**林杰 (Hubert)**

* **應徵職位**

全年實習（每周3至4天佳）

機器學習、資料科學相關為主，後端、軟工相關亦佳

* **簡介**

我是一個自我要求很高、勇於挑戰困難的學生，開始學寫程式至今僅約兩年，已有多項專案成果，也已能與研究所學生一同修課、合作、受到教授肯定，接近完成畢業門檻。目前學習較偏重機器學習、資料科學與資安。

而我在課餘則較重視前後端技術，也會關注Kaggle並參與比賽，自學Python、JS、前後端框架、Chrome extension、iOS app、Arduino …等，並關注尖端技術發展，除了興趣之外，我也認為這些知識與理念能提供不一樣的新觀點與想法。

我對教授說過的一句話印象很深刻：「Know your data.」，我認為了解自己所學知識的實際應用、要面對的問題、處理真實的資料，是自我提升非常重要的環節。

此次申請實習希望能應用所學的知識與技術，並實際解決問題、做出貢獻。

* **個人特質**

擅長學習新知識與技術

能正視失敗並持續改進、挑戰

擅於解決問題與思考解決方案

* **生涯規劃**

**Short Term**：

短期之內並沒有繼續進修的計畫，將把握大學剩餘一年努力學習新技術，接著直接進入職場累積經驗、處理真實問題、理解自身不足，隨後才會再次考量是否進修。

**Long Term**：

希望能在不同的經驗中尋找自己的原點、確定想要花費一生鑽研的領域、找到一群一起奮鬥的團隊。



**清華大學資工系大三C/C++/Python/Java/JS**

**專長：機器學習、前端**

**學習中：資安、DB、後端**

**Github：** [**hubert0527**](https://github.com/hubert0527)

**英文能力：TOEIC 870分**

手機：

0963243221（簡訊佳）

E-mail：

[hubert052702@gmail.com](mailto:hubert052702@gmail.com)

個人網站（by myself）：

<hubert0527.github.io>

線上履歷網址：<goo.gl/Bb7kne>

* **相關專案經驗**

**News Popularity （Large-Scale Machine Learning）**

相關連結：[Github](https://github.com/hubert0527/MachineLearning-NewsPopularity)、[Kaggle](https://inclass.kaggle.com/c/datalabcup-2016-news-popularity)

研究所課程比賽，根據新聞的HTML原始碼預測受歡迎度。

我負責進行preprocessing、feature尋找與抽取、model設計、training、tuning、testing等所有工作，並在該次競賽中得到第二名的成績。

我能證明自己對機器學習有正確的認識，有卓越的實務能力（獨自完成三人份程式且成果僅次於第一名小組），多次自行實作演算法（Boosting、Stacking、CrossValidation、Ensembling），並在private score的階段驗證 model 具有generalizability（public score: 0.572, private score: 0.579）。

**Image Caption （Deep Learning）**

相關連結：[Github](https://github.com/hubert0527/Image_caption)、[Kaggle](https://inclass.kaggle.com/c/datalabcup-image-caption)

研究所課程比賽，使用MS COCO image set解captioning的問題。

使用word-embedding、CNN、RNN作為基本框架，實作attention、beam search、curriculum learning。過程中我設計新的演算法，提升beam search精準度，讓整體效能額外提升10%。

在CIEDr-D評分下得分0.58（在2015 MS-COCO比賽中0.4以上便能躋身前40名），在這次比賽中得到了第5名。經反省後找出問題的癥結，但無奈於硬體限制無法實際驗證。

**Data Crawling （MLB game information crawling）**

相關連結：[Github](https://github.com/hubert0527/DataCrawler--MLB)

此小組專案的目標是MLB賽事的勝負預測，我負責進行 data parsing，並共同負擔model training的工作。

我共抓取橫跨103年、17萬場比賽、652項feature、總計389MB、經過清理的MLB比賽資訊，並根據需求實作feature擷取的API，大幅減輕後續training phase的時間壓力。

使用python的request與selenium進行一般網頁與JS render網頁資料的擷取，針對所有可能的例外狀況進行錯誤處理、清理、重新抓取 … 等。

**-- Thanks for reading my resume ! --**

**-- See more projects on Github ：** [**hubert0527**](https://github.com/hubert0527) **--**