型是一種短期預測的模型，可以預測90分鐘後的天氣狀況，在小區域下，近乎及時預報的特性，幾乎可以匹敵當代使用物理特性為基礎的氣象預測方式，並且有超越其準確度的趨勢。（電腦王，2021年11月04日）

機器學習之所以被作為天氣預測一種方式，因為其可以整理大量資料，並且對資料進行歸納整理，透過神經元的權重，將一段天氣現象的特徵進行捕捉，使得在短時間內產生的氣象資料可以得到快速的預測，機器學習通過學習時間越長得到的結果也就越準確，也就是說在現今資料雲端化的趨勢下，機器學習將有更多資料可以優化其預測的結果。

* 1. 影響降水量的因素

**影響台灣氣候變化的主要因素為季風與大型天氣現象，類似聖嬰現象、北極震盪等。**（中央氣象局全球資訊網），從上述說明可以發現台灣的天氣主要受到氣溫、壓力以及季風的變化影響。

1. 研究方法
   1. 程式實作（AI學習流程圖）

圖一：機器學習流程圖（來源：研究者自製）

* 1. 研究器材

1. selenium

selenium是網頁爬蟲使用的自動化網頁工具，主要可以用於自動化網頁測試與網頁抓取的功能，並且由於其使用的是自動控制真實的瀏覽器，因此不必解析網頁所需的cookie、request等資料，使得抓取資料的過程被簡化，不過相對來說也使得速度較慢。

1. pandas、numpy

此論文中使用pandas中的Dataframe作為解析與更改csv檔的工具，並且使用numpy中的功能製造符合機器學習模型需求的array。要使用python進行資料分析及繪圖時，最為人稱道的就是pandas模組了!（鄧文淵， 2018）

1. Keras、SciKit-learn

本小論文使用Keras作為機器學習的基礎，並使用SciKit-learn作為機器學習的資料歸一方式。因SciKin-learn有專門進行資料前置作業的函式庫，並且可以通用於numpy的array和python原版的list，因此方便我們進行資料後續的處理，而Keras作為較為簡易且視覺化的機器學習軟體，可以用模組化的方式拼接機器學習模型，並有預先寫好的機器學習節點，使得建置上比較不用擔心機器學習模型有錯誤的情形。

1. 觀測資料查詢系統CODiS

此系統可以查詢至1995年的資料，並且可以下載csv之機器學習或資料分析等通用檔案，方便進行資料的彙整。

1. matplotlib 繪圖
2. glob 批次讀取資料夾內的資料
   1. 研究流程

圖二：研究流程圖

1. 研究分析與結果
   1. 程式碼展示
2. 資料下載

CODiS為線上的資料庫，雖然其有提供csv的下載連結，不過它是利用網頁的程式碼將呈現出來的Html的格式進行轉換，若直接使用網頁爬蟲的request或curl會產生需額外撰寫解析Html，因此在討論後我們決定選用selenium加上網頁瀏覽器，使得電腦操控瀏覽器進行下載。





圖三：資料下載程式

1. 將下載後資料轉換為.csv檔案

.csv檔為通用的檔案格式，以逗號分隔的文字形式儲存資料，因此可以以程式的方式進行編輯，並且使用pandas讀取成Dataframe之後便可以改變內容並重新存為.csv，引此我們將下載後的.csv檔轉換後合併成新的csv檔案方便後續機器學習。



圖四：將.csv合併並傳存為.csv

1. 切分資料

此次論文所要建成的模型為LSTM模型，因此會多一個時間維度（time steps）的資料，切分資料時就必須注意其時間切分是否正確，這次建置的是多對一的模型，因此y都為一個二維陣列，第一維為總共的資料數，第二維則為一個相對應的資料，x則為三維陣列，分別為「總量、時間維度、7種不同的資料」，相比於卷積神經網路多了一個維度的資料。

