

輔仁大學金融與國際企業學系

Linebot 介面應用於 P2P 借貸信用評分模組

指導教授： 楊雅薇博士

組員姓名： 蔡皓宇
蕭孟岳
賴易明
陳聖方
黃功裕
郭聖皿

中華民國一百零九年五月

摘要

院(系)所組別: 輔仁大學金融與國際企業學系學士班

論文題目: Linebot 介面應用於 P2P 借貸信用評分模組

學士生: 蔡皓宇 蕭孟岳 賴易明 陳聖方 黃功裕 郭聖皿

指導教授: 楊雅薇 博士

論文頁數: 49 頁

論文摘要:

此論文目標是欲建立 P2P 信用評分分析服務 API。

P2P 網貸在金融業的數位化、人工智能化浪潮下逐漸興起。過去 P2P 網貸的介面普遍以"信用評分模型+網站"方式來操作,而為了讓使用者可以更方便使用,我們選擇將其建立在手機 LINE APP 上,也就是"信用評分模型+LineBot 聊天機器人介面"的方式。並且為進一步提高便利性,我們在介面上嵌入 Google Speech 深度學習、類神經網路機器學習演算法 API (Application Programming Interface)服務,將儲存後的資料轉譯成語音,實現輸出語音化。

而為了增加信用風險預測準確率,在借貸違約風險預測技術上,我們透過不對稱樣本的處理,多種機器學習方法的導入,以及樣本分群預測等三種方式來強化。本研究使用 Lending Club 公開資料做測試資料。在使用 ROS 方法處理樣本的不對稱性後,導入多種機器學習演算法模型進行預測發現其預測準確率約為 79.9%;為了進一步提高準確率;我們發現以職業作為分群因子,將可有效提升評估借貸違約風險的準確率 5-15%。但因為 Lending Club 在收集職業資料是自由填寫,所以同一種職業可能有很多種不同表達方式,使得分群困難。為了實現以職業分群,我們只能利用人工來分群數萬筆資料,成功將資料進行切割後,再次導入模型進行二次建模訓練並預測,測試結果顯示預測準確率確實也提高不少。因此我們發現切割職業後分群,可以有效提升評估借貸違約風險時的準確率。

關鍵詞: P2P 借貸、API 串接、信用評分、AI 語音

Abstract

Title of thesis : LineBot Interface Application P2P Credit Scoring Module

Department of Finance and International Business(Bachelor of Finance)Fu Jen University

Student : Hao-Yu,Tsai , Meng-Yueh,Hsiao , Ian,Lai , Sheng-Fang,Chen ,

Kung-Yu,Huang , Samuel Solomon,Kok

Advisor : Dr.Yang,Ya-Wei

Total pages : 49

Abstract:

The goal of this study is to build a P2P credit score analysis service API.

P2P Lending is gradually emerging under the wave of digitization and artificial intelligence in the financial industry. The interface of P2P Lending in the past is generally operated with a credit scoring model on website. In order to make it more convenient to the users, we connect credit scoring model to the LINE APP. We also used Google Speech deep learning, neural network machine learning algorithm Application Programming Interface service to translate the stored data into speech. Then the translated stored data will be sent back to LineBot output as a voice file.

We enhance predicted accuracy via data imbalance technique, Multiple Machine Learning algorithms, and Sample clustering to conduct our analysis using data from Lending Club. Our predicted accuracy rate can increase on average 15% from 79.9%. However, the collected information of Lending Club can be filled in freely by the applicant so there were various expressions about the occupation and this gave us some difficulties when it came to grouping the data. So we can only group tens of thousands of data by reviewing them one by one by ourselves. The data were being imported into the model again for the second prediction after we successfully separated the data. The test result shows that the prediction accuracy increased successfully. Therefore, we found that grouping different jobs can effectively improve the accuracy of predicting the risk of loan defaults.

Key words : P2P Lending, API Connection, Credit Score,AI Voice

誌謝辭

「時間就像一張網，撒在哪收穫就在哪」，在這次的專題報告中我們完全體會到這其中的奧妙。一開始的我們都還懵懂無知，尤其對 P2P 這個領域幾乎不認識，完全不知道到底該從何著手，甚至不知道現在金融業到底發展如何，但在我們的指導老師楊雅薇老師的細心指導下，我們也逐漸成長茁壯，讓這次的專題報告慢慢步上軌道，也讓我們可以更加確信自己的目標，以及該為此作何種準備。

我們在這一路上跌跌撞撞，但老師在我們每次跌倒遭遇挫折時都會伸出援手，不厭其煩地解決我們的問題，尤其在我們跑預測模型及開發語音介面的時候，因為成果一直沒有很好，我們也數次想要放棄，但每次在我們想要放棄時，老師便會給予我們勉勵，讓我們繼續堅持下去，並告訴我們若可以撐過去，最後必可以享用到最甜美的果實，也讓我們可以更加清楚自己的價值，彷彿船在夜間靠港找不到方向時的照明燈指引著我們，因此這次的專題報告楊雅薇老師真的功不可沒，可以說若是沒有老師的幫助，我們就不會有如此好的表現。

再來就是要謝謝我們的父母，很感謝父母在我們做選擇時可以相信我們，當我們在因為專題而煩惱時，父母願意傾聽我們並支持我們的決定，而且每當我們因為討論專題到很晚才回家時，打開門後還可以看到他們坐在客廳等待我們的歸來，讓我們可以沒有後顧之憂地做我們真正想做的事情，因此在此要特別感謝我們的父母，也十分感謝一路上幫助過我們的人，若沒有你們這段日子的照顧，也不會有今天的我們。

金融與國際企業學系學士班

蔡皓宇 蕭孟岳 賴易明 陳聖方 黃功裕 郭聖皿

中華民國 109 年 5 月

目錄

摘要	I
ABSTRACT	II
誌謝辭	III
目錄	IV
表目錄	V
圖目錄	VI
第一章 緒論	1
第一節 研究背景與動機	1
第二節 研究目的	4
第三節 研究貢獻	5
第四節 研究架構	6
第二章 文獻探討	7
第一節 P2P 相關文獻	7
第二節 LINEBOT AI 界面相關文獻	9
第三節 機器學習相關文獻	10
第四節 職業分群相關文獻	11
第五節 SWOT--P2P 借貸於 LINE 聊天機器人之應用	12
第三章 研究流程	13
第一節 P2P 資料輸入	15
第二節 機器學習信用評分分析	16
第三節 語音輸出	26
第四節 生活資訊	27
第四章 研究成果與應用	28
第五章 結論與建議	46
第一節 結論	46
第二節 建議	47
參考文獻	48

表目錄

表 3-1 整體資料基本統計量	16
表 3-2 違約資料之基本統計量	17
表 3-3 不違約基本統計量	17
表 3-4 模型預測力	20
表 3-5 不同模型，分群後預測力以及錯放率比較	21

圖目錄

圖 1-1 研究架構.....	6
圖 2-1 LINEBOT 介面關係圖.....	9
圖 3-1 簡易流程 1.....	13
圖 3-2 簡易流程 2.....	13
圖 3-3 簡易流程 3.....	14
圖 3-4 完整流程.....	14
圖 3-5 混淆矩陣.....	25
圖 4-1 平台功能.....	28
圖 4-2 使用說明.....	29
圖 4-3 我要借貸.....	30
圖 4-4 借貸申請頁面.....	31
圖 4-5 申請成功畫面.....	32
圖 4-6 申請失敗畫面.....	33
圖 4-7 我要取消.....	34
圖 4-8 成功取消.....	35
圖 4-9 誤觸取消.....	36
圖 4-10 關於我們.....	37
圖 4-11 位置資訊.....	38
圖 4-12 聯絡我們.....	39
圖 4-13 天氣查詢.....	40
圖 4-14 匯率查詢.....	41
圖 4-15 股價查詢.....	42
圖 4-16 用餐建議.....	43
圖 4-17 SERVER 資料庫.....	44
圖 4-18 變更資料庫內容.....	45

第一章 緒論

第一節 研究背景與動機

一、研究背景--Open API 與 P2P 網貸

Open Banking 於 2018 年從英國開始，其用意是讓金融科技服務新創公司能連接銀行資料庫進行身分驗證以縮短流程時間，也藉此更了解消費者需求。然而近期台灣金融業第一階段的 Open Banking 應用多集中於生活繳費、記帳 APP、身分認證、大數據分析等，較少應用於借貸領域，因此我們期望將 **Open API 的概念帶入信用評等服務**。

過去五到十年間，金融業的數位化、人工智能化浪潮逐漸興起，借貸領域中，則是衍生出了 P2P(Peer to Peer)網貸的新模式。如全球最大的 Lending Club P2P 網路借貸平台。相較於傳統銀行放貸，P2P 網貸模式能夠興起主要有兩層面的原因：從需求面來看，以 80/20 法則來解釋，傳統金融業貸放的客群多集中在信用狀況較良好的族群，另外沒服務到的 80%，便可作為 P2P 借貸的客源。從供給面來說，P2P 借貸者多為信用小白(如學生、工人、自由業者無固定薪資、較少與銀行往來者)，這對於信用評分有者極大困難。不過當人工智能與機器學習逐漸在各個跨領域應用中引發關鍵性突破，我們發現**機器學習演算法正是 P2P 網路借貸信用評分的關鍵**。此外，網貸模式中因多了對社群資料的利用，使商品服務有了彈性化、客製化的可能，也增加滲透市場的力度，進而服務到更多的族群以開發客源。

二、研究動機--信用評分服務：Open API

CRM(Customer Relationship Management)是由 Salesforce 提供的客戶關係管理服務，讓客戶有更彈性的介面快速自訂想要的表單，開創了 SaaS(Software as a Service)商業模式。科技巨頭如 Google 提供 Map API 服務讓第三方能輕易的串接 Google Map 使用其龐大的資料並嵌入第三方自己的網路中；Amazon 提供 AWS(Amazon Web Services)雲端伺服器代管服務，讓新創公司能快速建立 IT infrastucture 藉此降低初創期技術投入成本。而我們期望最終能提供**信用評等 API 服務**。

三、研究動機--便利體驗：LineBot 結合信評模型及 AI 語音

台灣借貸市場中已有 P2P 網貸如最有名的是鄉民貸。他利用網站申請方式將借貸流程簡化和加速，解除了申請時間、地點的限制，讓服務相較於傳統銀行更為便利。

不過 P2P 網貸崛起的年代是賈伯斯尚未發表 iPhone 的年代。十年後的現代人生活離不開手機，所以 App 或許是不錯的選擇，如同普匯 Influx 他們全心投入在

App 媒介。然而我們細思後認為，多數人手機中大量的 App 卻只有少數是經常使用，而最常使用的即社群媒體 App。因此我們希望能有比 App 更為便利的媒介，便轉向聊聊天機器人去探索，我們發現在台灣應用 App 的專業金融服務多集中在繳費支付以及投資領域，較少應用在借貸領域。至於聊天機器人之選擇，則因為 Line 是台灣幾乎人人使用的重要通訊 App，因此 **LineBot** 聊天機器人很自然成為我們 P2P 借貸媒介首選。

