

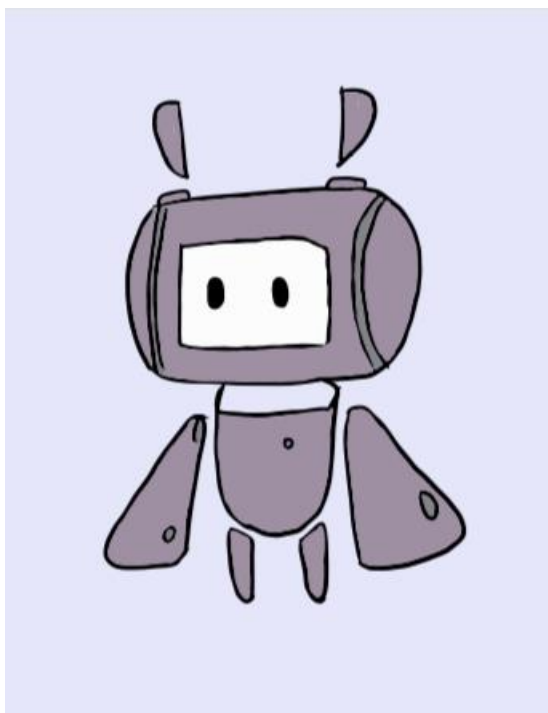


中原大學資訊工程學系

專題實驗 成果報告

CynTi

-結合音樂查詢與音樂性 MBTI 分析之聊天機器人



指導老師：吳宜鴻 教授

專題成員：

資訊四甲 11027114 徐梓薰

資訊四甲 11027125 謝旻軒

目錄

壹、摘要.....	3
貳、動機與目的	4
參、系統設計和方法	6
一、軟體設計架構及使用工具	6
二、MBTI 與音樂的相關性分析	7
三、MBTI 模型預測分析	12
四、機器人介面的設計和功能	14
肆、專題成果.....	20
一、主選單	21
二、分類搜尋功能	21
三、演唱會收藏功能	25
四、預測 MBTI 功能	25
五、推薦功能	27
伍、結論和未來方向	28
陸、參考資料.....	29

壹、摘要

每逢大型演唱會或音樂活動期間，歌迷們常會為了查詢演唱會資訊或獲取心儀的門票而煩惱。然而，目前市面上的音樂資訊查詢系統多缺乏對行動裝置的支援，無法滿足使用者快速搜尋及個性化管理的需求。

我們設計了一款名為 CynTi 的音樂聊天機器人，旨在協助用戶輕鬆查詢與管理與音樂和演唱會相關的資訊。此系統具備多元的查詢功能，不僅支持依日期或關鍵字（如地點、演出者）篩選演唱會資訊，更設有收藏功能，讓用戶能快速訪問收藏的演唱會。系統還提供藝人和歌曲之相關資訊，並根據用戶的音樂偏好推薦相似的曲風或藝人。此外，我們利用 Azure 平台進行自動化 ML 訓練，使用該平台將音樂特性與 MBTI 進行分析及連結，使用者可根據音樂偏好進行 MBTI 性格預測，讓用戶可以得知自己的音樂型 MBTI。上述功能透過 Line 介面和 Python 後端搭建，並配合 nginx、Docker、MongoDB 等技術，提供即時、穩定的使用體驗，使音樂資訊的獲取與管理更加便捷有趣。

貳、動機與目的

(一)、專題研究動機

對現代許多人來說，音樂已是日常生活的重要部分，而 MBTI 測試也常被用來代表個人的人格特質。在經過分析後，我們發現音樂品味與人格特質之間存在關聯性。此專題希望 CynTi 能讓使用者在享受音樂的同時，可以進行一些簡單的音樂查詢，甚至掌握自己的音樂型人格。

