**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、軟工相關亦可。

* **簡介**

目前就讀資工系大三，擅長於實作，至今絕大多數專案都有獨立重現的能力。因偏好修較嚴格的課程而使得GPA較低，CS平均GPA3.69，等同A-。

擅於學習是每個人都會自述的特質，但我有諸多實例可以表述自己**特別**擅長，並且也非常努力。其中之一是我在資工領域僅花費兩年半時間，便已近乎達成畢業門檻，並有自信比許多畢業生更有實務能力。

此外，在課餘也非常積極學習，會關注Kaggle並參與比賽，自學Python、JS、前後端、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。

我的學習範疇不僅限於資工，在小說、圍棋、音樂、棒球上也都有投注非常多時間精力，有許多傑出表現。

在實習中，我希望能獲得更多的「多人協作」、「處理實際問題」、「實作可靠、可信賴的程式」經驗，積極學習之外，也相信自己的能力能回饋公司、實際做出貢獻。

* **個人特質**

擅長學習新知識與技術

能正視失敗並持續改進

擅於解決問題與思考解決方案

* **生涯規劃**

短期之內並沒有繼續進修的計畫，將把握大學剩餘一年充實各方面的知識，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。

長期而言，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**

**目前專長：機器學習、前端**

**學習中：網路安全、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](file:///C:\Users\user\Desktop\hubert0527.github.io)

線上履歷網址：

goo.gl/YCgCBN

* **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，並有卓越的實務能力（獨自完成三人份程式且成果僅次於一個小組），使用python自行實作多個演算法（Boosting、Stacking、Cross Validation、Ensembling），在不斷失敗中持續思考改進方式，並持續驗證與修正自己的理解。

**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，根據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，但無奈於硬體限制暫時無法實際驗證。

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

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

「資料科學概論」課程的一項專案，取得資料是資料科學的重要基礎，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，並實作feature擷取的簡易API，大幅減輕後續training phase的時間壓力。

使用python的request與selenium分別進行一般頁面資料與JS render頁面資料的擷取，針對所有可能的例外狀況進行錯誤處理、清理，增加了對於資料擷取與清理的實務經驗。