為了實現 P2P LineBot 聊天機器人借貸優於 P2P 網路借貸的便利性，我們原認為應只需簡單的將"信用評分模型+網站"改成"信用評分模型+LineBot 聊天機器人介面"即可。不過實際建構時發現還必須架設伺服器容納模型，才能讓模型與 LineBot 之間搭起溝通的橋樑。而**建構出信評分析模型並串接上便利的使用介面**正是本研究最困難也是最主要的目的。

針對便利性，除了利用 Line 在台灣普遍使用的優勢及聊天機器人簡化借貸流程和解除時間地點限制，為了能更接近實質意義上的"聊天"，讓繁忙如貨車司機也能在駕駛過程中以語音方式完成手續，我們也嘗試加入了 **AI 語音技術**來增加使用的便利性。

四、研究動機--信評模型：機器學習演算法及模型再優化

如前所述，P2P 借貸的客群多數是信用小白，而這些人可能缺乏聯徵資料。人工智能、機器學習演算法的出現，為我們提供了解決資訊不對稱問題方法。本研究將 Lending Club 資料利用機器學習方法分析並建立信用評分模型。**機器學習演算簡單來說**，就是將資料標上各種特徵標籤和一個結果標籤，來把每一筆資料向量化，接著將所有向量切割成訓練集和測試集兩部分，讓訓練集內的向量用演算法推算其可能關係並與測試集對照正確性，接著極大量地反覆訓練和對照，以自動生成一個對此種資料的理解。機器學習演算法的好處是能夠對大量且乾淨的非線性資料非常逼近真實情況地去模擬並預測出其他類似資料可能的特性，而這是傳統統計方法無法做到的。

為了提高模型預測準確性和降低錯放率，本研究使用以下方法：

(一) 樣本資料中，不違約的數量遠高於違約的數量，這讓模型訓練時會產生偏誤。因此我們利用 ROS(Rodom Over Sample)隨機過採樣處理樣本不對稱問題。

(二) 不同演算法有各自的架設以多個演算法訓練多個預測模型，比較並選用準確率高且錯放率低者：我們使用了 KNN、Random Forest、XGBoost 三種機器學習演算法。

(三) 參考馬若雅(2019)的結論，以公開資料變數(如:人口結構、犯罪類型情況等)將樣本資料分群來進行預測能夠提高預測準確率。所以我們以職業為分類，針對不同類型的職業各自訓練專為該類型職業設計的信用模型，以藉此提高預測準確率。

為**進一步提高模型預測準確性**，除了上述我們針對**分析方法**嘗試的兩種優化預測力之方案，我們也希望能從更多角度了解借貸者的**行為特徵**來藉此更完整的捕捉其信用風險。傳統信用分析使用的主要是結構性資料，而藉著聊天機器人的優勢，我們嘗試讓機器人有更多關於生活面向的服務(如天氣服務、股價及匯率查詢等)以同時藉此累積關於借貸人更多面向的非結構性資料(如社群貼文、影像、聲音)，期望能由提供不同生活面向的服務，嘗試挖掘更多新的信用影響因子。

第二節 研究目的

為了將服務放在更便利的媒介，將模型與 LineBot 串連是我們主要研究目的之一。由於 LineBot 在網頁版開發環境僅能提供簡單對話和簡單服務，所以我們無法直接在網頁上進行如信用評分模型這種複雜的開發。因此我們需要利用到 LineBot 提供的 API 和自己開發的程式進行串接，並且於網際網路中架設一個伺服器，才能實現使用者端於自己的行動裝置上透過全球網際網路使用我們的信評分析服務。

另一研究目的為：我們嘗試讓機器人與我們對話，因此我們利用深度學習、類神經網路演算法將體驗過程語音化。

最後一研究目的則是將信用評分分析模型進行準確率提高及錯放率降低的優化。我們透過樣本不平衡的處理、多種機器學習演算法模型及職業變數對資料分群，等三種方式。

第三節 研究貢獻

本研究實作最終結果，達成以下貢獻：

1. 設計並開發一個 P2P 信用評分聊天機器人介面雛形，並利用 AI 語音讓體驗更為便利。
2. 建立完整的流程將信用分析模型與介面串接。
3. 利用不平衡樣本處理、多模型測試、職業變數分群成功提升預測準確率和降低錯誤放貸率。

第四節 研究架構

本研究之研究架構如下，第壹章為緒論，介紹本研究的動機及目的；第貳章文獻探討，回顧所參考的文獻對本研究的影響；第參章是研究流程，解釋本研究所使用之樣本、變數及方法；第肆章為使用者介面，將會呈現本研究所開發介面，最後第伍章是結論與建議。

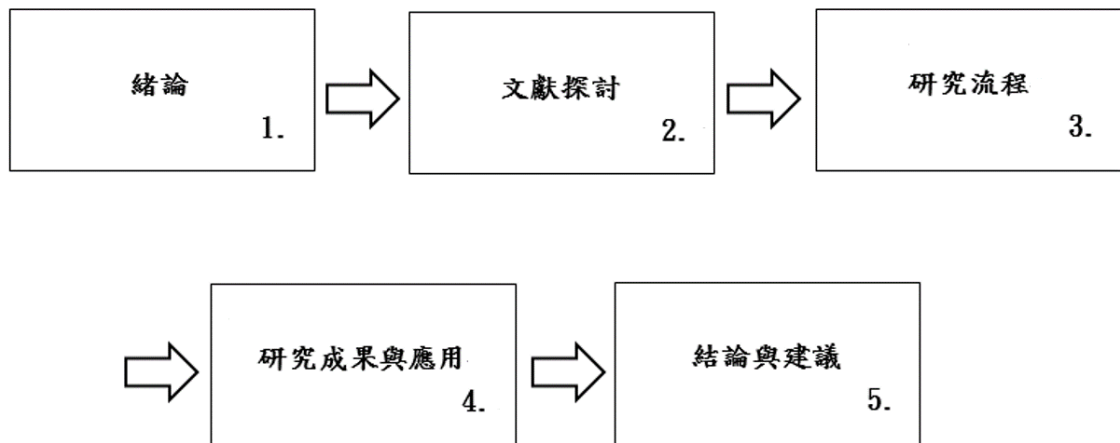


圖 1-1 研究架構

第二章 文獻探討

第一節 P2P 相關文獻

參考央行理監事會後記者會（2018）有提到說 P2P 借貸可以幫助那些經濟弱勢或急需資金的人得到貸款，因為他們無法短時間內或完全不通過傳統金融機構放貸的審核標準。除此之外，P2P 借貸還對社會創造了 3 個效益。第一，**P2P 借貸促進普惠金融**。在一些發展中或新興市場的國家例如：中國大陸、巴西和印度。這些國家地域廣闊，傳統金融機構設施不足造成中小企業甚至是普通人民都無法輕易借貸，P2P 借貸的出現正好可以補上這個需求，提供他們便利且快速的貸款方法。第二，**P2P 借貸增進資金使用效率**。因 P2P 借貸是利用網路的平臺，非但沒有實體通路，還可以大量減少人事和設備的成本。這也是 P2P 平臺吸引投資者的地方且因為運營平臺的資金較少可以直接使資金利益最大化，投資者的收益也較高。第三，**P2P 借貸發展了新的商業信用模式**。P2P 借貸平臺通常都會開創出有別於傳統金融機構的信用評分系統，它們會使用先進的科技取得和分析借款者的資料例如美國的 Lending Club 平臺。

P2P 借貸（Peer to Peer Lending）的出現改變了許多人借貸的方式。它是一種新的模式也屬於科技金融的一部分（Financial Technology）。P2P 借貸還提供給許多被傳統金融機構拒絕借貸給他們的人一個機會獲得資金；一個資料透明的投資平臺給投資者。P2P 借貸流程十分快速且簡單，比一般傳統金融機構有相當大的優勢。

P2P 借貸發展迅速除了有所述的 3 大效益以外，其實最大的原因是它們願意提供資金給傳統金融機構所拒絕的資金需求者，這也代表著它們的投資者所面臨的風險是較高的。因此 P2P 借貸平臺有責任去分析借款者的訊息，務必排除投資者和借款者出現**信息不對稱的問題**。

廖理、張偉強（2017）有說明不論是傳統金融機構或是 P2P 借貸的借款者，他們都會有意地隱瞞對其不利的訊息，尤其 P2P 借貸平臺遇到這類型的客戶機率更高。因此他們說明 P2P 平臺除了需要瞭解借款者的目的、經濟實力和債務等第三方認證的訊息之外，還需要瞭解借款人的**人口統計學特徵**（性別、年齡、教育程度、婚姻、是否有孩子）和一些**非認證訊息**（例如朋友圈、社群和網路社交賬號活躍度）以降低信息不對稱問題。

解決了投資者和借款者訊息不對稱的問題後，P2P 借貸絕對能為市場帶來非常大的好處。根據台灣的聰明貸款借貸平臺知識篇就有提到 P2P 借貸在台灣市場的潛力。雖然台灣的銀行眾多，但借款門檻非常高，只有 3 成的借款申請者成功取得貸款，這是因為銀行不想冒太大的風險。被拒絕的借款者往往本身沒有信用問題，但因為傳統金融機構的規則和政策使得許多人無法獲得貸款。而被銀行拒絕借款的人通常會再向其他銀行申請貸款，但因為短時間內聯徵信用記錄查詢的次數過多會被傳統金融機構視為信用不佳，這些人被稱為“弱勢貸款族群”。

最後，這些資金需求者可能只能尋找代辦業者或是地下錢莊。但往往這些代辦業者的訊息和利率非常不透明，導致借款者所面對的風險較高。因此，**這些“弱勢借貸族群”可以成為台灣 P2P 借貸平臺的服務客群**。讓借款人提供齊全的自身相關資訊，媒合提供資金的投資者，達到雙方利益最大化。因此我們也選擇了 P2P 借貸作為研究的主題。

第二節 Linebot AI 界面相關文獻

在這個科技發展蓬勃的時代，手機已成為每個人不可缺少的必需品。而手機除了基本的通訊的功能之外，手機和手機程序開發商也為手機增添愈來愈多的功能。當談到手機程序時當然不能不提 Line 這個網絡社交程序，尤其 Line 在台灣市場占據了非常大的市占率。現今的 Line 除了提供用戶基本的通訊服務之外還提供其他功能例如：Line 支付、利用 Line 架行動官網、宣傳頁面、調查功能、統計資訊等。

在林義旻（2017）有提到 Line 除了上述的功能以外，用戶還能利用 Line 上的開發平臺 Message API 來設計客服機器人。而 Line 平臺所提供的這個技術帶來了非常大的便利性以讓人們可以開發更完善的**客服機器人**。Messaging API 的優勢在於它是 Line 所提供給所有人的免費資源，而藉由此所開發的 Line 客服機器人可以在 Line 24 小時地運行。Line 客服機器人的形成減少了公司端許多人力的成本和時間並且可以為客戶端有效提供資訊化的服務。

