

Final Project Proposal

B10632014 陳品羚

B10632026 吳苡瑄

B10632032 黃千綾

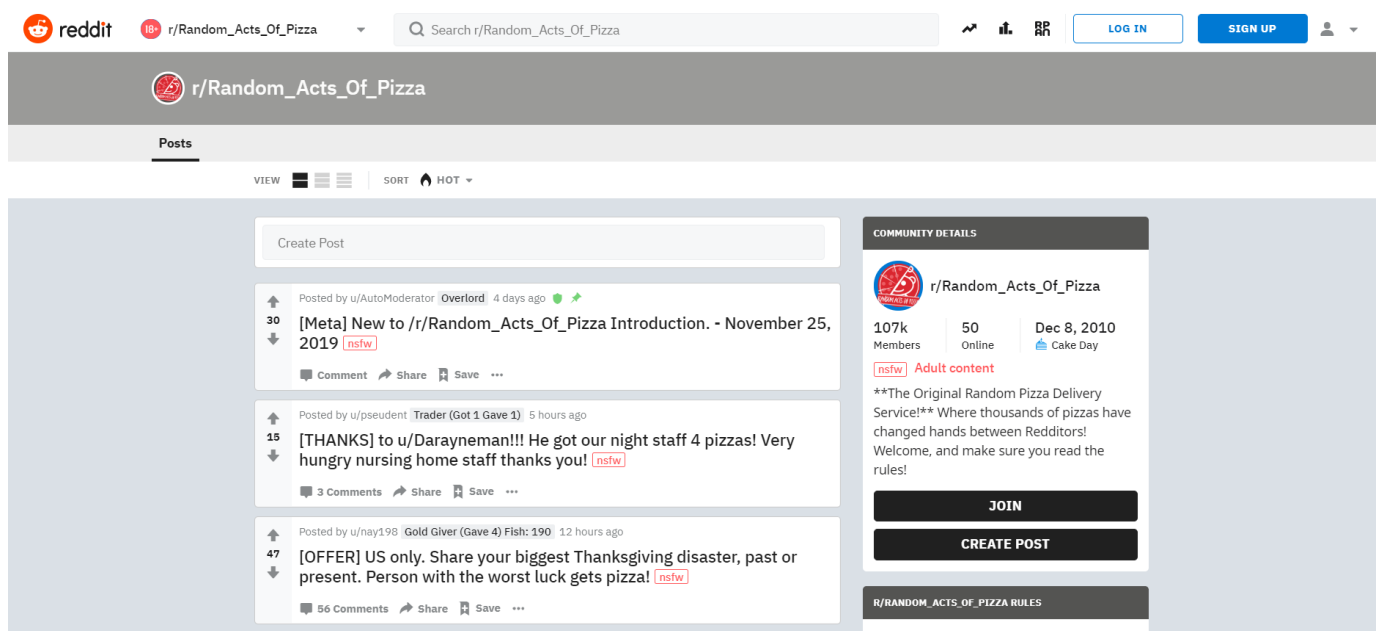
簡介

資料科學可以應用在許多生活情境中，我們經常能從其中得出意外的有趣結論，或是藉由視覺化的資料整理並理解大量的資訊。因此我們希望能找到有趣且能容易呈現的資料來製作期末報告，並希望這些資料是我們的能力範圍所能及，不會太過困難或有太多不確定因子。最後我們在 Kaggle 網站找到了 Random Acts of Pizza(RAOP) 這個資料集。

Althoff 等學者（2014）曾撰寫分析 ROAP 的論文，而 Kaggle 上的資料即是其中所使用的資料集。網路上舉辦過許多與 ROAP 相關的比賽，也有許多前人使用這些資料集的範例。因此我們決定使用這份資料集，並從中分析出具有意義的結果。

RAOP

Random Acts of Pizza(RAOP) 是社群網站 Reddit 的一個 subreddit，類似於 Twitter。



在這個社群中，Reddit 用戶會以 Request 標籤貼文徵求披薩。這些貼文的內容並無限制，可能是一些詩、一首歌，或是說明為什麼現在需要披薩。而其他用戶可以依據這些貼文贈送披薩。徵求的用戶收到披薩後，會發一篇 Thanks 標籤的貼文，裡面通常為附上披薩的照片及手寫的身名稱。

Kaggle 上的 RAOP 資料集便是提供相關的資訊，包含徵求者的詳細資料，例如：帳號創立時間、文章內容、按讚按噓數、是否收到披薩等。

實驗目的

我們希望藉由分析徵求者相關的資料，建立模型，以預測特定貼文是否能夠徵得披薩。

採用技術

我們會先使用課程上所學習的演算法分析非文章的內容。然而若僅靠非文章內容的預測效果不佳，我們會使用一些自然語言處理的方法進行斷詞等處理。此外，我們會使用 **XGBoost** 及 **NLTK** 等 Python 套件。

預期結果

- 能呈現出怎麼樣的條件下容易徵得披薩。
- 能呈現效果明顯的圖表。
- 做出可以呈現 **Request** 貼文性質的文字雲。

預期應用

- 可以作為商業行銷策略的參考資料。

Reference

- RAOP - Restoring Faith in Humanity, One Slice at a Time
(https://www.reddit.com/r/Random_Acts_Of_Pizza/)
- Random Acts of Pizza | Kaggle (<https://www.kaggle.com/c/random-acts-of-pizza/data>)
- Tim Althoff, Cristian Danescu-Niculescu-Mizil, Dan Jurafsky. How to Ask for a Favor: A Case Study on the Success of Altruistic Requests, Proceedings of ICWSM, 2014.
- [資料分析&機器學習] 第5.2講: Kaggle機器學習競賽神器XGBoost介紹
(<https://medium.com/jameslearningnote/%E8%B3%87%E6%96%99%E5%88%86%E6%9E%90-%E6%A9%9F%E5%99%A8%E5%AD%B8%E7%BF%92-%E7%AC%AC5-2%E8%AC%9B-kaggle%E6%A9%9F%E5%99%A8%E5%AD%B8%E7%BF%92%E7%AB%B6%E8%B3%BD%E7%A5%9E%E5%99%A8xgboost%E4%BB%8B%E7%B4%B9-1c8f55cffcc>)
- NLTK 初學指南(一)：簡單易上手的自然語言工具箱－探索篇 (<https://medium.com/pyladies-taiwan/nltk-%E5%88%9D%E5%AD%B8%E6%8C%87%E5%8D%97-%E4%B8%80-%E7%B0%A1%E5%96%AE%E6%98%93%E4%B8%8A%E6%89%8B%E7%9A%84%E8%87%AA%E7%84%B6%E8%AA%9E%E8%A8%80%E5%B7%A5%E5%85%B7%E7%AE%B1-%E6%8E%A2%E7%B4%A2%E7%AF%87-2010fd7c7540>)