1. 緒論

APP(Application)是指應用程式的意思，原本不限於個人電腦上或是手機上的應用程式，只是在這個人手一支智慧型手機年代，APP有時特指手機應用程式，APP的型態包羅萬象，包括社群軟體、遊戲、生產性工具、音樂、教育等等，族繁不及備載。APP的存在無疑豐富了人們的生活使人們的生活更加便利。

分析APP市場是一件重要的事情，根據Paolo Roma等人的研究，APP市場在2016年就超過了460億美元的營收，而在2020年1月8日的CNN報導中也指出Ios系統的消費者在短短一個禮拜內就消費了14億美元，種種數據顯示APP市場占據全球消費領域重要地位。

而APP的收費大致可以分成三種，一種是免費型的APP，一種是買斷型的APP，最後一種則是基本免費，但進階功能要收費的APP(以手機遊戲為主，也就是俗稱的儲值或是課金行為)，第一種的營收主要來自於廣告，但不管是哪一種APP，下載量均為APP開發商最關注的一環。

智慧型手機的作業系統主要分成兩種，第一種是以Apple公司開發的iOS作業系統，另外一種則是Google公司研發的Android作業系統，前者下載APP的商店稱為APP Store而後者下載APP的商店則稱為Google Play。根據Gartner公司的報告，上述兩種的作業系統在2016年合計占了智慧型手機99%的市佔率，所以可以知道分析上述兩種平台上的APP情形，便可以大致掌握APP市場。

本研究有兩個主要的資料集，一個是Google play商店的APP資料，另一個則是APP商店的APP資料，兩者均來自Kaggle(一個全球性的數據建模和數據分析競賽平台，並於2017年被Google公司收購)，

近年來有許多文獻分析APP的市場，包含分析APP製造商的投資策略、分析何種要素會影響APP的rating等等，但較少文獻在同時分析兩種平台的APP有何差別，是故本研究將會專注在這塊分析。

本研究將會專注在以下課題：

1. iOS的APP和Android的APP在APP的標題上用字偏好是否有不同的趨勢
2. 兩個平台影響rating的關鍵字是否有不同
3. 兩個平台的定價是否有不同
4. 分析和預測Android市場的APP留言為正面還是負面

本研究的研究流程如下，首先會在第二章的文獻回顧中回顧有關APP市場相關的文獻，以及簡介使用的機器學習和計量模型方法，接著在第三章變數說明和敘述統計中，詳細說明兩個資料集的變數以及他們大致的資料分布，在第四章模型設計和實驗分析中，則會對相關問題使用適合的模型來分析，最後則在第五章的討論與建議中針對研究的不足給予建議

隨機森林:

隨機森林為Leo Breiman於2001年提出的機器學習算法，其優點為可以應對