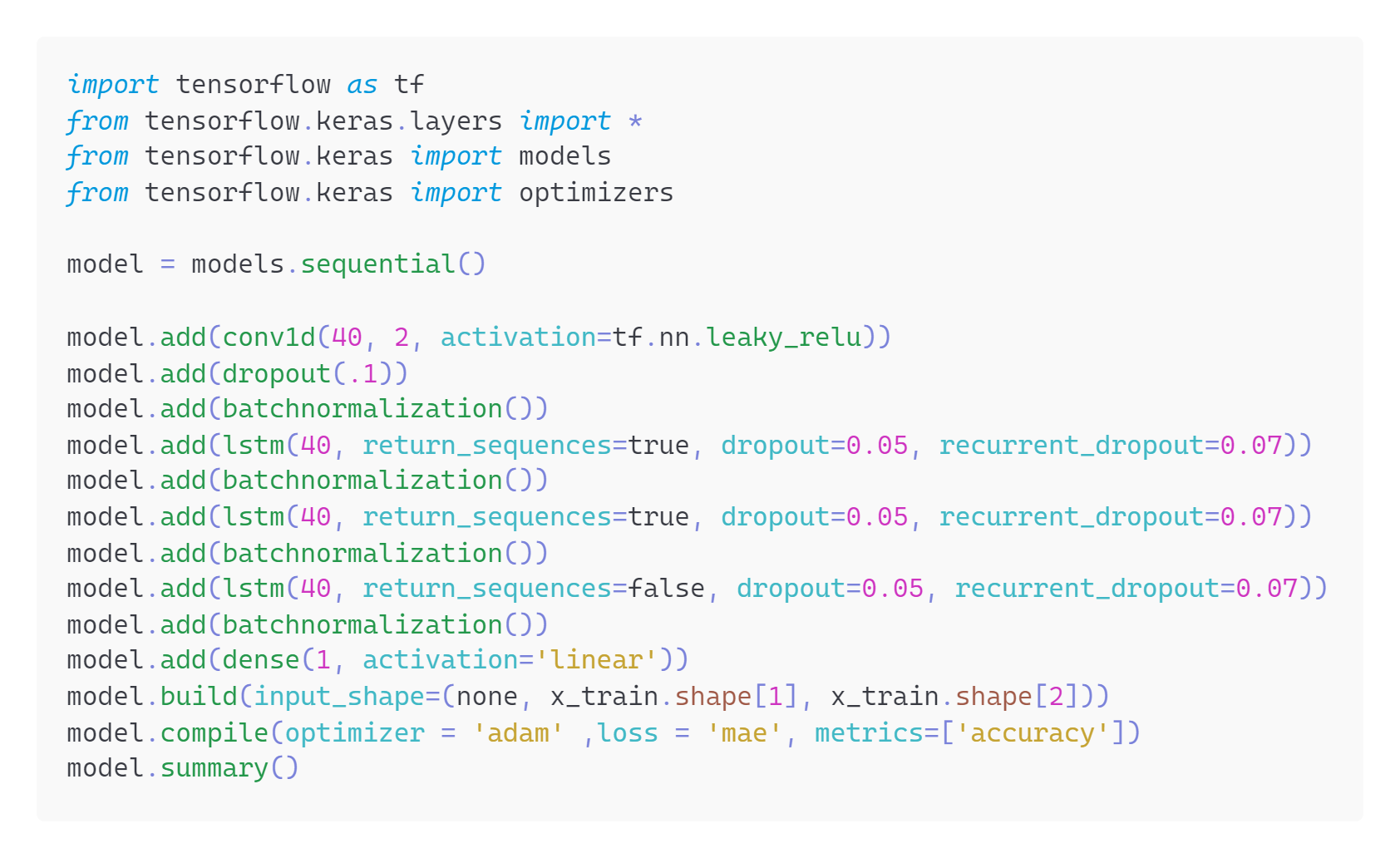
從csv轉換過來的Dataframe由於有在降水量有T的資料表示微量資料，由於其為一個字串，因此我們直接將其取0，以方便後續的機器學習。

