**2019 Autumn Information Retrieval and Text Mining**

**Big Five 人格與社群媒體文章分析**

**Prof. Chien Chin Chen, Information Management Dept.**

**Project by:**

**B06406009 陳姵如 B06705026 林語萱**

**B05502058 趙禹誠 B05106010 黃冠文**

**一、分析主題**

以人類說話用詞結合心理學 BIG FIVE 性格特質，預測分類公眾人物個性。公眾人物會分以職業分群，猜測職業與人格或相關。

註：BIG FIVE 五大性格特質，即

* [經驗開放性](https://zh.wikipedia.org/wiki/%E7%BB%8F%E9%AA%8C%E5%BC%80%E6%94%BE%E6%80%A7)（Openness to experience）
* [盡責性](https://zh.wikipedia.org/wiki/%E5%B0%BD%E8%B4%A3%E6%80%A7)（Conscientiousness）
* [外向性](https://zh.wikipedia.org/wiki/%E5%A4%96%E5%90%91%E6%80%A7%E4%B8%8E%E5%86%85%E5%90%91%E6%80%A7)（Extraversion）
* [親和性](https://zh.wikipedia.org/wiki/%E4%BA%B2%E5%92%8C%E6%80%A7)（Agreeableness）
* [情緒不穩定性](https://zh.wikipedia.org/wiki/%E7%A5%9E%E7%BB%8F%E8%B4%A8)（Neuroticism）

**二、方法**

（一）發放 BIG FIVE 心理學人格分類問卷於NTU交流版、取得同學人格分析與社群媒體資料使用權

（二）使用官方 API 或自己寫爬蟲爬他們的社群媒體發文用字資料

（三）以個性分成五類並以上述資料 likelihood（暫定）方式建模

（四）使用官方 API 或自己寫爬蟲爬目標名人的社群媒體發文用字資料，用 Naive Bayes分析目標名人的五大性格特質

註：社群媒體來源：Facebook**、**Instagram

**三、資料說明**

（一）中文繁體為此次資料探勘報告之分析對象

（二）[名人列表](https://docs.google.com/spreadsheets/d/1cfyPRklZypx-U_tfVaaNzN-WMnmbnrUsr3khRKUmcR8/edit?usp=drive_web&ouid=115721435623430820746)（尚在新增，預計100筆）分為政治 / 娛樂 / 運動 / 文學 四類

（三）[問卷](https://forms.gle/kk3QCyBBAqxf3hXn8)包含填答人臉書與instagram的帳號蒐集以及五大人格特質測驗兩大部分；五大人格特質測驗內容參考Jones, G., George, J. (2019). Values, attitudes, emotions, and culture: the manager as a person. In Essentials of Contemporary Management. New York, NY: McGraw-Hill Education.