我們會把關注放在設計人員端和客戶端因為這次研究裏並沒有公司端的需求。設計人員負責程序的撰寫、開發速度和功能完整性。而客戶端的所關注的問題是服務品質、回復的速度和功能性，所以我們使用機器學習分析 P2P 借貸客戶的資料後結合 Line 的 Message API 開發的客服機器人來服務客戶也符合了這個時代的趨勢。

在吳淑美（2018）也有提到新科技已經衝擊了金融 6 大產業——存貨、支付、市場資訊提供、保險、籌資和投資管理，尤其是改變了這些產業的服務流程。利用新科技結合金融可以提供更專業、優質的諮詢服務，讓客戶擁有更良好的服務體驗。例如：英國巴克萊銀行和富國銀行運用人臉辨識和語音的生物辨識來驗證企業用戶、高盛投資銀行在 app 中使用機器學習來自動化企業系統管理、西班牙國際銀行對電話錄音進行情感分析、星展銀行使用機器人審核客戶聊天記錄來提升溝通品質與設計客服機器人回答客戶問題和西班牙對外銀行開發網頁上的虛擬助理提供其客戶理財建議。

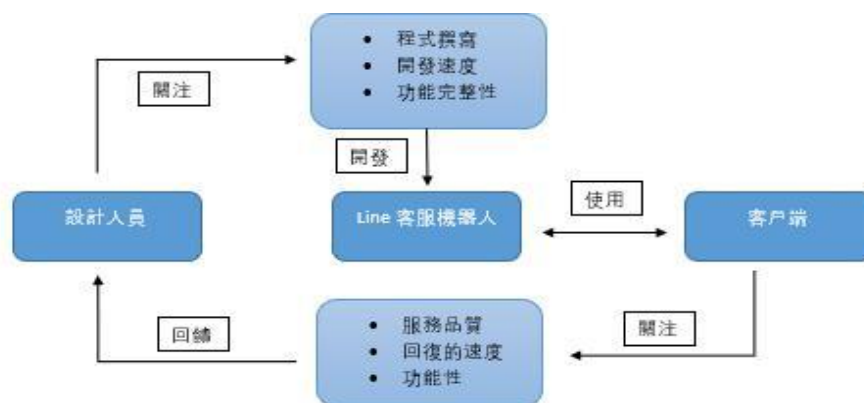


圖 2-1 LineBot 介面關係圖

第三節 機器學習相關文獻

由於近幾年的電腦和網際網路的高度普及和發展，使得許多新科技如機器人學習（Machine Learning）與人工智能（Artificial Intelligence）在各個產業廣泛地使用。

機器學習是一種人工智能的技術，有別於傳統程序，機器學習是通過處理和學習數據後，利用機率、統計等數學原理來解決問題。但相較於統計，機器學習較注重預測結果，而不是解釋變數間的關聯。所以當新的數據出現時，機器學習能透過新的資料來更新其學習模式和改變對於問題的認知。

金融領域更是順著趨勢利用這些新技術來進行許多演算法的運算分析數據和資料。

金融科技（Financial Technology）這一詞的出現頻率屢見不鮮就是因為金融領域結合了機器人學習與人工智能。這可以幫助人們分析複雜的資訊、業務自動化例如風險管控、貸款服務的諮詢與啟動和與客戶互動。

在 Bart (2017)對機器學習在金融領域的應用研究就有提到機器學習在金融領域的引用已經是非常普通的事。一般能收集的資料是非線性，而機器學習可以分析並預測非線性資料，機器在學習過程中利用各種特徵變數的訊息與預測目標變數進行整合，經由訓練多次後，只要有一筆新資料進來即可預測其是屬於哪一類，因此是屬於樣本外(out-of-sample)預測。機器學習可分為非監督式學習(Unsupervised)與監督式學習(Supervised)，非監督式學習是透過不同的特徵變數且沒有目標變數進行訓練模型，目標是將一筆資料進行分群，觀察資料可以分為幾群；監督式學習是由特徵變數與目標變數的資料數據進行訓練模型，以達到預測的目的。非監督式學習主要用於揭露一大群凌亂數據的情況，並進行深入分析，而監督式學習透過不同的分類模型(決策數、支持向量機)進行資料分類，並找出在違約情況發生時的關鍵違約變數以改善預測風險的能力。**本研究的重心在於監督式學習，主要是想探討那些重要的特徵變數會影響違約與否結果。**

本研究使用了監督式學習裡的 KNN、隨機森林、XGBoost 來分析 P2P 借貸平臺 Lending Club 的客戶數據來預測未來客戶違約與否的準確度。相信利用機器學習分析 P2P 客戶的數據將會使預測水準提高許多。

在李振華（2019）就有提到機器學習屬於人工智慧技術價值鏈的核心技術層，負責以各種數量方法、統計模型與仿生物模擬等演算法提供演算法調校和建模等。而 P2P 借貸平臺透過機器學習分析數據可以發展出一個獨有的信用評價機制，這除了可以幫助 P2P 借貸平臺篩選客戶以外，還能有效的降低風險。

第四節 職業分群相關文獻

從古至今，職業和工作往往都代表著一個人的信用和財務能力。從傳統金融機構來看，良好的信用和財力可以幫助人們快速地通過和取得貸款。雖然我們的研究有提到要幫助“貸款弱勢族群”，讓他們也有可以取得貸款的空間。但申請者也需要通過我們設計的信用審核機制。而在這信用審核機制，職業也是我們一個非常重要的參考變數。

在 Thi Huyen et al.(2007) 就有表示職業在銀行貸款是非常重要的評估要點。當一個人擁有固定收入，他得到貸款的機率就越高。但這也表示沒有制式財力證明的人例如學生、攤販和店商很難通過銀行的審核機制。Benjamin Kojo Otoo(2012)就有提到他認為攤販的收入並不固定和其他薪水不穩定者將在銀行貸款受到很多阻礙。

Jonathan D.Glater(2008) 更是提到學生在取得貸款是多麼困難的事。他認為大部分學生本身已經有背負學生貸款的負債，這使得他們在這前提下獲得新的貸款在傳統金融領域裏是一個不可能的事情。

一個人擁有越高收入的職業，往往他們違約的機率也就越低，就如 Wang et al. (2017)所指出職業與貸款違約有最顯著的關係，因此在信用評分中職業類別的占比應該更重要。因此我們設計的信用審核機制中也利用了 Lending Club 所提供的資料去做職業分群，成為審核申請者信用一個非常重要的變數。

除了職業分群，就如我們第一節所提到參考申請者的一些統計數據如性別、年齡、教育程度、婚姻、是否有孩子等和一些非認證訊息例如朋友圈、社群和網路社交賬號活躍度，也會作為申請者信用審核的變數以達到最完善和準確的分析。

第五節 SWOT--P2P 借貸於 Line 聊天機器人之應用

其實 P2P 借貸在台灣發展已經有好幾年的時間，在我們做這項研究時參考了在台灣的一些 P2P 借貸網站，例如鄉民貸和 imb 借貸媒合平臺。我們發現這些 P2P 借貸平臺都較少在手機的應用上有任何的發展，因為他們多數都已經有自己的網站。而我們發現唯一有把 P2P 借貸應用在手機 app 的只有普匯 Influx，他們將操作的重心都放在手機上，借貸和投資都是透過 app 完成。

這裡我們列出 P2P 借貸在 Line 客服機器人之應用的強弱危機分析(SWOT)：

首先我們的**優勢 (Strengths)**是我們運用了許多技術包括機器學習和多種演算法如 KNN、隨機森林 和 XGBoost，並利用職業分群來分析客戶的資訊，更加準確地判斷客戶是否符合借款的資格，我們最高可以達到 96% 準確率及 0.6% 錯放率。第二個優勢是我們使用 Line 機器人和獨特的語音系統來服務客戶。截至 2019 年 10 月的統計數據，Line 目前在台灣的滲透率有九成以上，擁有活躍用戶人數是 2100 萬。這龐大的用戶都可能成為我們 P2P 借貸的服務客群。第三個優勢則是我們的機器人架設在 Line 上，操作方便且簡易，用戶只要點開我們的 Line 客服機器人，服務項目一目瞭然，而語音系統更是讓客戶有更好的使用體驗。

我們的 P2P 借貸 Line 客服機器人的**劣勢 (Weaknesses)**是我們開發客服的時間比起其他同業的客服較短，因此客服可能存在少許瑕疵或是滿足不了所有用戶的需求。第二個劣勢是客群的上限。雖然應用在 Line 的關係可以接觸龐大的 Line 用戶，但是我們也間接的限制了我們 P2P 借貸客服的高度，如果未來我們在台灣市場的客戶已經達到飽和，會很難把這款 P2P 借貸 Line 客服機器人發展到其他沒有使用 Line 的地方。

接下來 P2P 借貸 Line 客服機器人的**機會 (Opportunities)**是目前在 Line 上並沒有跟我們一樣關於 P2P 借貸的客服機器人。這代表著我們具有先占優勢，可以首先把這種服務推廣給廣大的 Line 用戶。第二個機會就是我們有跟著金融科技和時代的趨勢，我們結合了機器學習分析 P2P 借貸數據和 Line 客服機器人，加上人們對於手機和 Line 的黏著度，這將會是一個極大的機會。第三個機會則是因為近期新冠病毒的肆虐，所有人都大量地減少了與他人直接接觸的次數，這也將會成為 P2P 借貸和網絡 APP 發展的契機。

最後，我們 P2P 借貸 Line 客服機器人的**威脅 (Threats)**，是已經出現類似的借貸 API 服務，例如凱基銀行提供 API 給 iCHEF 去針對餐廳提供借貸服務。第二個威脅是未來可能會出現跟我們相同性質或是比我們更好的 P2P 借貸 Line 客服機器人。雖然我們占據先占優勢，但未來可能會出現其他競爭對手開發相同的 Line 客服機器人來搶占市場。而談到整個產業的威脅是產業未來的不確定性。P2P 借貸或許是現今的趨勢，但無法確定未來會不會有新的金融服務可以直接取代或替代 P2P 借貸。

第三章 研究流程

由於我們研究目標是欲建立一個借貸違約風險預測模型，並為模型包裝一個使用者介面，於是我們先將整體流程簡單分為：輸入、預測判別及輸出。接著，我們設想如何的包裝呈現方式是最能為現代人快速接受。原先欲使用實體機器人 Zenbo 呈現。不過從使用者觀點並不會帶來便利，只有帶來新鮮感；而從借貸平台提供者角度來說，實體機器人的方式不但成本過於高，且無法多人同時使用一台機器，並不符合經濟效益，因此我們轉向尋求其他呈現方式。

Lending Club 是本研究公開資料來源，同時也是 P2P 借貸平台中的佼佼者。於是原先欲仿效起架設網站的方式作為我們的介面包裝方式，但隨著科技進步演進，我們生活離不開的智慧行動裝置中真正使用的是一個又一個的 App，因此 App 開發或許是最直覺的包裝方式。但我們認為多數人手機所安裝的眾多 App 中，經常使用的只有**社群網路通訊媒體 App**。許多商家會利用聊天機器人即時回覆一些預設的簡單回覆來加速或自動化其服務。這成為我們的靈感來源，令**聊天機器人**不僅只會聊天，更能提供像借貸這種專業化服務。

