

# 111\_1 資訊管理

## 期末專案書面報告

基於機器學習預測募資課程熱門程度並探討關鍵變數

— 以 Hahow 好學校為例

第三組

R11725004 陳柏言

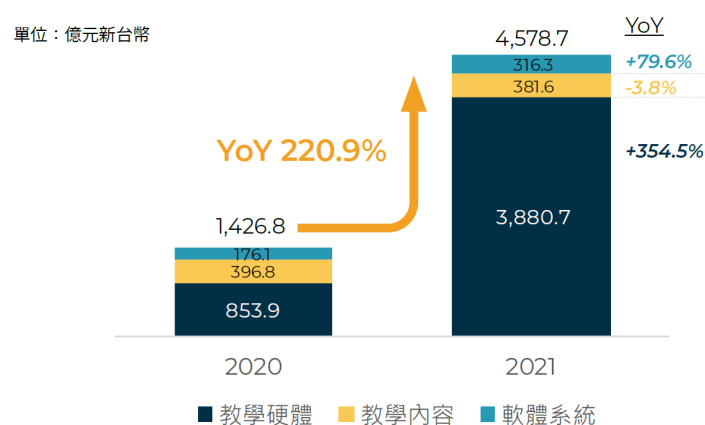
R11725015 胡予瑄

R11725039 林宜璇

R11725040 楊佳真

## 一、研究背景與動機

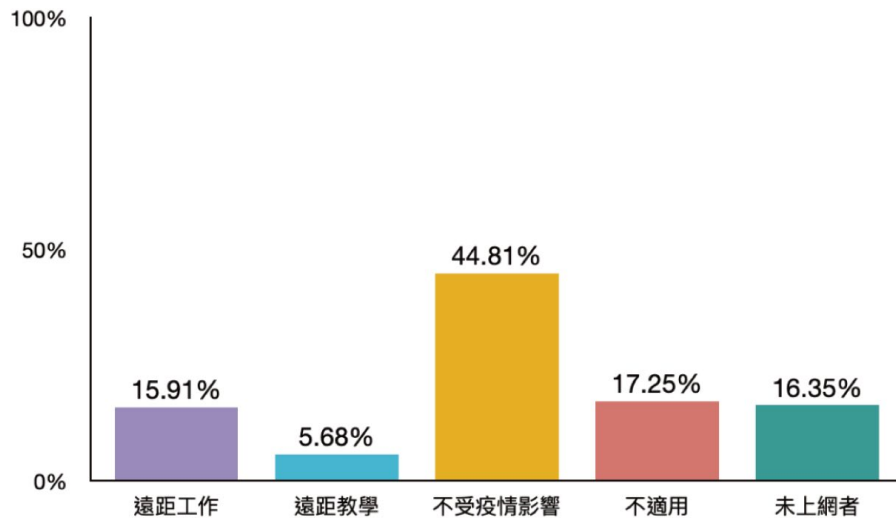
隨著科技不斷發展，網路已深入影響生活的各個層面，其即時性與方便性，使得人們已習慣從網路取得各種資訊。根據 2022 年台灣網路調查報告<sup>[1]</sup>指出，台灣 18 歲以上、近三個月有上網經驗的民眾為 84.3%，且在全部人口中，有五分之一的人每天無時無刻不在上網。而網路也改變了學習的形式，促使線上教學及影音學習等上課方式興起。2021 年美國研究機構 HolonIQ<sup>[4]</sup> 調查指出，2025 年全球教育市場規模將達到 7.3 兆美元，其中數位內容和線上教育市場預計於 2025 年將達到 2 倍以上的成長。另外，綜觀台灣市場的表現，根據 2021 年台灣智慧學習產業調查報告<sup>[5]</sup> 中顯示，智慧學習產業總產值來到 164 億美元，約台幣 4578 億元，與 2020 年相比，整體產值年成長率高達 220.9%，如圖一所示。



(圖一)

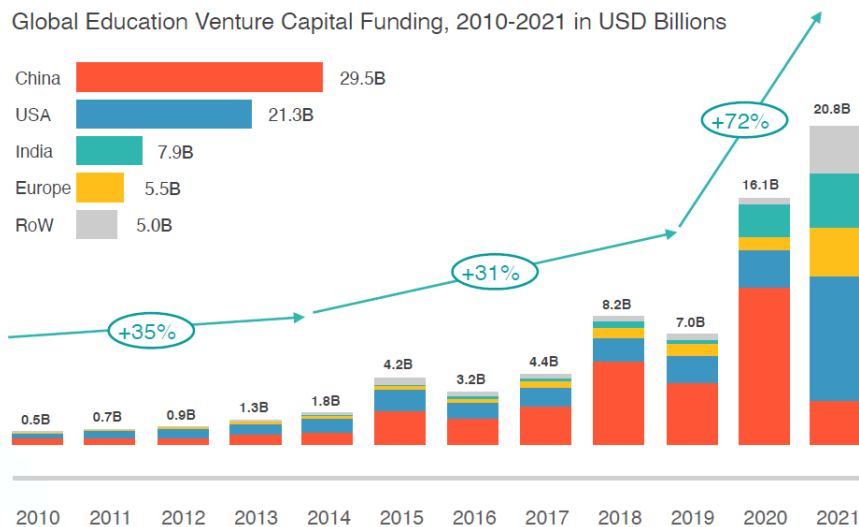
自從 2020 年以來，全球持續受到 COVID-19 疫情影響，不僅產業經濟發生變化，也影響了我們的生活樣貌，根據 2022 年台灣網路調查報告<sup>[1]</sup> 顯示，台灣民眾因疫情採取遠距工作或遠距教學的比例達 21.6%，其中遠距工作為 15.91%、遠距教學為 5.68%，如圖二所示。而教育科技產業在疫情影響下，也成為投資人高度關注的產業，2021 年 HolonIQ 調查報告<sup>[4]</sup> 顯示，自從 2020 年疫情開始，全球教育科技創投金額突破新高，2021 年來到 208 億美元，與 10 年前整體投資額相比，成長超過 30 倍，使得數位學習資源變得更加豐富，如圖三所示。如今，透過社群媒體、線上課程平臺或相關軟體，教學者便能藉由手機、電腦等裝置進行授課，而學生同樣能透過這些智慧型裝置進行學習。

約二成二的台灣民眾因應疫情採取遠距工作或遠距教學



資料來源：2022台灣網路調查，執行時間2022年2月14日至3月15日，加權後數值。  
樣本數：1113（市話樣本）

（圖二）



（圖三）

伴隨科技發展、自主學習觀念逐漸盛行，人們學習的管道變得更加多元，根據 2018 年 Pearson 的調查<sup>[7]</sup>指出，Z 世代年輕人更傾向於透過 Youtube 及其他應用程式進行學習。同樣地，創作者的呈現管道也變得更加多元化，除了在 Facebook、Youtube、Podcast 等社群平臺或影音平台發布創作，許多 KOL、領域專家也投入線上教學的行列，在線上課程平臺上發布課程，透過傳播知識收

取更多獲利，如圖四所示。因此，在創作平臺多元發展下，更多的創作者紛紛投入創作者市場。2022 年 Influencer Marketing Hub 的報告<sup>[3]</sup>顯示，創作者經濟的市場規模於 2022 年預估高達 1040 億美元，且全球約有 5000 萬人將自己視為內容創作者。



(圖四)