(二)、專題研究目的

我們希望設計一套系統，可以協助使用者查詢並管理與演唱會及音樂相關的資訊。具體目標包括以下幾點：

- **演唱會查詢與管理功能：**

演唱會查詢功能可讓使用者依類別查詢演唱會資訊。

(a) 日期查詢：根據日期範圍搜尋當天或即將舉行的演唱會。

(b) 關鍵字查詢：根據關鍵字（如地點、場館、演出者）搜尋演唱會資訊。

(c) 收藏功能：支援使用者將喜愛的演唱會收藏，方便日後快速訪問。

- **藝人與歌曲資訊查詢功能：**

提供詳細的藝人資料與歌曲資訊，使用者能夠方便地瀏覽藝人簡介、最新專輯、相似藝人等。

- **音樂 MBTI 推薦與預測功能：**

從使用者喜歡的歌曲提取音樂特性預測其 MBTI 人格，並依 MBTI 人格推薦同類型的音樂。

(a) 推薦功能：推薦與使用者相同 MBTI 用戶喜歡的歌曲

(b) 音樂型 MBTI 預測：根據使用者喜歡的歌曲特徵，預測其音樂型 MBTI，讓使用者更了解自己的音樂風格偏向。

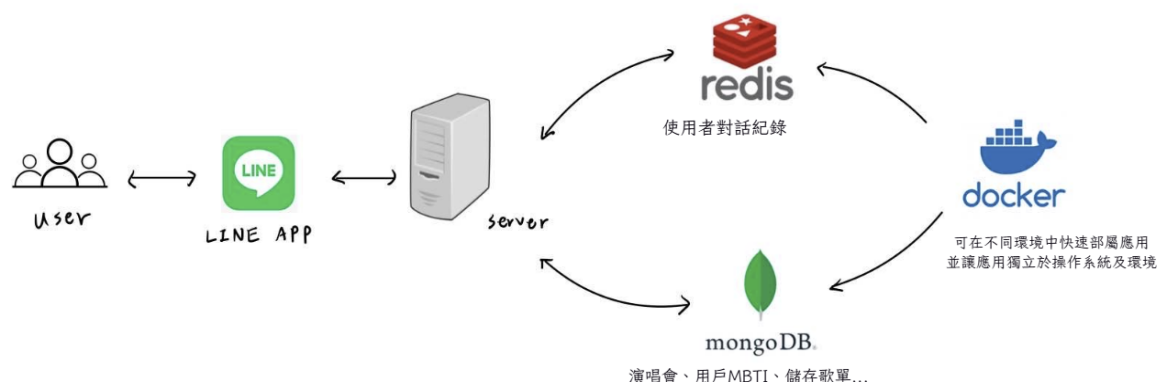
參、系統設計和方法

此章節將詳細說明本專題所採用的系統設計架構與方法，旨在分析 MBTI 性格類型與音樂屬性之間的相關性，並進一步利用機器學習模型進行預測分析。系統設計包含四個主要部分：軟體設計架構及使用工具、MBTI 與音樂的相關性分析、MBTI 模型預測分析及機器人介面的設計和功能。

(一)、軟體設計架構及使用工具

我們的系統採用了多種技術工具來確保穩定運行並提供最佳使用體驗。依圖一所示，使用者透過 Line 平台與系統互動，server 端選擇用 nginx 並透過 Docker 處理應用容器化，方便在不同環境中進行部署與維護，後端及資料儲存部分使用 Python 與 MongoDB 進行開發，其中 MongoDB 適合儲存半結構化的音樂及演唱會資料。

此外我們也利用 Azure 中的雲端服務進行自動化 ML 訓練 MBTI 分類模型，借助 BeautifulSoup(針對 tixcraft、kktix)爬取演唱會相關資訊，利用 selenium(連接 Spotify developer)進行自動化授權。使用 Flex Message Simulator 和 flask 框架進行卡片式訊息設計與後端 API 溝通，讓我們的系統可以即時、準確地呈現最新的演唱會與音樂資訊給使用者。



圖一 系統架構圖

(二)、MBTI 與音樂的相關性分析

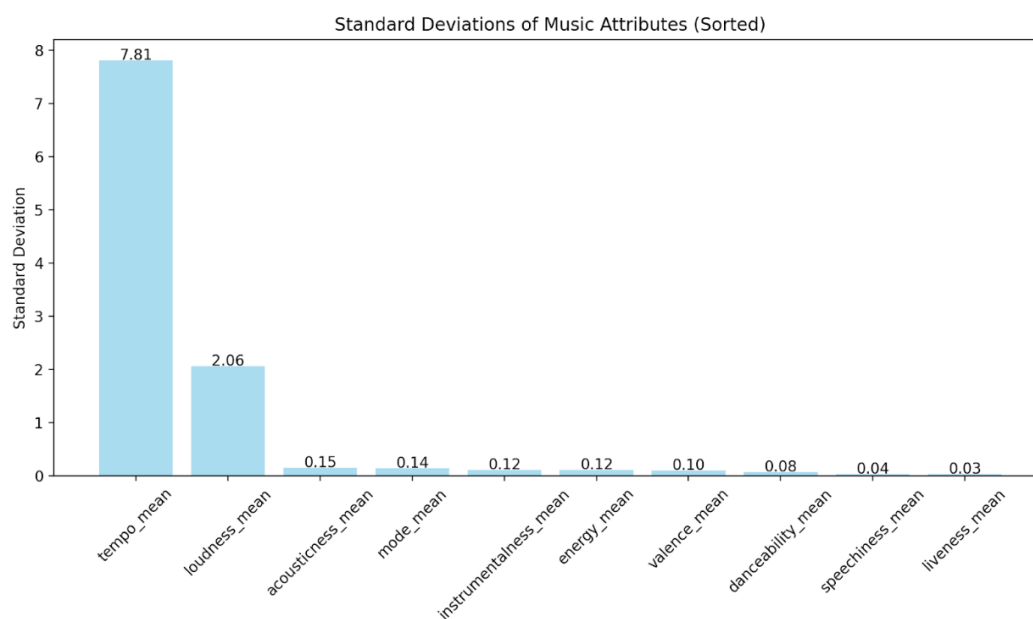
訓練 MBTI 的預測模型是使用來自 Kaggle 的資料集[1]，包含了 4081 筆資料，每筆資料代表一個 MBTI 類型對應的 Spotify 播放清單，其音樂特性經過聚合計算（平均值與標準差），每個特徵的詳細說明可參照表一。