簡單來說：

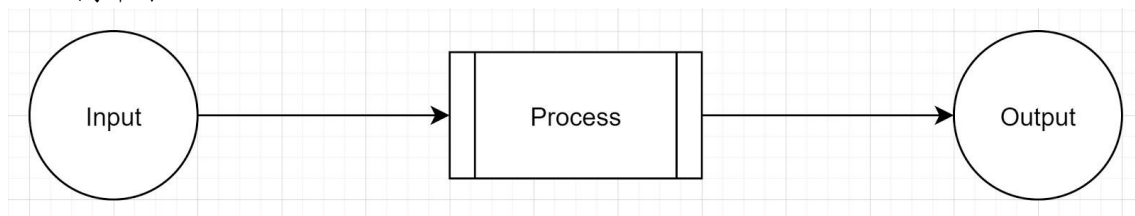


圖 3-1 簡易流程 1

以為的串聯：

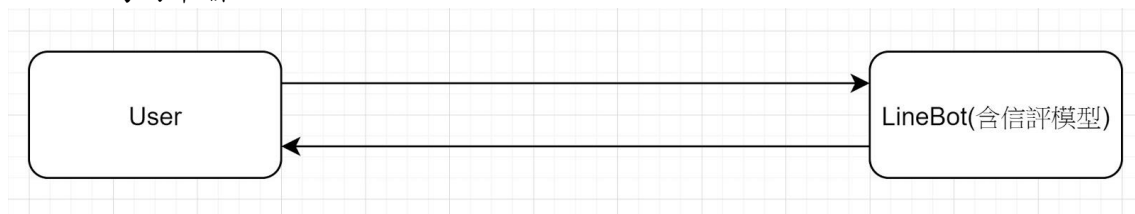


圖 3-2 簡易流程 2

實際需要的串聯：

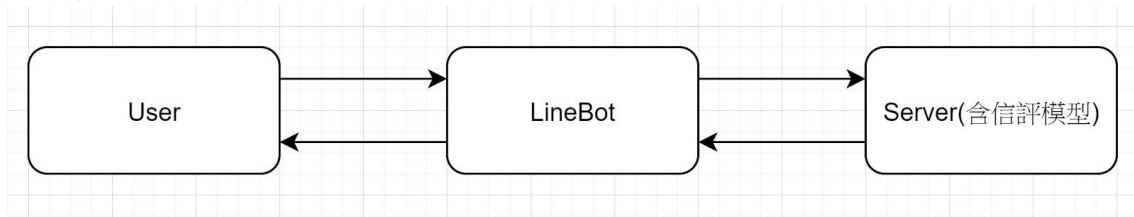


圖 3-3 簡易流程 3

server 的內涵：

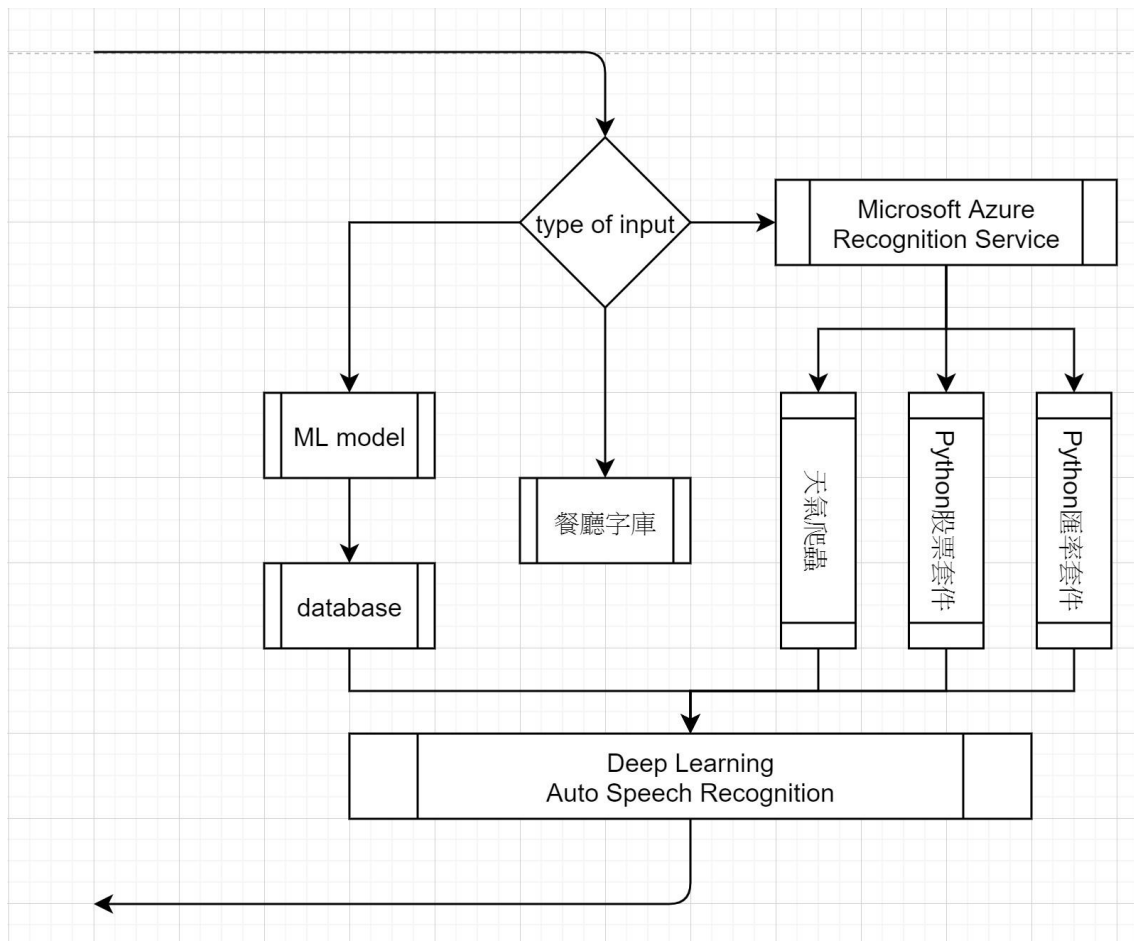


圖 3-4 完整流程

第一節 P2P 資料輸入

一、LineBot 與信用評等分析模型之串聯

為了利用聊天機器人實現我們的信用評分分析，我們須先向借貸者取得個人資料，再將資料輸入到模型之中。不過 **LineBot** 僅是單純的使用者介面，也就是說只有外表而無內在，只能簡單的對話但無法讓我們的信用評分模型直接放在 LineBot 中做分析處理。(如圖 3-1)

因此除了建構核心的機器學習違約預測模型程式本身之外，仍須架設一個伺服器容納模型，模型才能夠透過伺服器與 **LineBot** 溝通。(如圖 3-2)

而伺服器中容納(1)違約預測模型程式、(2)儲存使用者資料之資料庫、以及(3)生活資訊相關的程式等，再套上 LineBot 外衣，才能實現使用者端於自己的行動裝置上透過全球網際網路使用我們的信評分析服務。(如圖 3-4)

但我們在研究過程遇到的最大問題，正是伺服器的架設並非如此容易。需要有一台固定的電腦主機，運作著專門的作業系統來管理伺服器內部與外部之任何溝通、需要申請購買一個網域讓伺服器擁有一個固定的門牌號碼，才能順利在網路上與外界聯絡、需要有保護伺服器不受外網隨意訪問的防火牆……等。幸運的是我們身處在對於架設伺服器有大量需求的時代，因此有專門提供架設伺服器服務的公司幫助我們開發。本研究使用 **Ngrok** 將置於內網的應用程式和資料庫包裝成一個能對外網溝通的伺服器，並將伺服器與 **LineBot** 服務串聯，搭起 **Line** 聊天機器人與應用程式(違約預測模型程式、生活資訊程式)間的橋梁。

二、資料輸入

任何借貸之決定，首先當然是取得申請人的資料，以有審核的基礎。對於審核資料我們採用邱敬桓(2019)之研究成果做為變數選擇的依據。其中包含：欲貸金額、收入、負債、房屋的持有(自有、租、抵押)、帳上餘額和現金卡額度比例、總信用額度。

而為了確保伺服器後端預測模型應用程式的運行無誤，使用者所輸入的申請資料型式必須是違約預測模型能夠理解的。我們的解決方式是使用 **LIFF(Line Front-end Framework)** 的特殊 HTML 語言，寫了一個讀取資料的表單 WEB APP 讓使用者清楚的按照表單填寫，再將表單資料由 **LIFF WEB APP** 轉換為我們設計的特定格式，並在特定格式中加入作為辨別用的特殊字元，以正確的呼叫出伺服器中的違約預測模型並將特定格式的資料回傳給預測模型程式。

第二節 機器學習信用評分分析

在機器學習方面，主要分成四個大步驟：

1. 資料選取 2. 資料預處理 3. 使用不同機器學習模型 4. 以職業變數對資料做切割。

一、資料選取

本專題的資料來源是知名借貸平台 Lending Club 的資料，我們挑選 2016 年的資料。在資料預處理的部分，先將尚未到期之借貸狀況予以刪除，在一兩百個變數中為了減去搜尋重要變數之時間成本，我們參考邱敬桓(2019)，此研究利用了隨機森林、極限梯度提升、支持向量機以及羅吉斯回歸四個機器學習模型進行特徵的篩選，將其結果進行交叉比對，最後剩下個特徵包括：bc_util：帳上餘額和現金卡額度的比例、dti：負債收入比、funded_amnt：貸款金額、grade：評級、insrallment：每月付款金額、tot_hi_cred_lim：總信用額度、home_ownership：房屋持有狀況。

基本統計量：

本專題總樣本數經過整理後所採用後共有 80,526 筆，違約筆數為 15,405 筆，不違約比數為 65,121 筆。

表 3-1 整體資料基本統計量

	count	mean	std
funded_amnt	80,526	13,866.050	8,846.460
installment	80,526	434.380	279.250
grade	80,526	2.750	1.180
dti	80,526	17.800	8.100
bc_util	80,526	58.090	28.170
tot_cred_lim	80,526	172,434.000	174,016.400
mortgage	80,526	0.470	0.500
own	80,526	0.120	0.320
rent	80,526	0.420	0.500

表 3-2 違約資料之基本統計量

	count	mean	std
funded_amnt	15,405	15616.600	8,909.330
installment	15,405	48,248.000	280.550
grade	15,405	3.369	1.260
dti	15,405	19.150	8.040
bc_util	15,405	60.870	27.340
tot_cred_lim	15,405	151,143.100	152,615.200
mortgage	15,405	0.410	0.500
own	15,405	0.110	0.320
rent	15,405	0.470	0.500