綜觀以上背景，在創作熱潮及線上學習趨勢下，如何設計出學生感興趣的熱門課程，對於線上課程教學者或創作者而言十分重要！

## 二、研究目標與重要性

### 專案研究目的

本專案研究目的分為二點：

1. 蒐集線上課程平臺上的課程資料，透過資料探勘、文字探勘技術，針對不同類型資料進行前處理與特徵工程，建立基於機器學習之預測模型，以便在課程上架前，達到預測一課程之熱門程度的目的。
2. 藉由此預測模型，找出影響模型預測表現之重要特徵，作為衡量課程熱門程度之關鍵變數。

對創作教學者而言，可根據本研究找出的重要特徵，調整課程介紹與課程內容。對線上課程平臺而言，預測結果可協助廣告內容的調整，將預測為熱門之線上課程放入廣告中，提升平臺知名度並增加接觸潛在學員的機會，協助平臺獲取更多收益。

### 以台灣線上課程募資平臺 Hahow 好學校為研究對象

近 10 年來，許多線上課程平臺紛紛出現，提供多元化的課程供學員學習，根據 2020 年 OpView 線上課程平臺聲量分析報告<sup>[6]</sup>顯示，討論度最高的前三大跨領線上課程平臺分別為 Hahow 好學校、YOTTA 及 Udemy，並在疫情爆發後，線上課程的需求增加，整體線上學習聲量有所增長，其中以遠距教學為主，帶來了自主學習新熱潮。在討論度最高的 3 大線上課程平臺中，Hahow 好學校及 YOTTA 為台灣創立的平臺，皆使用募資的形式進行開課，Udemy 則為全球最大的線上課程平臺，開課條件寬鬆，學員及課程數量十分龐大。

Hahow 好學校<sup>[2]</sup>為國內最大的線上課程募資平臺，成立於 2015 年，由七年級生團隊所創，取名自台語的「學校」發音。Hahow 透過全球首創的「群募開課」機制，成為涵蓋多元領域的線上課程平臺，其學員人數達 65 萬以上，講師創作者達 600 多位，共涵蓋 700 多堂的線上課程。除了線上預錄課程之外，

Hahow 亦提供多元的學習方式，包含直播學習、圖文學習及學習專班等，滿足不同學員的學習需求。而 Hahow 目前也將服務對象拓展至企業，透過多元跨領域的課程，協助企業提升員工職場技能及跨領域知識，目前企業學員達到 6 萬多名，涵蓋 15 種以上的產業類別，提供 250 堂以上的多元課程。因此，本專案以線上課程募資平臺 Hahow 為研究對象。



### 研究課程「熱門」的重要性

熱門課程不僅可以提升教學者或創作者的口碑與收益，線上課程平臺也可將課程作為廣告宣傳內容，增加平臺自身的知名度，觸及更多的潛在顧客並提升未來的收益。若能透過蒐集線上課程平臺上的資料，利用資料分析技術，挖掘出哪些關鍵變數與課程熱門相關，並建立預測課程是否熱門的工具，便能協助教學者或創作者調整課程介紹與教學內容，提升課程熱門程度，平臺方也能透過預測結果，調整廣告宣傳內容，並進一步給予教學者課程內容修改建議。

### 三、研究文獻探討

在眾多線上課程中，最常見的是大規模開放線上課程（MOOCs），學生能在 MOOCs 中獲得高品質的學習機會，只需要電腦和網路等設備，世界各地學生就可以獲得來自全球最好的學校和機構的公開高品質課程<sup>[15]</sup>，而本研究所鎖定的線上課程平臺 Hahow，亦是屬於大規模開放線上課程的一種，Hahow 透過全球首創的群募開課機制，打造自身平臺的獨特性，與其他平臺相比更具有其特色<sup>[11]</sup>。

#### 探討有關購買意圖之相關研究

在過去與線上課程相關的研究中，許多研究聚焦於線上課程的購買意圖，如蘇暘升等人，以 SOR 理論探討線上課程之購買意圖<sup>[13]</sup>，以及鄧楷馨等人，以信號理論結合擴充評價架構，探討線上課程的購買意圖<sup>[12]</sup>。另外，林慶儒等人，採用修正後的 ECT 模型作為架構，探討消費者加入線上募資開課的線上課程後，有哪些因素會影響消費者再次購買相同平臺中的其他課程<sup>[9]</sup>。

#### 預測群眾募資平台上之專案募資成功與否之相關研究

在預測群眾募資平台上之專案募資成功與否的相關研究中，R. S. Kamath 等人透過爬取 Kickstarter 產品募資平台上超過 120 個產品專案的頁面，其中頁面的架構包含影片、募資達標的資金目標、產品專案的描述、社群平台的連結、支持者人數、產品專案隸屬的類別及子類別、資助期限等資訊，利用這些特徵來對募資結果進行預測，而該研究中所訂定的預測變數是將募資到的資金百分比依據比例多寡，分為五大類別：0 funded、less funded、partial、successful、above goal。使用分類器則包括 Random Forest、NAÏVE BAYES、ANN 來進行分類的預測。最終結果發現若產品專案的頁面上有放置影片相較於沒有放置影片的產品專案，對於募資結果可以達到較高的成功率，是攸關成功與否的一項重要特徵；至於分類器則是神經網路表現最好，可達到九成的準確率。

而在 Chen 等人的研究中，他們使用 Kickstarter 上的多達 20,000 個產品專案資訊，利用 SVM 模型，針對 Kickstarter 上的產品專案在完成前是否會成功來進行預測（隨著時間變化會有不同的預測準確率）。他們在處理特徵時，還加入了一些外部資料，包括附在產品專案頁面上的社群媒體資訊，例如，一項產品專案若有使用 YouTube 影片介紹，該影片的觀看次數為何，以及若該產品連結有在 Twitter 上被轉發或提及的次數多少，加入這些作為新的特徵，最後發現對預測結果影響較大的特徵仍是產品專案頁面本身的資訊，例如隨著時間的募資金額變化，而他們根據 SVM 分類器預測的準確度有達九成。

### 使用機器學習探討線上課程之相關研究

過去研究中，李妮燁等人針對 Hahow 課程平臺，蒐集用戶課程購買的資訊及平台上的文本資料，透過文字探勘技術對不同種類的資料進行前處理及特徵選擇，建立了機器學習混和模型，用以預測用戶的課程花費總金額<sup>[10]</sup>。而 Liu, Z., & Zhang, Y. 等人，提出了一種混合方法，挖掘 MOOC 討論區的學員角色，並納入其他學習活動共同評估學員的學習品質，其中，他們更關注於提取論壇中貼文和評論的語義特徵，也表示這有助於提升模型預測性能<sup>[17]</sup>。

### 熱門預測之相關研究

在預測「熱門」的相關研究中，研究目標主要為基於網路內容的熱門預測，例如 Hongjun Jeon 等人，使用混和式機器學習模型，針對影音串流平臺上尚未發布的影音內容進行熱門預測<sup>[16]</sup>，以及 Arie Rachmad Syulistyo 等人使用多層感知器 MLP，針對新聞內容進行熱門預測<sup>[14]</sup>。然而，較少文獻使用機器學習研究線上課程的熱門預測。