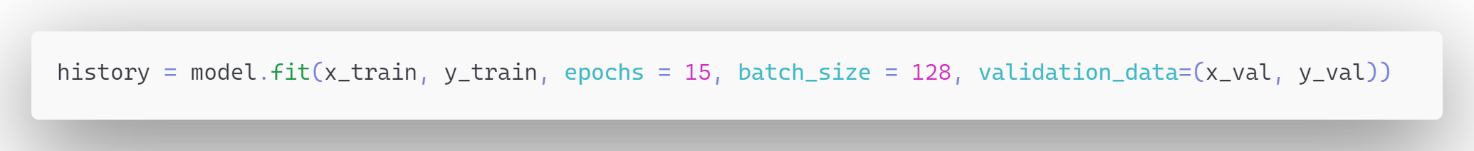


圖五：切分資料程式

1. 模型建置

在處理影像的範疇，幾乎都可以使用CNN。（斎藤康毅, 2017）使用keras的sequential作為模型基礎，先以一個CNN減少資料的輸入LSTM的大小，後在每個資料後添加dropout防止模型過擬合，以Leaky\_ReLU取代ReLU作為激活函數防止節點死亡，最後以線性函數為激發函數的全連接層輸出預測的資料。We find the performance of normal leaky ReLU （a = 100） is similar to that of ReLU, but very leaky ReLU with larger a = 5.5 is much better.（Bing Xu, 2015）。

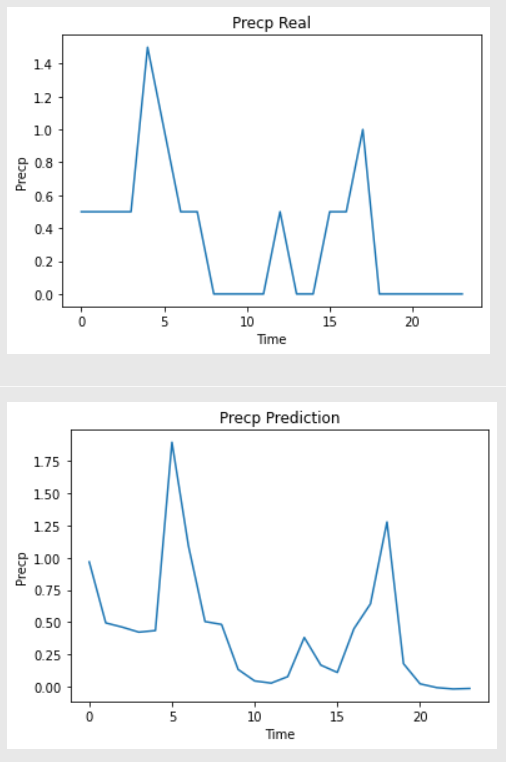
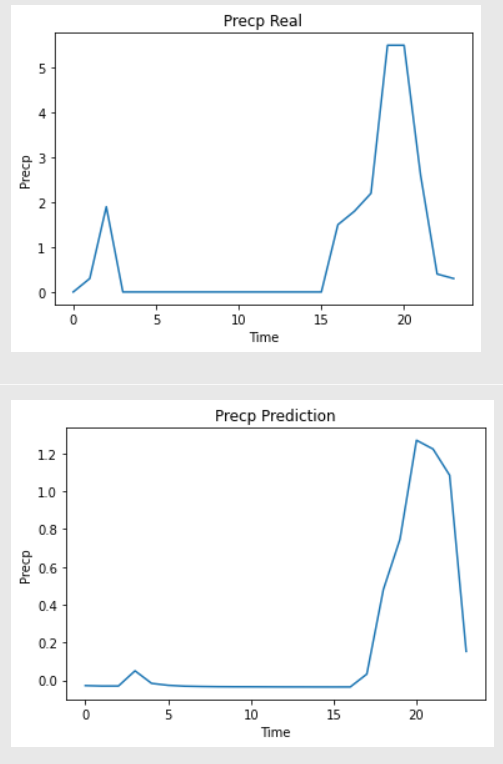
圖六：模型建置