表一 音樂特性與內容

特徵	內容
mbti	MBTI 性格類型(共 16 種)
danceability_mean / danceability_stdev	表示歌曲的節奏適合跳舞的程度
energy_mean / energy_stdev	歌曲能量的平均值和標準差，代表音樂的強度和活力
loudness_mean / loudness_stdev	曲目的整體音量，以分貝計
mode_mean / mode_stdev	歌曲的音調模式（大調或小調）
speechiness_mean / speechiness_stdev	曲目中語音的占比
acousticness_mean / acousticness_stdev	歌曲是否偏向原聲音樂
liveness_mean / liveness_stdev	錄音中有現場觀眾的可能性
valence_mean / valence_stdev	曲目音樂的正面情感表達程度
tempo_mean / tempo_stdev	曲目的每分鐘節拍數，表示速度
instrumentalness_mean / instrumentalness_stdev	樂器音的占比，越接近 1.0 表示越可能沒有人聲

此外，資料集中還包含 12 個調性的記數值(0_count-11_count)，這些欄位表示不同調性的計數值，反映了播放清單中特定音樂特性的分佈狀況。

將音樂特性的標準差排序後，由錯誤! 找不到參照來源。可得知 Tempo 與 Loudness 標準差相較於其他特性數值較大，表示數據的變異性越高，由此可得知不同的 MBTI 類型在這兩種音樂屬性上的表現差異大，以下利用 ANOVA 表格[3]對此兩種音樂屬性做進一步的分析。



圖二 音樂特性之標準差排序

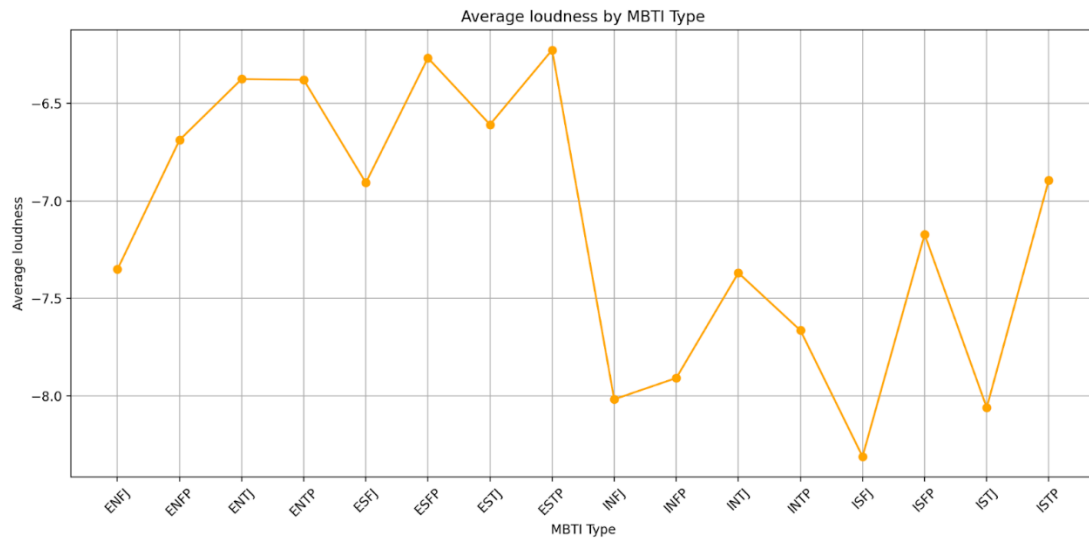
由表二表二 Loudness 之 ANOVA 表格可得知在 loudness 特性中，E&I 類型的 R^2 值雖然是四者中最高，但仍顯示出音量對於其解釋的變異性相對有限。儘管 N&S 的 R^2 值最低，但其 P 值仍顯示出在統計上是顯著的，這意味著 S 和 N 型在音量偏好上存在某些可辨識的差異。由上述可得知音量在所有類型中均顯示出顯著性。