表 3-3 不違約基本統計量

	count	mean	std
funded_amnt	65,121	13,414.800	8,773.840
installment	65,121	421.930	277.560
grade	65,121	2.590	1.100
dti	65,121	17.450	8.070
bc_util	65,121	57.370	28.330
tot_cred_lim	65,121	177,922.000	178,709.600
mortgage	65,121	0.480	0.500
own	65,121	0.120	0.320
rent	65,121	0.400	0.500

二、資料預處理

(一) 遺失值處理:

因為 Lending Club 的資料多屬使用者自行填入，而非經過縝密檢查及整理，因此在資料中會出現一些空值，使得在執行機器學習時會有程式無法執行的問題，因此空值必須補上或是刪除，在整個資料裡遺失值個數未超過總樣本百分之五十，因此本研究採取以**平均數填值的方式處理空值**。

(二) 資料切割:

在機器學習的過程中，需要將資料切分為訓練資料以及測試資料，本專題依照 3：7 之切分法來進行機器學習，**30% 預測，70% 訓練**，將訓練資料進行建模後，再用測試資料去進行預測來分析模型的準確率。

(三) 資料標準化:

資料集中每個特徵的數值大小差異過大，為了避免在預測中被數值較大的特徵影響預測結果，本研究採取 **Z-score 進行標準化**，將資料之值壓縮在 0 附近，即資料會符合平均值為 0，標準差為 1 之標準常態分布。以下為 Z-Score 公式：

$$x^* = \frac{x - \mu}{\sigma} \quad (3-1)$$

(四) 資料不平衡處理:

大多數現代的機器學習系統都屬於分類器，系統將輸入資料分類到不同的種類中，像是特定動物的照片，而**理想上來說，當一個系統被訓練用來辨識多個種類時，每個種類的訓練樣本數大致上要相等**，例如違約與不違約資料數量不能相差太多。若訓練樣本數量差異太大，系統的辨識結果可能偏向訓練樣本數較多的種類。本專題採用 Haixiang et al.(2017) 與 Amin et al.(2016) 提到在處理資料不平衡可以從外部資料層面，透過減少資料較多的類別或增加資料較少的類別的資料，以改變不平衡數據集的分佈。在參考 Namvar et al.(2018) 之後**本專題採 ROS(Random Oversample) 隨機上採樣¹**，將數據較少的資料類別增加到與數據較多的資料類別相等的方式，來處理資料不平衡的問題。

¹ 上採樣會把小眾樣本複製多份，一個點會在高維空間中反復出現，這會導致一個問題，那就是運氣好就能分對很多點，否則分錯很多點。為了解決這一問題，可以在每次生成新數據點時加入輕微的隨機擾動，經驗表明這種做法非常有效。

三、多模型簡介

本專題在模型上採取一個多模型的一個比較，多模型的目的是在於每個模型演算方式有其特性，每種模型都有適用的情境，因此需要多種模型以因應不同情境下都能優化預測準確率的需求，因此**本專題希望在不同的模型中找出最適合本專題的模型**，來進行研究。

(一) KNN 機器學習演算法:

KNN 可以說是機器學習中最簡單的演算法之一，因為它不需要複雜的計算就能夠實現。首先，我們要先知道 KNN 分類運作的模式，KNN 的目標在於找出鄰近距離最短的 k 個點，並透過「多數決」的方式決定該點屬於哪一類。

(二) Random Forest 機器學習演算法:

隨機森林是決策樹的一種衍伸方法，它集合了多個決策樹，同時加入隨機分配的訓練資料，以提升最終的結果，此方法為一種集成式學習，主要運用在分類與迴歸分析。首先要決定以幾棵決策樹來建構隨機森林，再以 Bagging(引導聚集算法)方式將訓練資料生成新的訓練資料，如果有 M 個自變數，每個節點都將隨機選擇 $m(m < M)$ 個自變數來進行分支，然後運用這 m 個自變數來確定最佳的分裂點，重複步驟直到指定的數量為止。

(三) XGBoost 機器學習演算法:

XGBoost 是一種可擴展式的 Gradient Boosting 的演算法，為 Gradient Boosting Decision Tree(GBDT)演算法的變種，可以用於分類也可以用於迴歸問題。其中 Gradient 是指在添加新模型時使用了梯度下降演算法來最小化的損失，Boosting 是將多個弱分離器組合起來形成強分類器，進而使目標損失函數達到極小。Gradient boosting 就是通過加入新的弱學習器，利用損失函數的負梯度方向在當前模型的值作為殘差的近似值，進而擬合一棵 CART 回歸樹(決策樹演算法的一種)做為學習器，而 GBDT 會累加所有 CART 回歸樹的結果，作為最終預測結果，會比單棵樹的預測效果來的好。

XGBoost 通過對目標函數做二階泰勒展開，從而求出下一步要擬合的樹的葉子節點權重（需要先確定樹的結構），從而根據損失函數求出每一次分裂節點的損失減少的量，從而根據分裂損失選擇合適的屬性進行分裂。

(四) 最後模型選擇:

在 2015 年的時候 29 個 Kaggle 優勝隊伍中有 17 隊使用 XGboost 作為他們的演算法，由此可知 XGboost 是非常強大的機器學習模型，而在經過本研究測試之後也發現 XGBoost 的表現最為優異，因此我們**最後選擇使用 XGBoost 演算法**。

表 3-4 模型預測力

	KNN	隨機森林	XGboost
原來預測力:	0.716	0.698	0.799
錯放率:	0.320	0.320	<i>0.210</i>

(五) 職業變數切割分群

馬若雅(2019)、何韻儀(2019)參考過去文獻提出的"信用評等較低區間的預測力較差"問題，嘗試以將樣本切割成子樣本的方式提升準確性，並且兩篇研究論文中分群後重新訓練後對大部分的群體預測力都有提升的改變。因此我們決定利用此改年對我們信用評分模型優化。

至於樣本如何切割的依據在 Thi Huyen et al.(2007)、Jonathan D.Glater(2008)、Wang et al. (2017)等論文中都曾提出職業在銀行貸款是非常重要的評估要點，當一個人擁有穩定的收入，其成功貸款的的機率就會越大。因此本專題採用此文獻提出的要點針對各個職業進型分群且建模並預測。

上述兩點，皆支持我們進行分群訓練預測且在職業類別上進行分群的動作。

表 3-5 不同模型，分群後預測力以及錯放率比較

	KNN	隨機森林	XGboost
A:管理職業	0.724	0.699	0.868
錯放率:	<i>0.310</i>	<i>0.320</i>	0.220
B:商業和金融運營職業	0.757	0.773	0.893
錯放率:	<i>0.280</i>	<i>0.260</i>	<i>0.140</i>
C: 計算機和數學職業	0.809	0.890	0.951
錯放率:	<i>0.260</i>	<i>0.120</i>	<i>0.070</i>
D:建築和工程職業	0.774	0.838	0.955
錯放率:	<i>0.260</i>	<i>0.200</i>	<i>0.070</i>
E:生活，體育和社會科學職業	0.787	0.853	0.938
錯放率:	<i>0.260</i>	<i>0.230</i>	<i>0.120</i>
F:社區和社會服務職業	0.781	0.860	<u>0.967</u>
錯放率:	<i>0.270</i>	<i>0.120</i>	<u><i>0.060</i></u>

	KNN	隨機森林	XGboost
G: 法律職業	0.781	0.948	0.961
錯放率:	<i>0.300</i>	<i>0.080</i>	<i>0.080</i>
H: 教育指導與圖書館職業	0.756	0.819	0.891
錯放率:	<i>0.290</i>	<i>0.220</i>	<i>0.150</i>
I: 藝術設計, 娛樂, 體育和媒體職業	0.794	0.928	0.956
錯放率:	<i>0.260</i>	<i>0.100</i>	<i>0.080</i>
J: 醫療從業人員和技術職業	0.728	0.801	0.860
錯放率:	<i>0.310</i>	<i>0.260</i>	<i>0.190</i>
K: 醫療保健支持職業	0.663	0.802	0.819
錯放率:	0.360	<i>0.230</i>	0.210
L: 保護性服務職業	0.767	0.909	0.935
錯放率:	<i>0.280</i>	<i>0.240</i>	<i>0.090</i>
M: 食物製備及服務相關職業	0.685	0.804	0.859
錯放率:	0.320	<i>0.260</i>	<i>0.200</i>
N: 建築物及地面的清潔和維護職業	0.687	0.872	0.930
錯放率:	0.390	<i>0.130</i>	<i>0.080</i>
O: 個人護理和服務職業	0.740	0.876	0.925
錯放率:	<i>0.300</i>	<i>0.180</i>	<i>0.120</i>

	KNN	隨機森林	XGboost
P: 銷售及相關職業	0.717	0.831	0.877
錯放率:	<i>0.290</i>	<i>0.210</i>	<i>0.170</i>
Q: 辦公室和行政支持職業	0.794	0.792	0.897
錯放率:	<i>0.260</i>	<i>0.220</i>	<i>0.130</i>
R: 農業, 漁業和林業職業	0.818	0.909	0.818
錯放率:	<u><i>0.250</i></u>	<i>0.120</i>	0.250
S: 建築和提取職業	<u>0.834</u>	<u>0.967</u>	0.957
錯放率:	<u><i>0.250</i></u>	<u><i>0.030</i></u>	<u><i>0.060</i></u>
T: 安裝, 維護和維修職業	0.736	0.809	0.906
錯放率:	<i>0.310</i>	<i>0.250</i>	<i>0.120</i>
U: 生產職業	0.717	0.761	0.870
錯放率:	0.330	<i>0.310</i>	<i>0.150</i>
V: 運輸和物料搬運職業	0.672	0.777	0.847
錯放率:	0.350	<i>0.240</i>	<i>0.180</i>
W: 軍事特定職業	0.713	0.862	0.929
錯放率	0.327	<i>0.152</i>	<i>0.103</i>

粗體字為預測率提升, 斜體字為錯放率下降, 底線為表現最好

從表 3-5²可以看出，在進行多模型比較中，除了 S 群表現最好的是隨機森林以外，其他分群都是 XGBoost 為準確度最高以及錯放率最低的一個機器學習演算法。

在表 3-5 與表 3-4 比較後，可以得知：

第一，在 KNN 模型中 23 個職業群中有 18 個職業群的預測準確度確實提升。其中幅度最大為 S:建築和提取職業(0.716 升至 0.834)；另外有 17 群是錯放率下降的，下降最大幅度 R:農業，漁業和林業職業、S:建築和提取職業(0.320 降至 0.250)。

第二，隨機森林模型中 23 個職業群全部的準確度皆有大小不等的增加，最大為 S:建築和提取職業(0.698 升至 0.967)；錯放率部分在全部職業群皆是下降，下降最多 S:建築和提取職業(最低有從 0.32 降至 0.03)。

第三，XGBoost 模型所有職業群準確率都有提升，提升幅度最大為 F:社區和社會服務職業(0.799 升至 0.967)；而至於錯放率，有三個群不降反升。錯放率下降最大幅度 F:社區和社會服務職業、S:建築和提取職業(從 0.21 降至 0.06)。

