

Final project example list

Title: Analysis of Open Data Usage in Taiwan

Motivation:

In this project, we will collect these data of each dataset from the Open Data Platform of Taiwanese Government, and discover the relationship between these data in order to understand how open data is used and what kind of open data is mostly used in Taiwan.

Methods and Tools used:

Omitted.

Expected Results:

Relations between the number of views/downloads and some properties of a dataset, such as last updated date, provider, data format, some items in the metadata, etc.

Title:

《MovieLens Will Be A Hit!》 - 從投資者角度分析 Kaggle MovieLens Dataset 預測電影利
潤與賣座程度

Motivation:

，我們想換個

切入點，關注的不是客戶、而是電影本身，我們想從投資者的角度分析以預測這部電影的表

現。舉例來說：「雷神索爾3：諸神黃昏」真的會像影評所預言的大賣嗎？「奪魂鋸：遊

戲重啟」能不能帶來利潤？抑或是會虧損？諸如此類的問題。

Method (tools)

Sklearn, Tensorflow, Pandas, Numpy, Matplotlib, 也可能會用到現成的 infographics 套件（e.g. D3.js, Processing.js）。

Expected Results

我們暫定的目標有二：首先是建立迴歸模型，以預期某部電影的 revenue，在給定電影的類型、月份、語言、片商...等條件之下。第二，我們將建立有效的分類器，訂定 threshold value 來判斷某部電影是否會賣座、抑或是會讓廠商虧損。在 Infographics（資料視覺化）方面，除了月份或種類為軸的長條圖、圓餅圖，我們將用文字雲（Wordcloud）找最熱門的關鍵字、或是畫出 Confusion Matrix、集群分析...等。我們將深入探討此類問題，如果能準確分析賣座程度、電影利潤，相信這樣的專案能有很大的影響力。

Title:

主題(title): Board Games - What makes a game a good game

Motivation:

我們想找出人們究竟喜歡哪種遊戲，是因為參與的人數多寡，又或者是因為需要動腦筋？競爭類型還是合作類型比較受歡迎？這些是我們想要探究的，並試圖從中找出一些規則，或許未來要打造一款受歡迎的遊戲時可以作為參考。

Method:

資料集取自 Kaggle 上的 Board Games Dataset，原始來源為 boardgamegeek.com，以 80 個屬性描述 90,400 種桌上遊戲，包含發行年份、類型、玩家數、耗時、難度以及玩家給予的評分等。

- 資料前處理：

刪除未上市或是極少人給予評價的遊戲。此外，90400 筆涵蓋眾多名稱相同僅是版本不同的遊戲，亦可視情況刪除，而有些欄位屬於亂碼，也將刪減。藉由去掉 outlier 來提升分析的準確度。

- 資料探勘：

剛開始先用自身想法來想哪些 attribute 會影響受歡迎的程度，藉由分析這些屬性來確認想法是否正確，再由 python 3 實作相關係數矩陣，了解各變項之間的相關性，或是以 Matplotlib 顯示分佈圖、兩變項的散佈圖，藉由這些分析，也許可以組合出哪些維度是重要的，哪些其實並不影響。

用 Clustering 分類，來分析資料

LDA topic modeling

Expected results:

以敘述性統計分析各遊戲類別以及遊戲模式；分析變項之間的相關性，例如：難度與評比、普及率與評比、後悔人數與期待人數等。找出影響最大的 attribute 之後，再用 PCA 降維分析其結果是否正確。

Goal: To understand if a particular detail could bring a game on top

Data science title list

1. Analysis of Open Data Usage in Taiwan
2. Clinical-diagnosis assisting system
3. 《MovieLens Will Be A Hit!》 - 從投資者角度分析 Kaggle MovieLens Dataset 預測電影利潤與賣座程度
4. Encrypted Malicious Flow Detection and Analysis
5. Board Games - What makes a game a good game
6. The Secret Behind Blacklist
7. 糟了!!!是地震
8. 以神經網路作為數據挖掘架構之台灣各地區人口增減、遷移與其他因素之相關性分析
9. Facebook 粉絲專業貼文對閱聽者之情緒預測

10. KKTV Data Game - 預測使用者上線時間
11. Crime Analysis in India
12. 個人化的影劇推薦系統
13. KKBOX Churn Prediction
14. 什麼口味的蘋果即時新聞最好吃?
15. Hey, Terrorist
16. 台北車站交通擁擠預測
17. 明星高中學生上大學後適應狀況之研究
18. 利用台股五檔資訊來預測股票漲跌
19. 人才流動預測與探討
20. House price prediction in USA
21. 航班誤點與取消分析
22. Moovies !
23. 分析網路連線攻擊之特徵
24. Bike Sharing System User Number Prediction
25. 俄羅斯房地產分析與預測
26. Game of Thrones - Explore deaths and battles from this fantasy world
27. 德特·賽恩的煩惱 - 最大化 ig 讚數的秘訣