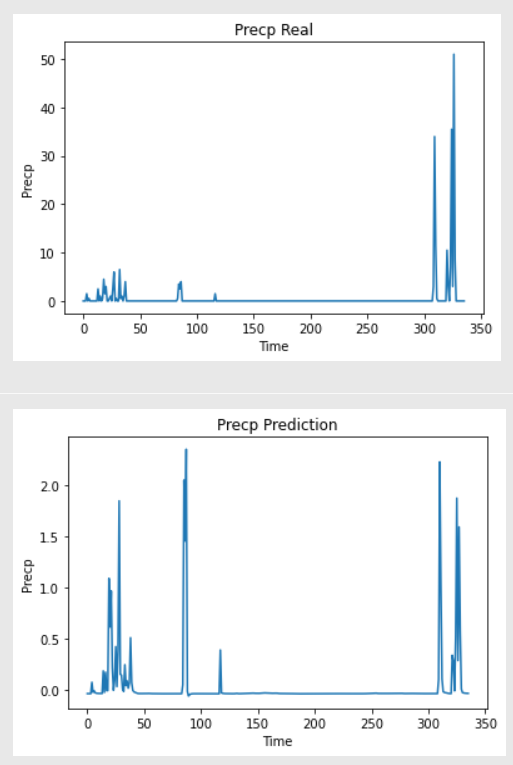
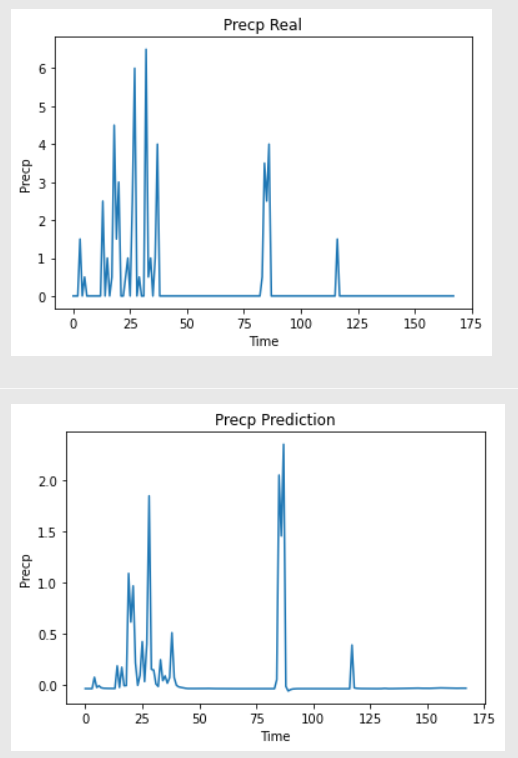
圖七：模型訓練

* 1. 預測與實際狀況比較

短時間的預測中，如圖八，一天資料的預測，可以發現機器學習模型的預測的曲線變化與實際的曲線變化相近，在圖九，一周與兩周的預測，也可發現模型計算出下雨的時刻，可見機器學習以學習到短期的資料預測，同時也可以發現資料的尺度在機器學習的模型上有不一樣的情形，我們推測機器學習由於是由大量資料學習取得的平均變化，因此導致對於資料的變化預測變化較小。