### 線上課程熱門預測之相關研究

與線上課程熱門預測相關的研究中，Zachari Swiecki 等人針對 Udemy 平臺上的線上課程，使用 4 種機器學習模型進行熱門預測<sup>[18]</sup>。而 Zachari Swiecki 等人蒐集的資料包含三大類，內容型（課程介紹、名稱等）、結構型（影片時長、作業數等）、描述型（價格、發布年月日等）。針對類別型資料，他們使用 One-



hot Encoding，針對文本資料則使用 BERT 轉換成向量。而在特徵選擇階段，由於他們研究目標在於預測尚未發布的線上課程，因此發布後所產生的特徵資料不會納入模型訓練，另外值得注意的是，該研究提及預測熱門可分為兩種類別，一為分類受歡迎與受歡迎，一為與熱門相關的指標數值迴歸分析，而該研究選擇兩種與熱門相關的數值型變數分別進行預測，分別為「平均每月上課人數」及「平均每月評論數」，目的是希望「熱門」能以兩種不同的角度評量熱門程度，而除以月份數的目的是在於減緩時間累積對於熱門的影響。由於預測變數為數值，因此他們使用線性迴歸、隨機森林、XGBoost、MLP 4 種機器學習模型進行迴歸分析，根據預測結果，XGBoost 的表現最好，預測每月平均上課人數及每月平均評論數的 RMSE 分別為 2.72351 和 1.9057。

## 文獻探討總結

在線上課程蓬勃發展的趨勢下，對於教學者(或創作者)及平臺而言，預測課程「熱門」並探討其關鍵變數十分重要，然而綜觀過去研究，僅少數研究使用機器學習針對線上課程熱門預測進行探討，且目前並未發現相關文獻針對 Hahow 募資課程平臺進行研究，故本研究欲藉由 Hahow 平臺上的課程資訊，建立機器學習模型，用以進行未發布之線上課程熱門預測，並透過影響模型之重要特徵，探討「熱門」的關鍵變數！

## 四、研究方法與步驟

### 第一階段 資料收集與前處理

使用 Hahow API 爬取 Hahow 官網中的「詳細課程資料」，進一步整理爬取後的 json 檔資料，可取得「詳細課程資訊」248 個欄位的資訊(共 915 筆資料)。



### 第二階段 資料前處理

刪除「詳細課程資訊」中無意義的欄位，並且移除零元課程、募資中課程及開課時間為 2023 年的課程。此外，考量疫情帶動線上課程購買的因素，故刪除疫情前 2015 年至 2019 年的課程，形成新資料集。

針對數值型態的資料，進行空值處理(補植)，並且將類別資料(如狀態、課程分類、課程子分類)，進行 One-hot encoding。針對文字型態的資料，則是進行斷詞等前處理(包含中文斷詞、英文轉小寫、移除 stop words、stemming、TF-IDF 向量轉換等)。資料集欄位如下：

前處理後的 Hahow 課程資料集			
變數分類	英文變數名稱	中文變數名稱	型態
價格	preOrderedPrice	優惠價	Numeric
	price	原價	
時間	expireDate	募資截止時間	Date
	startDate	募資起始時間	
	estimatedCourseStartTime	預計開課時間	
	status	狀態	Nominal
授課老師	ownerName	授課老師姓名	String
	ownerDescription	授課老師文字介紹	
	ownerSkill	授課老師技能	
	ownerInterest	授課老師興趣	
課程資訊	CourseTitle	課程名稱	String
	title	課程分類	
	subtitle	課程子分類	
	description	課程介紹	
	recommendedBackground	建議知識背景	
	requiredTools	需要的設備工具	
	targetGroup	目標學生	
	willLearn	可以學到什麼	
	totalVideoLengthInSeconds	課程影片總長度(秒)	Numeric
	averageRating	平均星星數	

本研究進行兩種變數的預測，以不同角度評量熱門程度。兩種預測變數分別為「平均每個月上課人數」及「是否熱門」。前者採取自 Zachari Swiecki 等人研究中的其中一種預測變數，以人數為熱門的依據，並除以時間以減緩時間累積的影響，而後者以平臺的判斷為熱門依據，根據課程是否為 Hahow 平臺上的熱門課程所產生。預測變數如下表所示：

預測變數名稱	預測變數說明	型態
平均每個月上課人數	上課人數除以開課至今之月份數	Numeric
是否為熱門課程	是否為Hahow上熱門課程	Boolean

### 第三階段 特徵工程

Zachari Swiecki 等人於其研究中，使用之特徵資料包含三大類，內容型（課程介紹、名稱等）、結構型（影片時長、作業數等）、描述型（價格、發布年月日等），其中並未涵蓋與老師介紹相關的特徵資料，然而老師本身是否有名氣、經驗豐不豐富，有可能與學生上課人數相關，因此我們將與老師資訊這個與「人」相關的特徵資料納入，並根據現有特徵，加入其他資訊或想法，創造出新的特徵進行探討。以下為進行特徵工程後新增之變數：

新增之特徵	變數名稱	說明	類型
適配性	「授課老師」與「課程介紹」之適配性	「授課老師」與「課程介紹」之 Cosine Similarity	Numeric
	「授課老師」與「可以學到什麼」之適配性	「老師介紹」與「可以學到什麼」文字之 Cosine Similarity	
	「授課老師」與「目標學生」之適配性	「老師介紹」與「目標學生」文字之 Cosine Similarity	
	「授課老師」與「課程子分類」之適配性	「授課老師」與「課程子分類」文字之 Cosine Similarity	
	「授課老師」與「課程分類」之適配性	「授課老師」與「課程分類」文字之 Cosine Similarity	
	「授課老師」與「課程名稱」之適配性	「授課老師」與「課程名稱」之 Cosine Similarity	
	「授課老師」與「建議知識背景」之適配性	「授課老師」與「建議知識背景」之 Cosine Similarity	
	「授課老師」與「需要的設備工具」之適配性	「授課老師」與「需要的設備工具」之 Cosine Similarity	

新增之特徵	變數名稱	說明	類型
價格	課程CP值	以影片時長除以課程價格算出	Numeric
	子分類中的課程數	考量不同子分類課程中的數量不同	Numeric
課程資訊	課程介紹文長	考量不同課程中的文字長度不同	Numeric
	圖表數量	課程介紹內的「圖表數量」	Numeric
	是否為平台名稱	課程的授課老師名稱是否為某個品牌、企業、平台，而非授課老師姓名	Boolean
	授課老師課程數	該授課老師過去共開設過多少課程	Numeric
	是否曾開課	該授課老師是否曾開課	Boolean
授課老師	授課老師課程星數	該老師所有開設課程的平均星數	Numeric
	授課老師社群熱門分數	該授課老師其 FB、IG、YT 粉絲人數的總和	Numeric

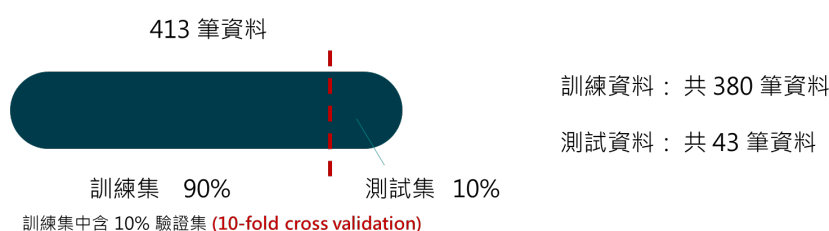
合併新增的欄位後，最終的 Hahow 課程資料集為：

新 Hahow 課程資料集			
變數分類	英文變數名稱	中文變數名稱	型態
價格	preOrderedPrice	優惠價	Numeric
	price	原價	
	courseCP	課程CP值	
時間	expireDate	募資截止時間	Date
	startDate	募資起始時間	
	estimatedCourseStartTime	預計開課時間	
	status	狀態	Nominal
三個欄位 合併成 aboutOwner	ownerName	授課老師姓名	String
	ownerDescription	授課老師文字介紹	
	ownerSkill	授課老師技能	
	ownerInterest	授課老師興趣	
	isBrand	是否為平台名稱	Boolean
授課老師	courseNumHisty	授課老師課程數	Numeric
	setUpCoursesAgain	是否曾開課	Boolean
	averageRatingHisty	授課老師課程星數	Numeric
	KOLScore	授課老師社群熱門分數	Numeric
課程資訊	CourseTitle	課程名稱	String
	title	課程分類	
	Subtitle	課程子分類	

