**Final project計劃書**

**組別：16**

**B07902034王昱凱、B07902126謝宗儒、B08902035彭光湞、B08901046詹侑昕**

**專案任務：**

利用YouTube影片的相關資訊(下面會列舉)，預測相似類型影片是否將成為受大眾認可的優質影片(預測讚好率)。影音串流平台佔據了現代生活重要的一塊，在此商機無限，若能建立一個預測模型，掌握影片類別、屬性與品質的關係，將有利於廣告投放，且能透過YouTube演算法加速推廣產出的影片。

**資料集來源：**

Kaggle上的Trending YouTube Video Statistics 裡的資料集

https://www.kaggle.com/datasnaek/youtube-new?select=CAvideos.csv

總共約為20萬至30萬筆資料

欄位數量暫定為11欄 (將視情況training狀況調整feature)

[country(發布者所屬的國家), Video\_ID, trending data(上發燒的日期), title, channel title, category\_ID(YouTube影片分類代碼), publish time, tags(影片相關關鍵字), views, likes, dislikes]

調整方向：

tags無固定種類，不易透過課堂所學處理，因此將改取用tags number、trending data將合併成同部影片上發燒的天數、publish time將改取用上傳至特定日期的天數、title及channel title會改取用兩者重複比率或字串相對長度(考慮語言差異)等。

範例資料：

[CA,

n1WpP7iowLc,

17.14.11,

Eminem - Walk On Water (Audio) ft. Beyoncé,

EminemVEVO,

10, 2017-11-10T17:00:03.000Z,

Eminem|"Walk"|"On"|"Water"|"Aftermath/Shady/Interscope"|"Rap",

17158579, 787425, 43420]

[DR,

LgVi6y5QIjM,

17.14.11,

Sing zu Ende! | Gesangseinlagen vom Feinsten | inscope21,

inscope21,

24, 2017-11-13T17:08:49.000Z,

inscope21|"sing zu ende"|"gesangseinlagen"|"gesang"|"gesungen"|"songs nachsingen"|"nachsingen"|"song...,

252786, 35885, 230]

[FR,

Ro6eob0LrCY,

17.14.11,

Malika LePen : Femme de Gauche – Trailer,

Le Raptor Dissident,

24, 2017-11-13T17:32:55.000Z,

Raptor"|"Dissident"|"Expliquez"|"moi"|"cette"|"merde,

212702, 29282, 1108]

[GB,

Jw1Y-zhQURU,

17.14.11,

John Lewis Christmas Ad 2017 - #MozTheMonster,

John Lewis,

26, 2017-11-10T07:38:29.000Z,

christmas|"john lewis christmas"|"john lewis"|"christmas ad"|"mozthemonster"|"christmas 2017"|"chris...,

7224515, 55681, 10247]

[JP, 5ugKfHgsmYw,

18.07.02,

陸自ヘリ、垂直に落下＝路上の車が撮影,

時事通信映像センター,

25, 2018-02-06T03:04:37.000Z,

事故|"佐賀"|"佐賀県"|"ヘリコプター"|"ヘリ"|"自衛隊"|"墜落"|"落下"|"現場"|"神埼市"|"ドライブレコーダー",

188085, 591, 189]