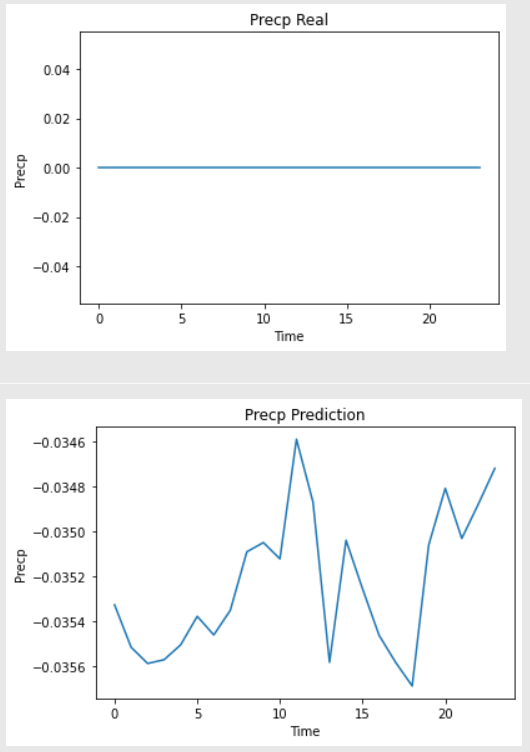
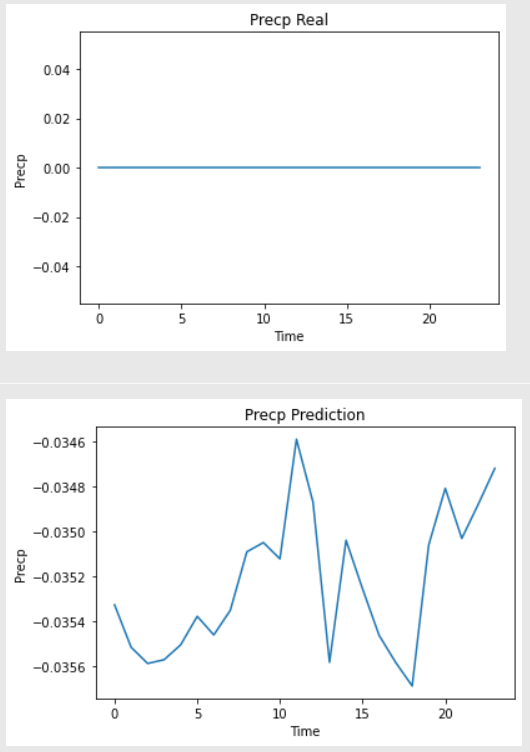


圖八：預測一天的降雨量（上圖為實際雨量，下圖為預測雨量）



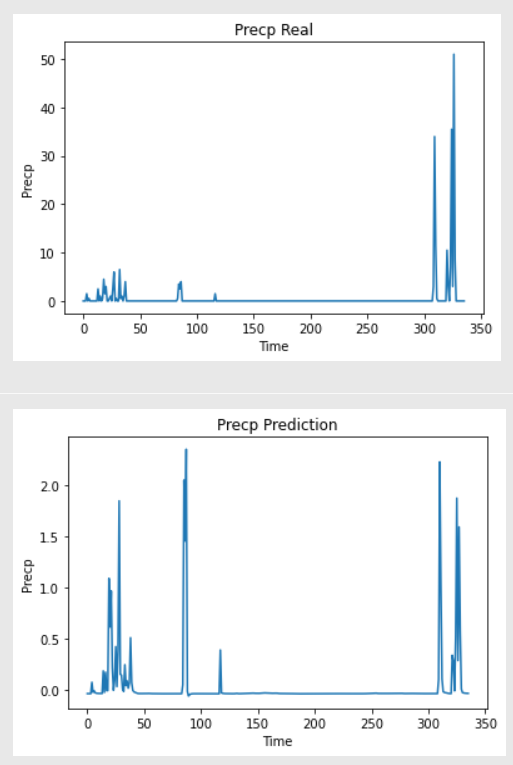
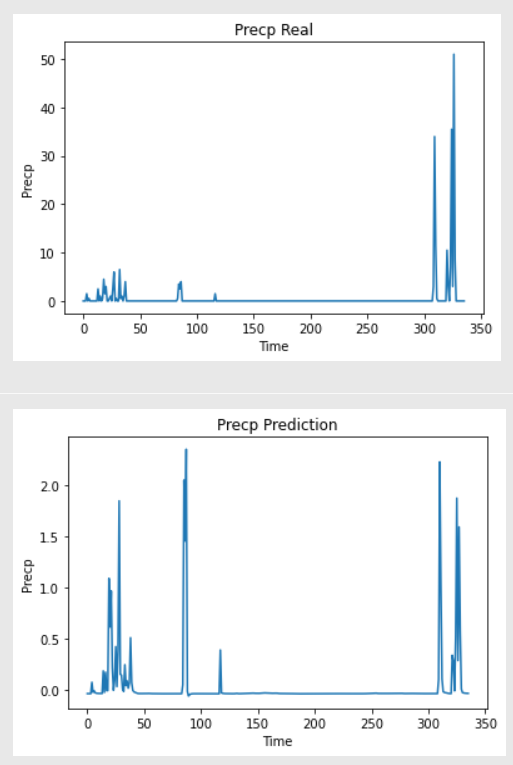
圖九：左圖為一週的預測，右圖為兩週的預測  
（上圖為實際雨量，下圖為預測雨量）