新Hahow 課程資料集			
變數分類	英文變數名稱	中文變數名稱	型態
課程資訊	Description	課程內容文字介紹	String
	recommendedBackground	建議知識背景	
	requiredTools	需要的設備工具	
	targetGroup	目標學生	
	willLearn	可以學到什麼	
	totalVideoLengthInSeconds	課程影片總長度(秒)	Numeric
	subtitleCourseNum	子分類中的課程數	
	lengthOfDescription	課程介紹文長	
	PictureNum	圖表數量	
適配性	teacher_metaDescription_similarity	「授課老師」與「課程介紹」之適配性	Numeric
	teacher_willLearn_similarity	「授課老師」與「可以學到什麼」之適配性	
	teacher_targetGroup_similarity	「授課老師」與「目標學生」之適配性	
	teacher_group__subGroup__title_similarity	「授課老師」與「課程子分類」之適配性	
	teacher_group__title_similarity	「授課老師」與「課程分類」之適配性	
	teacher_title_similarity	「授課老師」與「課程名稱」之適配性	
	teacher_recommendedBackgro und_similarity	「授課老師」與「建議知識背景」之適配性	
	teacher_requiredTools_similarity	「授課老師」與「需要的設備工具」之適配性	

#### 第四階段 切分訓練資料與測試資料

將 915 筆資料，切分為 90% 訓練集(訓練集中含 10% 驗證集)、10% 測試集。訓練資料共 380 筆、測試資料共 43 筆資料。



## 第五階段 特徵選擇 **\*從此頁開始修正**

在 Zachari Swiecki 等人的研究中，由於他們研究目標在於預測尚未發布的線上課程，因此發布後所產生的特徵資料不會納入模型訓練。本研究目的同樣希望預測未發布之課程，除了使用原本特徵，根據不同選擇方式進行特徵選擇，分別為人為選擇特徵、演算法選擇特徵，以進行模型預測表現之比較。

### 1. 人為選擇特徵：

人為選擇特徵		
第一種挑法	價格	preOrderedPrice、price、courseCP
第二種挑法	課程資訊	group_title、group_subGroup_title、subtitleCourseNum、pictureNum、lengthOfDescription、totalVideoLengthInSeconds
第三種挑法	授課老師	setUpCoursesAgain、owner_name、courseNumHisty、isBrand、averageRatingHisty、total_fans
第四種挑法	適配性	teacher_metaDescription_similarity、teacher_group_subGroup_title_similarity、teacher_targetGroup_similarity、teacher_title_similarity、teacher_group_title_similarity、teacher_recommendedBackground_similarity、teacher_willLearn_similarity、teacher_requiredTools_similarity
第五種挑法	全部特徵	結合上述四種挑法

### 2. 演算法選擇特徵：這部分會在第六階段做說明。

## 第六階段 模型訓練

在 Zachari Swiecki 等人的研究中，他們選擇線性迴歸、隨機森林、XGBoost 及 MLP 進行模型訓練，其中線性迴歸的預測表現最差，因此本研究不使用線性迴規模型，改採用 SVM 模型。另外，本研究欲預測數值型變數「平均每個月上課人數」及二元類別變數「是否為熱門課程」，使用 Random Forest、XGBoost、SVM，進行模型訓練，並針對兩種預測變數，比較各模型之間的預測表現。最後，本研究挑選表現最好的模型，並查看該模型的重要特徵。

**在此修正為使用 10-fold cross validation 進行模型訓練與驗證**



● 人為選擇特徵 / 二元類別變數「是否為熱門課程」

人為選擇特徵 / 二元類別變數「是否為熱門課程」													
挑法	特徵選擇	XGBoost				Random Forest				SVM			
		Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
挑法 1	價格	0.48	0.53	0.55	0.54	0.51	0.56	0.61	0.58	0.55	0.57	0.76	0.65
挑法 2	課程資訊	0.56	0.60	0.65	0.62	0.58	0.63	0.67	0.64	0.61	0.63	0.69	0.66
挑法 3	授課老師	0.58	0.62	0.61	0.62	0.58	0.59	0.82	0.67	0.56	0.57	0.90	0.69
挑法 4	適配性	0.53	0.58	0.61	0.59	0.53	0.57	0.67	0.62	0.55	0.65	0.89	0.69
挑法 5	全部特徵	0.56	0.60	0.63	0.61	0.58	0.60	0.79	0.67	0.58	0.60	0.75	0.67

首先在預測二元類別變數「是否為熱門課程」中，使用授課老師相關特徵的情況下，XGBoost 模型有最佳表現。而 Random Forest 模型的部分，在使用課程資訊相關特徵、授課老師相關特徵和全部特徵的情況下，預測表現較好。SVM 模型則在使用課程資訊相關特徵的情況下，表現最好，準確率為 61 %。

● 演算法選擇特徵 / 二元類別變數「是否為熱門課程」

在預測數值型變數「是否為熱門課程」的部分，為找出機器（模型）所選的重要特徵，分別透過 XGBoost 和 Random Forest 模型挑選 10 個及 20 個特徵，接著根據所挑選的特徵，分別使用 XGBoost、Random Forest 和 SVM 模型進行預測，以下為所模型選特徵及其預測結果：

		演算法選擇特徵 / 二元類別變數「是否為熱門課程」
XGBoost	挑選前10	['group_subGroup_title_應用設計', 'courseNumHisty', 'group_title_語言', 'group_subGroup_title_英文', 'group_title_生活品味', 'group_title_行銷', 'preOrderedPrice', 'averageRatingHisty', 'teacher_requiredTools_similarity', 'subtitleCourseNum']
	挑選前20	['group_subGroup_title_應用設計', 'courseNumHisty', 'group_title_語言', 'group_subGroup_title_英文', 'group_title_生活品味', 'group_title_行銷', 'preOrderedPrice', 'averageRatingHisty', 'teacher_requiredTools_similarity', 'subtitleCourseNum', 'teacher_willLearn_similarity', 'totalVideoLengthInSeconds', 'group_title_藝術', 'teacher_group_title_similarity', 'price', 'total_fans', 'pictureNum', 'teacher_title_similarity', 'teacher_targetGroup_similarity', 'lengthOfDescription']
Random Forest	挑選前10	['courseCP', 'total_fans', 'totalVideoLengthInSeconds', 'teacher_metaDescription_similarity', 'teacher_willLearn_similarity', 'lengthOfDescription', 'subtitleCourseNum', 'teacher_targetGroup_similarity', 'preOrderedPrice', 'price']
	挑選前20	['courseCP', 'total_fans', 'totalVideoLengthInSeconds', 'teacher_metaDescription_similarity', 'teacher_willLearn_similarity', 'lengthOfDescription', 'subtitleCourseNum', 'teacher_targetGroup_similarity', 'preOrderedPrice', 'price', 'teacher_title_similarity', 'pictureNum', 'teacher_recommendedBackground_similarity', 'teacher_requiredTools_similarity', 'teacher_group_subGroup_title_similarity', 'teacher_group_title_similarity', 'averageRatingHisty', 'group_title_生活品味', 'group_title_語言', 'courseNumHisty']