這個結論對本研究很有幫助，發現將職業分類後，對於各職業類別預測力的提升有助於借貸平台對於"借貸小白"預測力提升，對於這些族群的放貸也更有把握，也使更多人能進行借貸。測試三個模型以及職業分群後，確實呼應本專題在機器學習上的成果預期，在分群之後各個職業群體大部分都有預測力提升以及錯放率下降的趨勢，這也使得本專題不只提供更創新的使用介面平台，在平台的預測力方面也絲毫不遜色。

² 表 3-5 A-W 共 23 群，分群方式資料來源來自美國勞工局(U.S BUREAU OF LABOR STATISTIC)

混淆矩陣:

準確度代表所有樣本中預測正確的比例，計算公式如下:

$$P(A) = \frac{(TP+TN)}{(TP+FN+FP+TN)} \quad (3-2)$$

		Predict Label	
		P	N
True Label	P	True Positives(TP)	False Negatives(FN)
	N	False Positives(FP)	True Negatives (TN)

圖 3-5 混淆矩陣

錯誤放貸率代表預測不違約的情況下，實際上是發生違約的機率，特別強調此比率的原因在於，這是實際會讓公司直接虧損的，而這個比率越高代表公司在放貸上會有越多直接損失，也代表這個模型在預測違約的能力非常弱:

$$P(W) = \frac{FP}{TP+FP} \quad (3-3)$$

第三節 語音輸出

於前一步驟中我們利用機器學習之方法、以 Lending Club 之資料，訓練出違約預測模型，並將借款人之資料輸入預測模型中做違約預測判斷。

在分析結果輸出前我們將其儲存與資料庫中。而資料庫之建構中，我們利用 LineBot 提供的 Line 使用者特殊 ID³標記每一筆借貸資料作為資料庫的結構。資料儲存後我們再將預測結果輸出。

為實現如緒論中駕駛情境也能便利使用，我們利用 **Google Speech 深度學習、類神經網路機器學習演算法 API (Application Programming Interface) 服務**，將儲存後的資料轉譯成語音，再將語音檔回傳到 LineBot 輸出。

³與我們平時所用來加好友的 ID 不同，使用者無法自行輸入，是由系統自行產生給每支 Line 帳號的辨識碼，如(ID: U8955b5b423b1d6be9a9a6e902b0fcccc)

第四節 生活資訊

本研究除了探討如何降低 P2P 借貸的違約率，以及製作 P2P 介面之外，更在我們使用的 LineBot 介面上，增加一些可使一般生活使用上更便利的功能。我們增加了可以查詢天氣、股價、匯率的功能，甚至是一般人在為了晚餐到底該吃煩惱時，只要在我們製作的機器人中輸入對應的資訊後，我們的機器人都會給予即時的回應，以解決現代人常苦惱的問題。

而本研究選擇在原先的 P2P 介面上增加這些功能，目的就是為了增加在使用此介面時可以增加互動性，讓使用者在使用時感到更加親切。本介面跟其他市面上的介面不一樣的地方是，我們不單單只提供信用評分服務，更可以跟使用者有些簡單的小互動，以此來增加使用者的體驗感受，讓借貸不再只是有壓力的事情。因為即使被系統評分不合格後，還是可以使用其他生活化的功能，這就是本研究添加這個功能的目的。

上述查詢天氣、股價和匯率功能之**實現方式**，是將伺服器串接到 AZURE 語意辨識 API，因此當使用者在自己的行動裝置中於 Line 輸入關鍵字詞時，例如台北市天氣、台積電股價和美金匯率等，資訊將由 LineBot 傳送到伺服器再傳送到 AZURE 上。AZURE 辨識出對應的資訊後，再回傳給伺服器，伺服器得到辨識結果後，便會上網進行爬蟲，得到結果後再將結果傳送給使用者；而查詢食物的功能，是預先以人工的方式輸入字庫，在系統接受到相對應的字詞時，例如午餐、晚餐等，便會給予答覆，而此功能因為是用人工的方式輸入字庫，因此我們將輔大周邊的食物作為選項（約有數十種），若未來有其他好吃的餐廳出現，則會在系統上做出更新。

最後，我們收集的這些生活資訊，在未來若技術繼續發展，**或許可以成為一種全新的因子**，加入信用違約分析模型來提升預測準確率。例如平常都吃滷肉飯，小吃店的人，會比三餐都大魚大肉的人更難借到大數目的款項，原因可能是因為其薪資比較低；或者常查詢高風險高報酬股票的人，可能比都查詢低利低報酬股票的人更難借到款，可能因其屬於風險偏好類型。

第四章 研究成果與應用

本章針對我們開發之 API 所提供的服務進行示範，以下將會有詳述以及圖片說明。

圖 4-1 為平台的基本功能，分別是: 1.使用說明 2.我要借貸 3.我要取消 4.關於我們 5.位置資訊 6.聯絡我們。



圖 4-1 平台功能

在接下來的幾張圖片中，將會一一演示這些功能。

圖 4-2 為使用者按下使用說明之後所會出現的畫面，使用說明中會提到除了借貸平台的基本功能以外，還會提到附加的功能要如何進行操作。

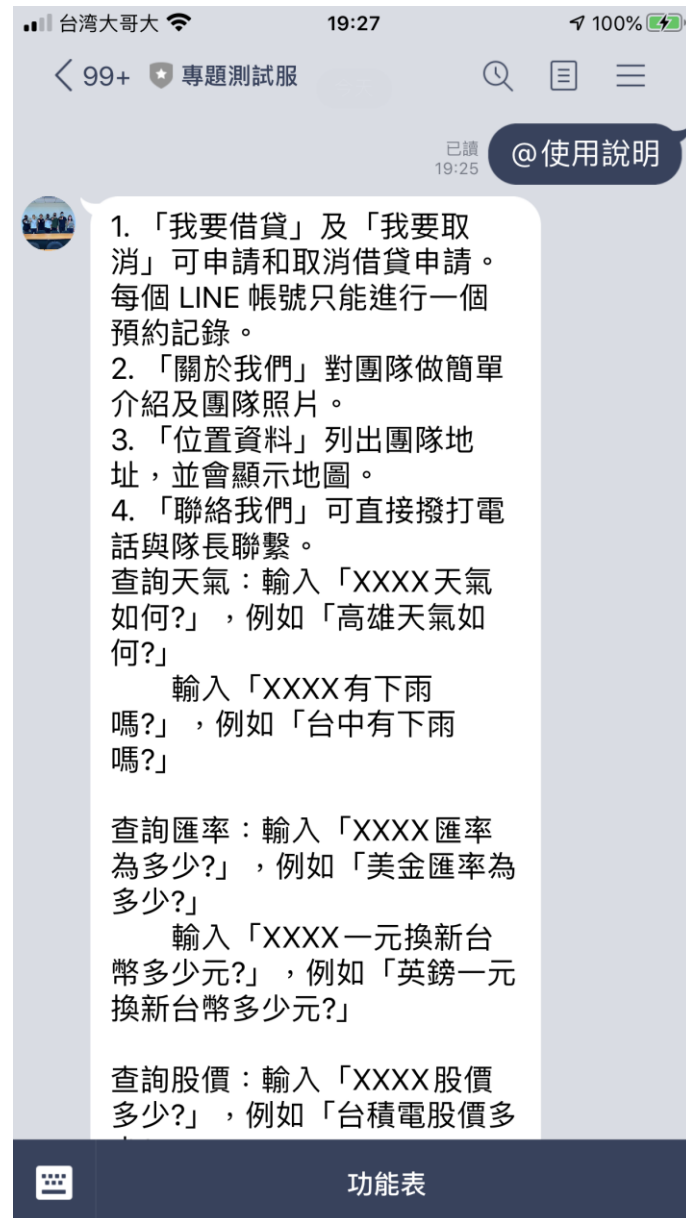


圖 4-2 使用說明

圖 4-3 為使用者按下我要借貸之後的畫面，按下填寫申請表單即可進入表單填寫畫面。



圖 4-3 我要借貸

圖 4-4 為表單填寫畫面，使用者需要輸入的資料有：

1.姓名 2.教育程度 3.房屋持有狀況 4.工作類別 5.預期借貸金額 6.預期借貸日期 7.FICO 分數下界 8.負債收入比 9.總信用額度 10.帳上餘額和現金卡額度比例