對於長時間不變化的資料，如圖十，一天未下雨的情形，機器學習的曲線雖然不相近，但預測的數值非常小，預測值近乎為0。



圖十：左側為實際雨量，右側為預測雨量

最後模型對於時間尺度為月份或年分即無法準確預測，如圖十一，推測是資料量變多，而模型預測的雨量沒有明顯的變化而產生的問題



圖十一：一年雨量預測（左圖為實際雨量，右圖為預測雨量）

* 1. 遇到的瓶頸

1. 資料選取範圍

在剛開始執行此篇論文時，我們選擇了NullSchool的氣象圖資料，後來發現資料涵蓋範圍太大，全球的資料不僅導致資料雜亂，更是導致資料會產生過多的運算，並且選用台灣地區的資料進行放大後，資料大小也過大，不方便使用一般家用主機進行機器學習，同時也增加資料整合上的困難。

因此我們之後選擇台灣一個氣象站的資料，初始選在平地的氣象站，不過平地氣象站的在每日的氣象資料卻不完善，因此最後我們選擇位於山地的氣象站作為研究的資料來源。

1. 分割帶有時間序列的資料

由於RNN的演算法有著時間序列的問題，因此建置資料時必須包含具有時間的資料，且以RNN設計的機器學習模型適合短時間的預測，故時間的選取範圍與切分方式最後選擇使用python內建的list並透過scikit-learn轉譯資料為0~1的範圍，選取3天的資料為範圍，之後再透過numpy統一轉譯型別為浮點數（float）的型別

1. 資料的時間細膩度

剛開始研究時選用的最小時間尺度為一天，發現模型無法妥善將資料整合，預測結果也不盡理想，因此改以半天，卻也發現資料在分析時若只使用半天的資料容易導致機器無法準確的預測資料當下的情形，因此最後我們選擇資料中最小的尺度，小時為單位，並進行小時間尺度的氣象預測，以確保短時間資料的密集度與關聯性。

1. 研究結論與建議
   1. AI預測的準確度

AI預測後的結果雖然準確度不高，但有注意到天氣變化的特徵，例如每過一段時間的雨季AI就有透過給予資料的變化，使預測產生明顯的起伏，並且在越短周期的資料，雖然預測出來的數值相差值偏高，但其與原本曲線的相似度越高，可見其有藉由資料預測到變化，不過此模型對於資料和預測的時間有差距，使得短時間的預測也存在誤差，因此模型可能可做為預測降雨機率的輔助工具。

* 1. 模型的改進方向

此模型目前尚有未擬合完全的問題，在測試時，測試與驗證集的損失函數無法持續的穩定下降，未來將可以對模型進行更多測試。

* 1. 天氣預測中AI的貢獻與未來發展

機器學習現已有運用在短期的氣象預測中，未來機器學習在對於短期防災的效果將隨著機器學習的發展繼續進步，正如Deepmind的氣象預測模型，我們在可見的未來將有望見到機器學習為人類無法處理的大量資料分析提供助力

1. 參考文獻

cnBeta（2021年11月04日）。DeepMind又贏了！AI天氣象預報準確度遠超氣象台：1張GPU1秒預測1小時天氣。電腦王。<https://www.techbang.com/posts/90385-deepmind-wins-again-ai-weather-forecast-hangs-weather-station>。

吳嘉芳（譯）（2017）。**Deep Learning｜用Python進行深度學習的基礎理論實作** （原作者：斎藤康毅）。歐萊禮出版社。

鄧文淵（2018）。**python大數據訓練班。**碁峰資訊。

Bing Xu, Naiyan Wang, Tianqi Chen, Mu Li（27 Nov 2015）。Empirical Evaluation of Rectified Activations in Convolutional Network。<https://arxiv.org/pdf/1505.00853.pdf>。

中央氣象局全球資訊網（無日期）**。**台灣氣候變異的主要影響因素**。**<https://www.cwb.gov.tw/V8/C/C/Taiwan/taiwan_5.html>**。**