首先在 XGBoost 所選特徵中，比較重要的前幾個是如「group\_subGroup\_title\_英文」、「group\_title\_語言」、「group\_subGroup\_title\_生活品味」等這些代表課程類別的特徵，表示這些

類型的課程可能可作為預測一課程是否熱門的重要變數。接著排序的是「preOrderPrice」，也就是募資時優惠的課程售價，顯示一個課程的優惠價格也可以作為判定是否熱門的重要特徵。

再來有幾個比較值得做討論的，就是我們這次研究中生成的新特徵或是有經過一些特別處理的，包括「averageRatingHistory（老師過去課程平均評分）」、「subtitleCourseNum（子分類的課程數）」、「total\_fans（總粉絲人數）」、「price（課程價格）」、「pictureNum（課程介紹中圖片數量）」、「lengthOfDescription（課程介紹文長）」、「teacher\_targetGroup\_similarity（老師與目標受眾的相似度）」、「teacher\_wiiLearn\_similarity（老師與將會學到內容的相似度）」、「teacher\_group\_title\_similarity（老師與課程類別的相似度）」、「teacher\_title\_similarity（老師與標題的相似度）」等相似度。

而在 Random Forest 所選特徵中，大致上都與 XGBoost 所選特徵相同，比較不一樣的地方在於多了「courseCP（課程 CP 值，為影片時長除以課程價格）」、「teacher\_metaDescription\_similarity（老師與課程介紹的相似度）」、「teacher\_recommendedBackgroud\_similarity（老師與建議知識背景的相似度）」。

以上模型所選重要特徵，說明了課程的價格、內容多寡、課程類型、老師過去課程評價、同類型課程數量及老師與課程的適配性（相似度），是影響課程熱門的重要變數。

XGBoost		演算法選擇特徵 / 二元類別變數「是否為熱門課程」			
挑選前10		Accuracy	Precision	Recall	F1
	XGBoost	0.60	0.63	0.66	0.65
	Random Forest	0.63	0.66	0.69	0.67
	SVM	0.61	0.62	0.77	0.69
挑選前20	XGBoost	0.62	0.67	0.68	0.67
	Random Forest	0.62	0.64	0.74	0.69
	SVM	0.60	0.61	0.76	0.68

使用 XGBoost 挑選的重要特徵後，模型表現有所提升，在選擇 10 個特徵情況下，Random Forest 模型表現最好，準確率為 63 %，而表現最差的 XGBoost 模型準確率為 60 %。而在選擇 20 個特徵情況下，XGBoost 模型表現最好，準確率為 62 %，SVM 表現最差，準確率為 60 %。

Random Forest	演算法選擇特徵 / 二元類別變數「是否為熱門課程」				
		Accuracy	Precision	Recall	F1
挑選前10	XGBoost	0.56	0.59	0.63	0.61
	<b>Random Forest</b>	<b>0.59</b>	<b>0.60</b>	<b>0.65</b>	<b>0.63</b>
	SVM	0.56	0.57	0.83	0.68
挑選前20	XGBoost	0.58	0.61	0.66	0.63
	Random Forest	0.61	0.62	0.73	0.65
	<b>SVM</b>	<b>0.62</b>	<b>0.61</b>	<b>0.85</b>	<b>0.71</b>

使用 Random Forest 挑選的重要特徵後，在選擇 10 個特徵情況下，所有模型表現皆不如前者使用 XGBoost 挑選重要特徵的表現，表現最佳的 Random Forest 模型準確率僅為 59 %。而在選擇 20 個特徵情況下，SVM 模型表現最好，準確率為 62 %，XGBoost 表現最差，準確率為 58 %。

● 人為選擇特徵 / 數值型變數「平均每個月上課人數」

人為選擇特徵 / 數值型變數「平均每個月上課人數」				
挑法	特徵選擇	XGBoost RMSE	Random Forest RMSE	SVM RMSE
挑法 1	價格	3990.983343	2871.267629	<b>2538.120174</b>
挑法 2	課程資訊	3053.780622	2969.508614	2538.433224
挑法 3	授課老師	<b>2543.313836</b>	<b>2510.536174</b>	2539.007990
挑法 4	適配性	5542.011635	3896.055153	2538.332320
挑法 5	全部特徵	2704.380569	2521.617219	2538.755982

在預測數值型變數「平均每個月上課人數」的部分，XGBoost 和 Random Forest 模型在使用授課老師相關的特徵情況下，RMSE 較低；而 SVM 模型則是在使用價格相關的特徵情況下，RMSE 較低。總體來看，使用授課老師相關特徵及 Random Forest 模型，擁有最低的 RMSE、表現最好。

● 演算法選擇特徵 / 數值型變數「平均每個月上課人數」

在預測數值型變數「平均每個月上課人數」的部分，為找出機器（模型）所選的重要特徵，分別透過 XGBoost 和 Random Forest 模型挑選 10 個及 20 個特徵，接著根據所挑選的特徵，分別使用 XGBoost、Random Forest 和 SVM 模型進行預測，以下為所模型選特徵及其預測結果：

演算法選擇特徵 / 數值型變數「平均每個月上課人數」	
XGBoost	挑選前10 <code>['total_fans', 'owner_name_Mu/s Craft毛線球牧場', 'owner_name_Avon', 'owner_name_王梓沅 X 創勝文教', 'owner_name_數位時代', 'owner_name_Sonny Chen', 'group_title_語言', 'owner_name_郭堯中 Nkwo', 'group_subGroup_title_英文', 'group_subGroup_title_資料科學']</code>
	挑選前20 <code>['total_fans', 'owner_name_Mu/s Craft毛線球牧場', 'owner_name_Avon', 'owner_name_王梓沅 X 創勝文教', 'owner_name_數位時代', 'owner_name_Sonny Chen', 'group_title_語言', 'owner_name_郭堯中 Nkwo', 'group_subGroup_title_英文', 'group_subGroup_title_資料科學', 'owner_name_窩廚房Joy'in Kitchen', 'owner_name_鏡頭教練洪婉禎', 'pictureNum', 'teacher_group_subGroup_title_similarity', 'preOrderedPrice', 'group_title_行銷', 'lengthOfDescription', 'teacher_title_similarity', 'owner_name_凱蒂Katie', 'teacher_recommendedBackground_similarity']</code>

挑選前 10 之 RMSE :	挑選前 20 之 RMSE :
XGBoost : 2598.874062	XGBoost : 2472.721445
<b>RF : 2532.962776</b>	<b>RF : 2440.852141</b>
SVM : 2536.235786	SVM : 2537.041847

其中透過 XGBoost 去產生的重要特徵主要有「total\_fans」各個開課老師在社群媒體上的粉絲人數，以及可以看到有幾個 owner\_name（授課老師名字）的變數，可以對應到所教授的課程類別「group\_subGroup\_title」。例如，「owner\_name\_王梓沅 X 創勝文教」是教英文的，對應到的重要特徵也有課程子類別「group\_subGroup\_title\_英文」，以及課程類別「group\_title\_語言」，所以接著我們也可以看到重要特徵有出現「teacher\_group\_subGroup\_title\_similarity」，表示老師跟課程子類別的適配性也是重要特徵。此外，也可以看到有像是表示課程介紹當中使用到的圖表數量特徵「pictureNum」、「preOrderedPrice（課程優惠價）」，以及「lengthOfDescription（課程描述長度）」這些都可以列為預測一課程的「平均每月上課人數」這一項數值型變數的重要特徵。