台灣大哥大 19:33 100%

< 99+ 專題測試服

今天

已讀 19:30 @我要借貸

借貸資料填寫
linetrytry2.herokuapp.com

姓名

教育程度

小學

房屋持有狀況

持有房屋且沒有貸款

工作類別

管理職業

預期借貸金額

預期借貸日期

FICO分數下界

圖 4-4 借貸申請頁面

圖 4-5 為申請成功畫面，系統將會重複使用者資料並且藉由語音告訴使用這申請合格。

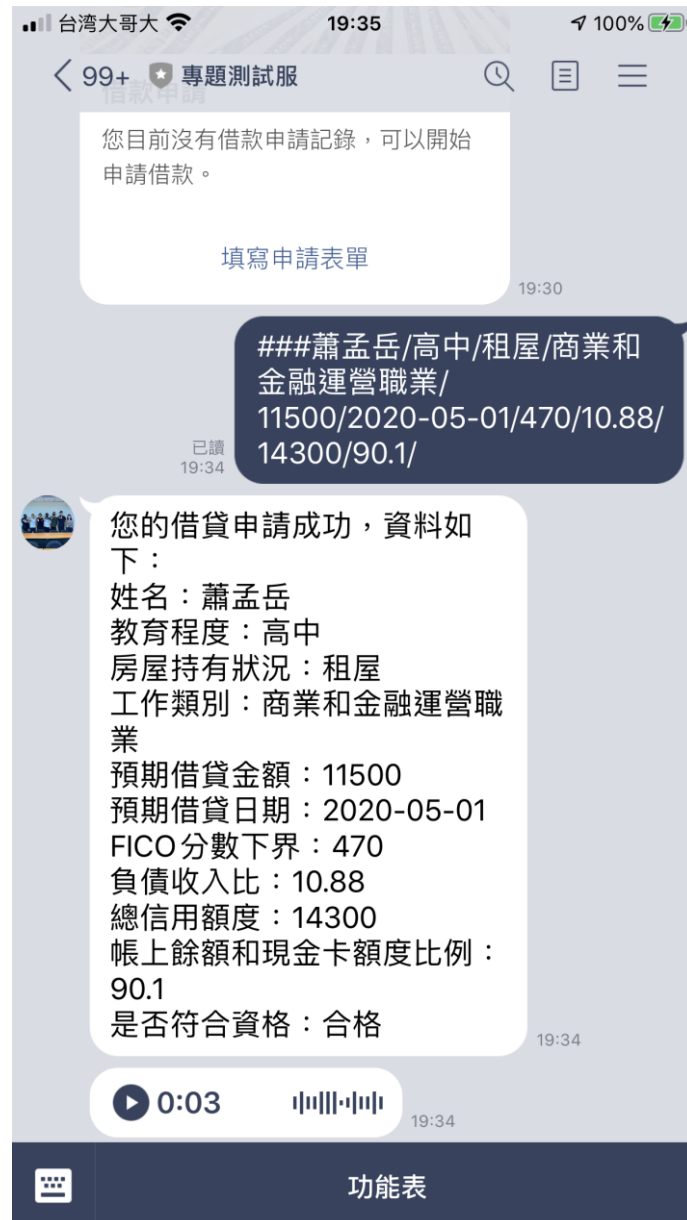


圖 4-5 申請成功畫面

圖 4-6 為申請失敗畫面，與上圖不同，申請失敗並不會重複完整申請資料，而是以一段文字還有語音來告訴使用者申請失敗。



圖 4-6 申請失敗畫面

圖 4-7 為如果申請成功的話，要取消的確認畫面。

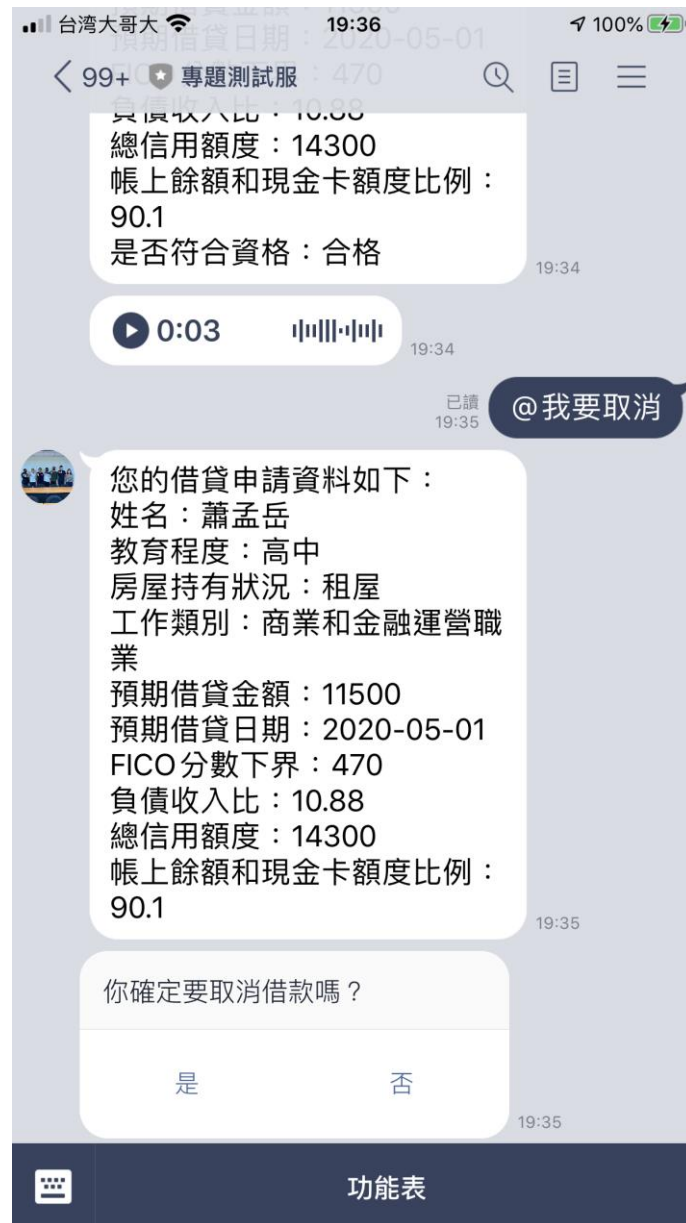


圖 4-7 我要取消

圖 4-8 為按下「是」取消成功以後，系統所會進行的回覆，另外若是按下「否」的話，系統則不會給予任何回覆。

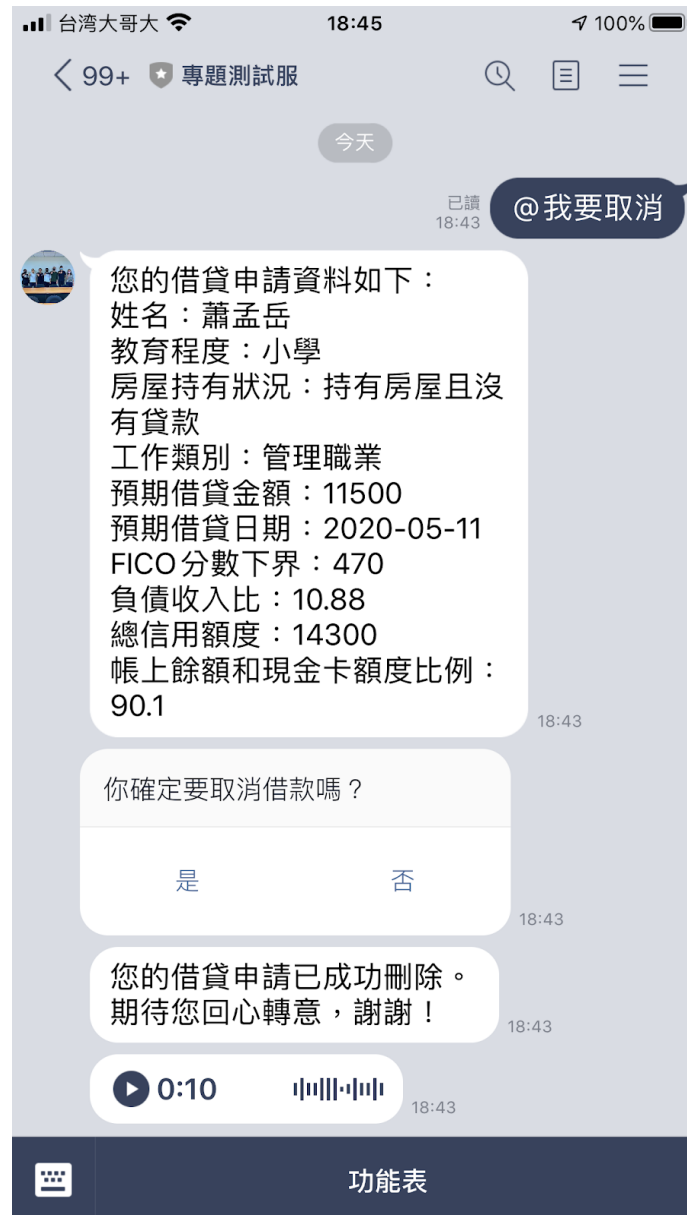


圖 4-8 成功取消

圖 4-9 為誤觸取消的話系統會做出的回應。



圖 4-9 誤觸取消

圖 4-10 為本專題團隊的合照與歡迎詞。



圖 4-10 關於我們

圖 4-11 為進行本專題研究的位置。



圖 4-11 位置資訊

圖 4-12 為要聯絡我們的電話。



圖 4-12 聯絡我們

圖 4-13 為詢問台北市的天氣狀況，實際上整個台灣的天氣都能查詢，系統會自動判別使用者詢問的內容是哪個縣市。

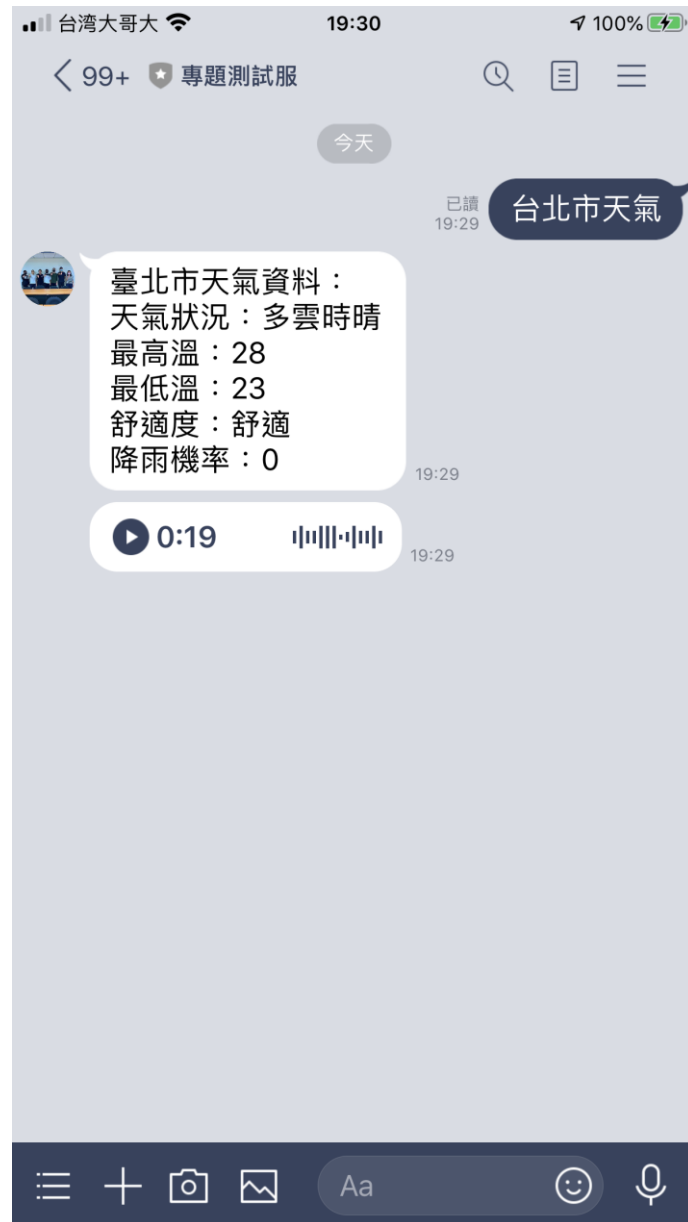


圖 4-13 天氣查詢

圖 4-14 為詢問美金的匯率，實際上台灣能換到的外匯都能查詢，系統會自動判別使用者詢問的內容是哪國外匯。



圖 4-14 匯率查詢

圖 4-15 為詢問台積電的股價，實際上整個台灣的上市股都能查詢，系統會自動判別使用者詢問的內容是哪支股票。



圖 4-15 股價查詢

圖 4-16 為詢問用餐建議，目前系統收錄了輔大附近本專題組員們覺得好吃的地點。

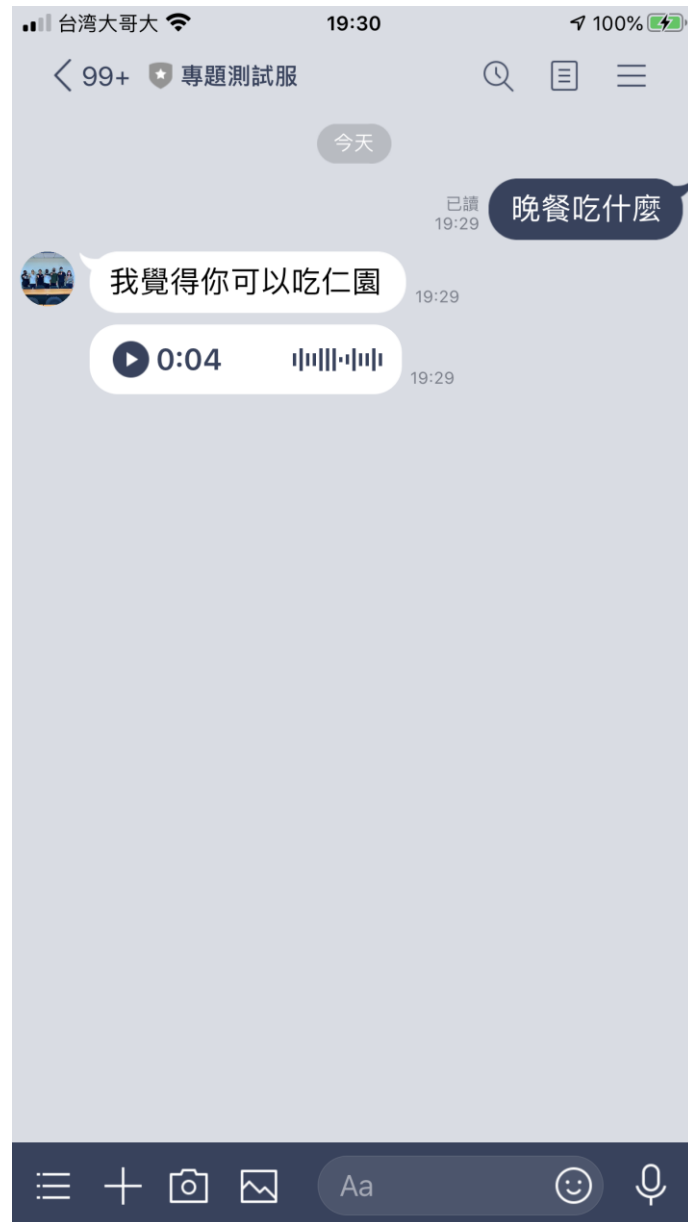


圖 4-16 用餐建議

圖 4-17 為申請成功之後，資料庫內部所有申請成功者的資料畫面。

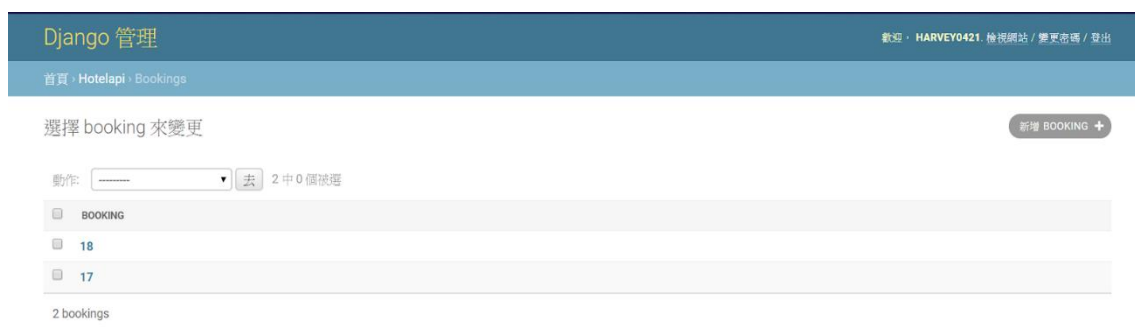


圖 4-17 server 資料庫

圖 4-18 為點進去個別資料中系統所會呈現的樣子，可以在系統內對個別資料進行修正。

Django 管理

歡迎 - HARVEY0421. 檢視網站 / 變更密碼 / 登出

首頁 > Hotelapi > Bookings > 2

變更 booking

Bid :

Uf061eb0b27d00c447f4a5c922107b4ab

User name :

蕭

Educationtype :

小學

Home ownership :

租屋

Job type :

管理職業

Funded amnt :

11500

Loan datetime :

2020-04-22

Fico range low :

470

Dti :

10.88

Tot hi cred lim :

14300

Bc util :

90.1

圖 4-18 變更資料庫內容

第五章 結論與建議

第一節 結論

近年來 P2P 借貸迅速崛起，由於相較於傳統金融機構借貸模式申貸時間較短、手續較為簡單，吸引的借款人大多為急用資金或無信用記錄，因此 P2P 借貸平台面臨較大的違約風險。當想要發展 P2P 借貸，會遇到的問題是部分目標客群缺乏相關的聯徵資料，難以進行信用評分。因此，平台會需要收集其他的資料來協助投資人做出判斷，解決投資人與借款人間的資訊不對稱的問題，降低投資人的違約風險。

本研究的目的之一是對信用評分分析之預測準確率提升。使用 Lending Club 資料，導入至機器學習模型進行建模。為了找出適合的演算法，我們採取多個機器學習的測試。為進一步提高準確率，利用職業變數對資料進行分群，再次導入機器學習進行建模。分群後的結果，也能發現準確率確實有提升，最大提升從 79% 提升到 96%，錯放率也有下降，最大降幅則是從 21% 降至 6%。

本研究主要目的是建立完整流程來串接機器模型和使用者介面，並期望最終能將開發結果變成 API 服務。而為了使服務有著比現有 P2P 網貸更為便利親民的體驗，我們除了使用 Line 來增加便利性，還串接 Google 提供的 API，加入了 AI 語音。最後也加上了生活資訊功能，除了來讓整個體驗更加親民，更希望未來能挖掘更多新的信用評分因子。

第二節 建議

一、未來發展

本次實驗有嘗試於資料輸入端融入 AI 語音，但由於在技術上 LineBot 無法與 Google Chrome 的語音輸入端相容，僅能做出語音輸出。在未來期望能將服務過程完全語音智能化。

對於信用評分新因子挖掘：有研究指出人類的個性及長相會間接影響其還款偏好，因不同的性格會影響其行為，進而影響其還款意願；也可從長相分辨出其性別、膚色，甚至是面相學之類的相關資訊，也會影響其還款意願，因此方面資訊未來可以加以探討。

在未來此系統若與其他人性生理特徵方面結合，如上述所說：記錄個人行為、加入性格評估、行為評比、外表分析等等，可令整體金融信用評分機制更加完善進而提高還貸預測準確率（可對應參考本論文章節 3-4）。甚至可以增加語音辨識、聲紋辨識、虹膜辨識等功能，多元整合 AI 智能，詳細記錄使用者的各項非結構因子（如：圖像、聲音），透過多種指標評級，可以統整出新的評等方式，令整體金融信用評分機制更加完善進而提高還貸預測準確率。

二、未來應用

因目前市場主流需求是將 AI 人工智能應用到生活中，在與居家生活場景融合，並透過互聯網以及大數據庫打造出更有未來科技感的智能社區。未來可以與生活居家型機器人做結合，倘若未來智能管家（市面上如：Amazon Alexa、Google Nest Mini、小米管家）可加入上述所有系統之功能，使其功能更加完善、對於使用者更為友善及便利，帶給使用者不同凡響的智能生活體驗。

參考文獻

一、中文部分

1. 馬若雅 (2019), 「公開資料對 P2P 平台現有違約模型預測績效之影響」, 未出版碩士論文, 輔仁大學金融碩士班, 1-8。
2. 邱敬桓 (2019), 「P2P 平台特徵變數選擇之探討」, 未出版碩士論文, 輔仁大學金融碩士班, 7-28。
