

# AI project proposal

---

## 成員

- 朱哲毅
- 吳維誠
- 王力得

## Problem Statement and Task Definition.

貴為偉大大學的莘莘學子，想偶爾騎個 Youbike 溜下去清夜吃宵夜，但等待我們的常常是空無一車的 Youbike 租借站。因此，我們的 AI 想要做到的是能夠預測學校附近 YouBike 的狀態，像是有沒有車可以借，有沒有地方可以還，雖然已經有軟體可以看到當前剩餘的車輛，但沒辦法**預測**某個時間點的狀態，若是需要透過 YouBike 來轉乘其他交通工具會比較難預估時間。因此，我們的 AI 輸入會是(時間點,站點)，輸出則會是預測站點在此時間點的狀態，若是可以的話我們想要做到可以預測出**多久之後可以去借車**，相當的 social impact。

## Description of the challenges.

1. 官方網站沒有已經整理好的 dataset，但我們發現 YouBike 的網站會定時更新及時的站點狀態，包含可借車數和可停車數，因此我們需要**透過爬蟲來蒐集**並累積各時間點的站點狀態，為了整理出足夠可用的 dataset，所以最近要趕快寫好爬蟲程式。
2. 會影響該時段的站點狀態的因素有很多，有可能是前幾個小時的租借狀況，也有可能是前一天同時段的狀況，亦也有可能是前幾天的單車總租借數量。但是，把所有的東西一股腦的丟進去 model 做訓練，顯然不合適。因此，我們需要找出**關聯性可能較高的因素**來訓練我們的 model。

## Input/Output Behavior with Concrete Examples.

input(2022/4/15/20:53,"交通大學(研三舍)")

ouput(預估可借車輛:39,預估可停空位:21,預估 2 hr 13 min 後可以去借)

## Related works.

- <https://www.techinasia.com/sensetimes-ai-shared-bike-properly>
- <https://dl.acm.org/doi/10.1145/3447988?sid=SCITRUS>
- <https://medium.com/@kinsg96sun/bike-sharing-demand-prediction-model-and-monitoring-ae2f9a3118c7>

## Methodology

我們使用 **Random Forest Regression**(隨機森林演算法)來完成這次的專案，這個演算法是一種 **ensemble learning**，有點類似作業一的 **adaboost**，只是 **random forest** 是用 **bagging**，也就是每個模型是用隨機的訓練資料子集來 **train**，好處是他對於預測非線性相關的資料有較高的準確度；但在實作時要注意 **overfitting** 的情形，不能矯枉過正。

## Evaluation Metrics.

想要評估我們模型的精準度，可以將我們預測的結果和站點實際的狀態進行比較，數字越相近則代表我們的預測越準確。一開始我們可以把一部分的 **training data** 作為 **test** 來測試，之後則可以利用爬蟲每天更新的資料來驗證我們 **AI** 的準確度。最後可以利用圖表來 **visualize** 我們的結果。

## Baselines.

我們的最終目標是要預測出某站點還有多少可以借用的 **YouBike**(或可以歸還的空位)，我們可以先將問題簡化到只判斷某個站點有沒有辦法借用(或歸還)車子，只判斷 **YES/NO**，而不是實際的 **value**，確認這個方法可以用之後再把問題拓展成原本的樣子。然後一開始可以使用 **kaggle** 上國外共享單車的 **dataset** 來測試我們的 **model**，確定預測有準之後再使用真正的爬蟲資料。

# Workplan

---

## Schedule

1. 上網學習隨機森林演算法
2. 利用 sklearn 簡單實作一個模型
3. 利用 kaggle 上的 dataset 先測試模型
4. 開始實作爬蟲的程式
5. 有足夠的資料後開始用爬到的資料訓練
6. 完成訓練確定模型有效後開始將每天的爬蟲的資料放進去比較準確度
7. 下結論和將結果轉換成圖表

## Discussion

- [DC 群組](#)(我們語音為主，文字討論不多)
- [HackMD 連結](#)

## Repo

- [Github 連結](#)

## Reference

- <https://levelup.gitconnected.com/random-forest-regression-209c0f354c84>
- <https://towardsdatascience.com/basic-ensemble-learning-random-forest-adaboost-gradient-boosting-step-by-step-explained-95d49d1e2725>
- <https://medium.datadriveninvestor.com/ensemble-learning-and-random-forest-7430ebf3da7e>
- [https://medium.com/@maggieliao.cm04g/ml 技術筆記-3-1-ensemble-learning-bagging-and-the-gradient-boosting-model-family-intro-250c59b9ab60](https://medium.com/@maggieliao.cm04g/ml-技術筆記-3-1-ensemble-learning-bagging-and-the-gradient-boosting-model-family-intro-250c59b9ab60)