使用 XGBoost 模型所挑選的 10 個重要特徵情況下，Random Forest 模型的表現最好（RMSE 最低），而 XGBoost 模型的表現最差（RMSE 最大）。在使用 XGBoost 模型所挑選的 20 個重要特徵情況下，同樣是 Random Forest 模型的表現最好（RMSE 最低），不過 SVM 模型的表現變為最差（RMSE 最大）。

演算法選擇特徵 / 數值型變數「平均每月上課人數」	
Random Forest	挑選前10 ['owner_name_阿滴 x 圖文不符上課囉', 'total_fans', 'owner_name_Mu/s Craft毛線球牧場', 'owner_name_Avon', 'owner_name_王梓沅 X 創勝文教', 'group_subGroup_title_手機程式開發', 'owner_name_數位時代', 'preOrderedPrice', 'owner_name_Sonny Chen', 'owner_name_高廚房Joy'in Kitchen']
	挑選前20 ['owner_name_阿滴 x 圖文不符上課囉', 'total_fans', 'owner_name_Mu/s Craft毛線球牧場', 'owner_name_Avon', 'owner_name_王梓沅 X 創勝文教', 'group_subGroup_title_手機程式開發', 'owner_name_數位時代', 'owner_name_Sonny Chen', 'preOrderedPrice', 'owner_name_莫影曦 Hailey', 'owner_name_高廚房Joy'in Kitchen', 'price', 'owner_name_鏡頭教練洪婉儀', 'teacher_metaDescription_similarity', 'teacher_title_similarity', 'pictureNum', 'totalVideoLengthInSeconds', 'courseCP', 'owner_name_郭堯中 Nkwo', 'owner_name_凱蒂Katie']

挑選前 10 之 RMSE :	挑選前 20 之 RMSE :
XGBoost : 2598.874062	XGBoost : 2472.721445
<b>RF : 2517.562171</b>	<b>RF : 2459.029466</b>
SVM : 2536.235786	SVM : 2537.041847

而透過 Random Forest 來產生的重要特徵除了也有「total\_fans」各個開課老師在社群媒體上的粉絲人數、課程介紹當中使用到的圖表數量特徵「pictureNum」、「preOrderedPrice」課程優惠價還有幾個「owner\_name」這些跟 XGBoost 產生的相似結果以外，還多新增了像是，「totalVideoLengthInSeconds（課程長度）」，還有我們產生的新特徵：「courseCP（課程 cp 值）」，是用影片時長除以課程價格得到的新變數以及「teacher\_metaDescription\_similarity」表示老師與課程文字介紹的適配性這項變數，對預測一課程的「平均每月上課人數」這一項數值型變數，上述都可被列為重要特徵。

而在使用 Random Forest 模型所挑選 10 個重要特徵情況下，Random Forest 模型的表現最好（RMSE 最低），而 XGBoost 模型的表現最差（RMSE 最大）。在使用 Random Forest 模型所挑選 20 個重要特徵情況下，同樣是 Random Forest 模型的表現最好（RMSE 最低），SVM 模型表現最差（RMSE 最大）。

### 額外作法：去除 owner\_name

在以上模型所選重要特徵中，我們發現 owner\_name（課程老師名稱）出現多次，為觀察 owner\_name 以外的重要特徵，在本專案中，我們去除了 owner\_name 特徵進行額外測試，並同樣測試不同的特徵挑法及機器（模型）所選重要特徵，使用 XGBoost、Random Forest 和 SVM 模型進行預測。以下是模型預測二元類別變數「是否為熱門課程」及預測數值型變數「平均每月上課人數」的結果：



● 人為選擇特徵 / 二元類別變數「是否為熱門課程」

人為選擇特徵 / 二元類別變數「是否為熱門課程」													
挑法	特徵選擇	XGBoost				Random Forest				SVM			
		Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
挑法 1	價格	0.48	0.53	0.55	0.54	0.51	0.56	0.61	0.58	0.55	0.57	0.76	0.65
挑法 2	課程資訊	0.56	0.60	0.65	0.62	0.58	0.63	0.67	0.64	0.61	0.63	0.69	0.66
挑法 3	授課老師	0.58	0.63	0.62	0.62	0.56	0.61	0.62	0.62	0.53	0.56	0.80	0.65
挑法 4	適配性	0.53	0.58	0.61	0.59	0.53	0.57	0.67	0.62	0.55	0.65	0.89	0.69
挑法 5	全部特徵	0.56	0.60	0.62	0.61	0.61	0.61	0.73	0.68	0.57	0.60	0.71	0.64

首先在預測二元類別變數「是否為熱門課程」中，與未去除 owner\_\_name（授課老師名字）特徵相比，去除 owner\_\_name（授課老師名字）特徵後，各模型的分類表現有些許變化。同樣在使用授課老師相關特徵的情況下，XGBoost 模型有最佳表現。而 Random Forest 模型的部分，則是在使用全部特徵的情況下，預測表現較好，準確率為 61 %。SVM 模型同樣在使用課程資訊相關特徵的情況下，表現最好，準確率為 61 %。

● 演算法選擇特徵 / 二元類別變數「是否為熱門課程」

去除 owner\_\_name（授課老師名字）特徵後，在預測數值型變數「是否為熱門課程」的部分，為找出機器（模型）所選的重要特徵，同樣分別透過 XGBoost 和 Random Forest 模型挑選 10 個及 20 個特徵，接著根據所挑選的特徵，分別使用 XGBoost、Random Forest 和 SVM 模型進行預測，以下為所模型選特徵及其預測結果：

演算法選擇特徵 / 二元類別變數「是否為熱門課程」	
XGBoost	挑選前10 ['group_subGroup_title_應用設計', 'courseNumHisty', 'group_title_語言', 'group_subGroup_title_英文', 'group_title_生活品味', 'group_title_行銷', 'preOrderedPrice', 'averageRatingHisty', 'teacher_requiredTools_similarity', 'subtitleCourseNum']
	挑選前20 ['group_subGroup_title_應用設計', 'courseNumHisty', 'group_title_語言', 'group_subGroup_title_英文', 'group_title_生活品味', 'group_title_行銷', 'preOrderedPrice', 'averageRatingHisty', 'teacher_requiredTools_similarity', 'subtitleCourseNum', 'teacher_willLearn_similarity', 'totalVideoLengthInSeconds', 'group_title_藝術', 'teacher_group_title_similarity', 'price', 'total_fans', 'pictureNum', 'teacher_title_similarity', 'teacher_targetGroup_similarity', 'lengthOfDescription']
Random Forest	挑選前10 ['total_fans', 'courseCP', 'teacher_metaDescription_similarity', 'totalVideoLengthInSeconds', 'lengthOfDescription', 'subtitleCourseNum', 'price', 'preOrderedPrice', 'teacher_title_similarity', 'teacher_recommendedBackground_similarity']
	挑選前20 ['total_fans', 'courseCP', 'teacher_metaDescription_similarity', 'totalVideoLengthInSeconds', 'lengthOfDescription', 'subtitleCourseNum', 'price', 'preOrderedPrice', 'teacher_title_similarity', 'teacher_recommendedBackground_similarity', 'teacher_willLearn_similarity', 'pictureNum', 'teacher_targetGroup_similarity', 'teacher_requiredTools_similarity', 'teacher_group_subGroup_title_similarity', 'teacher_group_title_similarity', 'averageRatingHisty', 'group_title_生活品味', 'courseNumHisty', 'group_title_語言']