3. 何韻儀 (2019), 「不同風險區間之違約與利潤模型配置設計」, 未出版碩士論文, 輔仁大學金融碩士班, 15-37。
4. 文淵閣工作室 (2019), 《Python 與 Line Bot 機器人全面實戰特訓班》, 台北: 碁峰。
5. 李震華 (2019), 「人工智慧技術發展趨勢與金融應用展望」, 《財經資訊季刊》95 刊 1-8 頁。available from URL: <https://www.fisc.com.tw/Upload/d56378d2-ff51-4552-9ab2-12ff1dd64a84/TC/9501.pdf>.
6. 藍國桐、蘇暉翔、古翔宇、林義旻、邱宇航 (2017), 「Line 機器人客服對話系統之設計」, 第九屆企業架構與資訊科技研討會。available from URL: <http://www.aea-taiwan.org/SEAIT2017/papers/039.pdf>.
7. 「AI & Blockchain 對金融服務業的影響與應用 及 AI in Fintech 人工智慧在金融科技應用趨勢」, 臺灣證券交易所企劃研究部專員 吳淑美整理, 證券服務 663 期。available from URL: <https://www.twse.com.tw/ch/products/publication/download/0001066327.pdf>.
8. 廖理、張偉強 (2017), P2P 網絡借貸實證研究: 一個文獻綜述 第 2 期 2-12。available from URL: https://www.pbcsf.tsinghua.edu.cn/Upload/file/20180410/20180410141940_4875.pdf.
9. 9 月 27 日央行理監事會後記者會參考資料 (2018)。available from URL: <https://webcache.googleusercontent.com/search?q=cache:tTI60PbiFacJ:https://www.cbc.gov.tw/tw/dl-24181-ec8e2bcf3cea4954b7c7fbf64b6d7965.html+&cd=4&hl=zh-TW&ct=clnk&gl=tw>.
10. 「什麼是 P2P 借貸? 臺灣有發展空間嗎?」。available from URL: <https://www.smartloan.com.tw/viewBlogDetail.do?kmlId=5>.
11. 「[30 天教你如何玩弄 Line bot API] 第 7 天: 教你如何自動嘴!」 available from URL: <https://ithelp.ithome.com.tw/articles/10219099>.

12. LineBot Developers - Messaging API 介紹。available from
URL:<https://developers.line.biz/zh-hant/docs/messaging-api/overview/#操作步驟>.

二、英文部分

1. Van Liebergen, Bart, (2017), " Machine Learning : A Revolution in Risk Management and Compliance ? ", *Journal of Financial Transformation,Capco Institute*, Vol. 45, 60-67.
2. Wang et al.(2017),"Rationality of the personal loan interest-rate markups of banks",*Corporate Management Review* ,Vol.37 No.1, 115-163.
3. Thi Huyen et al.(2007),"A credit scoring model for Vietnam's retail banking market", *International Review of Financial Analysis*,16, 471–495.
4. Benjamin Kojo Otoo(2012),"Micro-Credit for Micro-Enterprise: A Study of Women “Petty” Traders in Central Region, Ghana", *International Journal of Scientific Research in Education*,Vol. 5(3), 247-259.
5. Haixiang et al.(2017),"Learning from class-imbalanced data: Review of methods and applications", *Expert Systems with Applications*, Vol. 73, 220-239.
6. Amin et al.(2016),"Comparing Oversampling Techniques to Handle the Class Imbalance Problem: A Customer Churn Prediction Case Study", *Journals &Magazines IEEE Access*,Vol.4.
7. Namvar et al.(2018),"Handling Uncertainty in Social Lending Credit Risk Prediction with a Choquet Fuzzy Integral Model", available from
URL:<https://ieeexplore.ieee.org/abstract/document/8491600>
8. Jonathan D.Glater(2008),"That Student Loan,So Hard to Shake", *The New York Times* ,Aug.23,2008 available from
URL:<https://www.nytimes.com/2008/08/24/business/24loans.html>
9. 「 Standard Occupational Classification 」 , available from URL:
<https://www.bls.gov/soc/2018/#classification>