與未去除 owner\_name（授課老師名字）特徵相比，去除 owner\_name（授課老師名字）特徵後，重要的特徵沒有明顯差異，在這些特徵中，同樣比較重要的前幾個是「group\_subGroup\_title\_英文」、「group\_title\_語言」、「group\_subGroup\_title\_生活品味」等這些代表課程類別的特徵，表示這些語言、生活類型的課程可能可作為預測一課程是否熱門的重要變數。接著排序的同樣是「preOrderPrice」，也就是募資時優惠的課程售價，顯示一個課程的優惠價格也可以作為判定是否熱門的重要特徵。

接下來同樣重要的包括「averageRatingHistory（老師過去課程平均評分）」、「subtitleCourseNum（子分類的課程數）」、「total\_fans（總粉絲人數）」、「price（課程價格）」、「pictureNum（課程介紹中圖片數量）」、「lengthOfDescription（課程介紹文長）」、「teacher\_targetGroup\_similarity（老師與目標受眾的相似度）」、「teacher\_wiiLearn\_similarity（老師與將會學到內容的相似度）」、「teacher\_group\_title\_similarity（老師與課程類別的相似度）」、「teacher\_title\_similarity（老師與標題的相似度）」、「courseCP（課程CP 值，為影片時長除以課程價格）」、「teacher\_metaDescription\_similarity（老師與課程介紹的相似度）」、「teacher\_recommendedBackgroud\_similarity（老師與建議知識背景的相似度）」。

與未去除 owner\_name（授課老師名字）特徵前一樣，同樣說明了課程的價格、內容多寡、課程類型、老師過去課程評價、同類型課程數量及老師與課程的適配性（相似度），是影響課程熱門的重要變數。

XGBoost		演算法選擇特徵 / 二元類別變數「是否為熱門課程」			
		Accuracy	Precision	Recall	F1
挑選前10	XGBoost	0.58	0.62	0.64	0.62
	Random Forest	0.59	0.62	0.67	0.63
	<b>SVM</b>	<b>0.62</b>	<b>0.62</b>	<b>0.83</b>	<b>0.71</b>
挑選前20	XGBoost	0.55	0.58	0.63	0.60
	<b>Random Forest</b>	<b>0.62</b>	<b>0.63</b>	<b>0.72</b>	<b>0.68</b>
	<b>SVM</b>	<b>0.62</b>	<b>0.62</b>	<b>0.81</b>	<b>0.70</b>

使用 XGBoost 挑選的重要特徵後，分類模型表現有所提升，在選擇 10 個特徵情況下，SVM 模型表現最好，準確率為 62 %，而表現最差的 XGBoost 模型準確率為 58 %。而在選擇 20 個特徵情況下，Random Forest、SVM 模型表現較好，準確率為 62 %，XGBoost 表現最差，準確率為 55 %。

Random Forest		演算法選擇特徵 / 二元類別變數「是否為熱門課程」			
		Accuracy	Precision	Recall	F1
挑選前10	XGBoost	0.54	0.58	0.66	0.61
	Random Forest	0.54	0.61	0.68	0.66
	SVM	0.52	0.55	0.82	0.66
挑選前20	XGBoost	0.58	0.61	0.65	0.63
	Random Forest	0.60	0.61	0.71	0.67
	SVM	0.62	0.61	0.85	0.71

使用 Random Forest 挑選的重要特徵後，在選擇 10 個特徵情況下，所有模型表現皆不如前者使用 XGBoost 挑選重要特徵的表現，表現最佳的 XGBoost 和 Random Forest 模型準確率僅為 54 %。而在選擇 20 個特徵情況下，SVM 模型表現最好，準確率為 62 %，XGBoost 表現最差，準確率為 58 %。

### ● 人為選擇特徵 / 數值型變數「平均每個月上課人數」

人為選擇特徵 / 數值型變數「平均每個月上課人數」				
挑法	特徵選擇	XGBoost RMSE	Random Forest RMSE	SVM RMSE
挑法 1	價格	3990.983343	2871.267629	2538.120174
挑法 2	課程資訊	3053.780622	2969.508614	2538.433224
挑法 3	授課老師	3554.281333	3161.459493	2539.256718
挑法 4	適配性	5542.011635	3896.055153	2538.332320
挑法 5	全部特徵	2703.355795	2749.237738	2538.752467

在預測數值型變數「平均每個月上課人數」的部分，去除掉 owner\_\_name（課程老師名稱）特徵後，模型的表現有所變動，XGBoost 和 Random Forest 模型變為在使用全部特徵的情況下，RMSE 較低；而 SVM 模型則同樣在使用價格相關的特徵情況下，RMSE 較低。總體來看，使用價格相關特徵及 SVM 模型，擁有最低的 RMSE、表現最好。



## ● 演算法選擇特徵 / 數值型變數「平均每個月上課人數」

在預測數值型變數「平均每個月上課人數」的部分，去除掉 owner\_\_name (課程老師名稱) 特徵後，為找出機器 (模型) 所選的重要特徵，同樣分別透過 XGBoost 和 Random Forest 模型挑選 10 個及 20 個特徵，接著根據所挑選的特徵，分別使用 XGBoost、Random Forest 和 SVM 模型進行預測，以下為所模型選特徵及其預測結果：

演算法選擇特徵 / 數值型變數「平均每個月上課人數」	
XGBoost	挑選前10 ['total_fans', 'group_subGroup_title_手作小物', 'group_subGroup_title_手機程式開發', 'price', 'group_subGroup_title_英文', 'group_subGroup_title_數位行銷', 'courseCP', 'preOrderedPrice', 'teacher_group_subGroup_title_similarity', 'pictureNum']
	挑選前20 ['total_fans', 'group_subGroup_title_手作小物', 'group_subGroup_title_手機程式開發', 'price', 'group_subGroup_title_英文', 'group_subGroup_title_數位行銷', 'courseCP', 'preOrderedPrice', 'teacher_group_subGroup_title_similarity', 'pictureNum', 'group_subGroup_title_資料科學', 'teacher_recommendedBackground_similarity', 'lengthOfDescription', 'averageRatingHistory', 'group_subGroup_title_烹飪料理與甜點', 'teacher_title_similarity', 'group_subGroup_title_程式入門', 'teacher_group_title_similarity', 'group_subGroup_title_動態攝影', 'group_subGroup_title_商業攝影']

挑選前 10 之 RMSE :	挑選前 20 之 RMSE :
XGBoost : 2932.193189	XGBoost : 2960.436781
RF : 2657.502515	RF : 2754.660815
<b>SVM : 2538.263220</b>	<b>SVM : 2538.643956</b>

其中透過 XGBoost 去產生的重要特徵主要有「total\_fans」各個開課老師在社群媒體上的粉絲人數、「courseCP (課程 cp 值)」是用影片時長除以課程價格得到的新變數、「price (課程價格)」、「preOrderedPrice (課程優惠價)」、「lengthOfDescription (課程描述長度)」，課程介紹當中使用到的圖表數量特徵「pictureNum」、「averageRatingHistory」該老師所開設的課程星數，以及各種老師與其他相關資訊的適配性，例如老師跟課程子類別、老師跟知識背景等等。這些都可以列為預測一課程的「平均每月上課人數」這一項數值型變數的重要特徵。

去除掉 owner\_\_name (課程老師名稱) 特徵後，使用 XGBoost 模型所挑選的 10 個重要特徵情況下，變為 SVM 模型的表現最好 (RMSE 最低)，而 XGBoost 模型的表現最差 (RMSE 最大)。在使用 XGBoost 模型所挑選的 20 個重要特徵情況下，同樣是 SVM 模型的表現最好 (RMSE 最低)，XGBoost 模型的表現最差 (RMSE 最大)。

演算法選擇特徵 / 數值型變數「平均每個月上課人數」	
Random Forest	挑選前10 ['total_fans', 'courseCP', 'group_subGroup_title_手機程式開發', 'group_subGroup_title_手作小物', 'preOrderedPrice', 'lengthOfDescription', 'group_subGroup_title_英文', 'totalVideoLengthInSeconds', 'price', 'teacher_willLearn_similarity']
	挑選前20 ['total_fans', 'group_subGroup_title_手機程式開發', 'courseCP', 'group_subGroup_title_手作小物', 'price', 'lengthOfDescription', 'group_subGroup_title_英文', 'teacher_targetGroup_similarity', 'preOrderedPrice', 'teacher_willLearn_similarity', 'totalVideoLengthInSeconds', 'teacher_metaDescription_similarity', 'teacher_group_title_similarity', 'teacher_title_similarity', 'teacher_recommendedBackground_similarity', 'teacher_group_subGroup_title_similarity', 'teacher_requiredTools_similarity', 'group_subGroup_title_數位行銷', 'averageRatingHistory', 'subtitleCourseNum']

挑選前 10 之 RMSE :	挑選前 20 之 RMSE :
XGBoost : 2932.193189	XGBoost : 2960.436781
RF : 2648.590312	RF : 2717.968113
<b>SVM : 2538.263220</b>	<b>SVM : 2538.643956</b>

而透過 Random Forest 大體上跟 XGBoost 生成的重要特徵差不多，對預測一課程的「平均每月上課人數」這一項數值型變數，上述都可被列為重要特徵。

去除掉 owner\_\_name（課程老師名稱）特徵後，在使用 Random Forest 模型所挑選的 10 個重要特徵情況下，同樣是 SVM 模型的表現最好（RMSE 最低），而 XGBoost 模型的表現最差（RMSE 最大）。在使用 Random Forest 模型所挑選的 20 個重要特徵情況下，同樣也是 SVM 模型的表現最好（RMSE 最低），XGBoost 模型的表現最差（RMSE 最大）。

#### 預測小結：

##### 二元類別變數「是否為熱門課程」：

使用 XGBoost 挑選 10 個特徵情況下，Random Forest 模型表現最好，準確率為 63 %

##### 數值型變數「平均每個月上課人數」：

使用 XGBoost 模型所挑選的 20 個重要特徵情況下，是 Random Forest 模型的表現最好（RMSE 最低，為 2440.852141）

##### 模型所選重要特徵：

說明了課程的價格、內容多寡、課程類型（語言、生活、程式）、老師過去課程評價、同類型課程數量及老師與課程的適配性（相似度），是影響課程熱門的重要變數

**\*修正至此頁為止**

## 五、研究成果

透過機器學習方式，預測 Hahow 線上募資課程之熱門程度(以「平均每月上課人數」及「是否熱門」兩種變數評量)，並探討熱門預測模型之重要特徵，能發現哪些關鍵變數與熱門程度相關，並能針對「平均每月上課人數」及「是否熱門」兩種變數，找出各自的最佳預測模型。此研究成果能帶了以下價值：

### 針對平台方：

1. 針對未開課課程，即可參考預測之課程熱門狀況，協助授課老師針對課程進行調整
2. 針對已開課課程，平臺也可提供研究結果給教學者進行參考
3. 可利用研究成果之課程熱門關鍵變數，協助授課老師們多開發熱門趨勢課程
4. 將預測熱門之課程放入募資倒數廣告，協助教學者加快通過募資門檻人數（開課門檻為 30 人）

### 針對授課老師：

1. 可參考研究成果，了解課程設計應調整之處
2. 若在募資開課階段，可以根據課程熱門關鍵變數調整課程設計，以增加通過開課門檻的機會

### 針對學員：

1. 授課老師設計出令多數學生感興趣、熱門的課程，可以增加學生學習動力
2. 平臺方協助開發熱門課程並進行廣告宣傳，可以讓學生接觸到更多感興趣的線上課程

## 六、參考資料與文獻

- [1] 2022 年 台灣網路調查報告
- [2] 2022 年 Hahow 品牌白皮書
- [3] 2022 年 Influencer Marketing Hub Benchmark Report，連結為 <https://influencermarketinghub.com/influencer-marketing-benchmark-report/>
- [4] 2021 年 HolonIQ 調查報告，Sizing the Global EdTech Market. Mode vs Model，連結為 <https://www.holoniq.com/notes/sizing-the-global-edtech-market>
- [5] 2021 年 台灣智慧學習產業產值調查報告
- [6] 2020 年《OpView 社群口碑資料庫》線上課程平臺聲量分析報告
- [7] 2018 年 Pearson, Beyond Millennials: The Next Generation of Learners
- [8] 2019 ~ 2022 台灣百大網紅榜單，連結分別為 <https://www.kolradar.com/billboard/2019-tw-top100-kol>、<https://www.kolradar.com/billboard/2020-tw-top100-kol>、<https://www.kolradar.com/billboard/2021-tw-top100-kol>、<https://kol.bnext.com.tw/2022/>
- [9] 林慶儒，2022，採線上募資開課之付費線上課程的再購意圖研究，國立中山大學碩士論文
- [10] 李妮燁，2020，用機器學習分析並預測線上付費課程用戶的學習行為以及高價值用戶其特徵-以 Hahow 為例，國立臺灣大學碩士論文
- [11] 鄧語華，2021，以平台生態圈探討臺灣募資型數位學習平台的中介角色與商業模式：以 Hahow 為例，國立陽明交通大學碩士論文
- [12] 鄧楷馨，2022，探討線上課程之購買意圖，國立雲林科技大學碩士論文
- [13] 蘇暘升，2022，以 SOR 理論探討線上課程平臺之購買意圖－以 Hahow 好學校為例，國立高雄科技大學碩士論文
- [14] Arie Rachmad Syulistyo, Vira Meliana Agustin, Dwi Puspitasari, Predicting News Article Popularity with Multi Layer Perceptron Algorithm, Journal of Applied Intelligent System, Vol. 7 No. 2, August 2022, pp. 193 - 205
- [15] Fauvel, S., & Yu, H., A survey on artificial intelligence and data mining for moocs, 2016, arXiv:1601.06862
- [16] Hongjun Jeon, Wonchul Seo, Eunjeong Park, Sungchul Choi, Hybrid machine learning approach for popularity prediction of newly released contents of online video streaming services, Technological Forecasting and Social Change, Volume 161, 2020, 120303, ISSN 0040-1625
- [17] Liu, Z., Zhang, Y., A Semantic Role Mining and Learning Performance Prediction Method in MOOCs, APWeb-WAIM 2018: Web and Big Data, pp 259 - 269
- [18] Zachari Swiecki, Dragan Gasevic et al., Popularity Prediction in MOOCs: A Case Study on Udemy, January 2022, Artificial Intelligence in Education, pp. 607-613