表二 Loudness 之 ANOVA 表格

loudness	R-squared	R	P-value
E&I	0.06538	0.25569	6.248475e-62
N&S	0.00217	0.04658	2.918e-3
T&F	0.01178	0.10852	3.614035e-12
P&J	0.01446	0.12026	1.274014e-14

依圖三所示，在 16 種人格中，最喜歡音量大聲的歌的是 ESTP，最喜歡音量小的歌的是 ISFJ；且 E 人喜歡音量大的，I 人喜歡音量小的；N 人及 S 人並無顯著差異；F 人喜歡音量小的，T 人喜歡音量大的；J 人喜歡音量小的，P 人喜歡音量大的。

圖三 MBTI 與其平均之 Loudness 值的關係

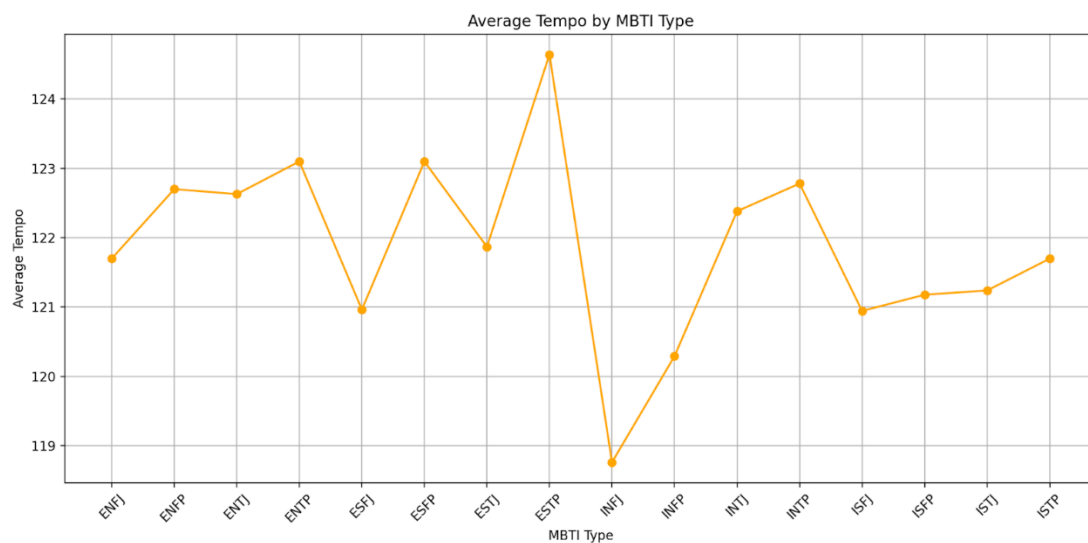


由表三 Tempo 之 ANOVA 表格可得知 N&S 的 P-value 為 0.1744(>0.05)，無法拒絕虛無假設，代表 S 人和 N 人在節奏偏好上並無顯著差異，E&I 和 T&F 之間的 P 值顯示出它們在節奏偏好上有明顯的差異，這也支持了 MBTI 類型對音樂偏好的影響。

表三 Tempo 之 ANOVA 表格

tempo	R-squared	R	P-value
E&I	0.2585	0.5084	2.064759e-11
N&S	0.0126	0.1123	1.74399e-1
T&F	0.2421	0.4920	6.274912e-10
P&J	0.1319	0.3631	5e-6

依圖四所示在 16 種人格中，最喜歡快歌的是 ESTP，最喜歡慢歌的是 INFJ，E 人喜歡快，I 人喜歡慢；N 人與 S 人不討論(無相關性)；F 人喜歡慢，T 人喜歡快；J 人喜歡慢，P 人喜歡快。



圖四 MBTI 與其平均之 Tempo 值的關係

依上述所言總結以下兩點：

1. Loudness 和 Tempo 的變異性：根據標準差的計算，Loudness 和 Tempo 的標準差相對較大，這表明不同 MBTI 類型在這兩個音樂屬性上的表現差異較大，顯示出顯著的音樂偏好差異。
2. 統計意義：儘管部分類型的 R^2 值較低，但 P 值仍顯示出有意義的結果，這說明音樂屬性之間的偏好差異仍具參考價值。

(三)MBTI 模型預測分析

由表四得知精確度的值為 0.27179，表示約 27.18% 的預測結果是正確的。F1-Score 是精確度和召回率的調和平均數，表示模型在平衡精確度和召回率上的表現，我們的 F1-Score 值為 0.26050 顯示出模型在正確的預測上存在困難，但由 AUC 值為 0.75921 得知，模型在區分正類和負類的能力上有一定的效果，尤其是在不平衡數據集的情況下，雖然模型在精確度和 F1-Score 上的表現不佳，但其在區分正負類別的能力上還是有一定的價值。

表四 執行計量

名稱	數值
精確度	0.27179
加權後 AUC	0.75921
加權後 F1 分數	0.26050

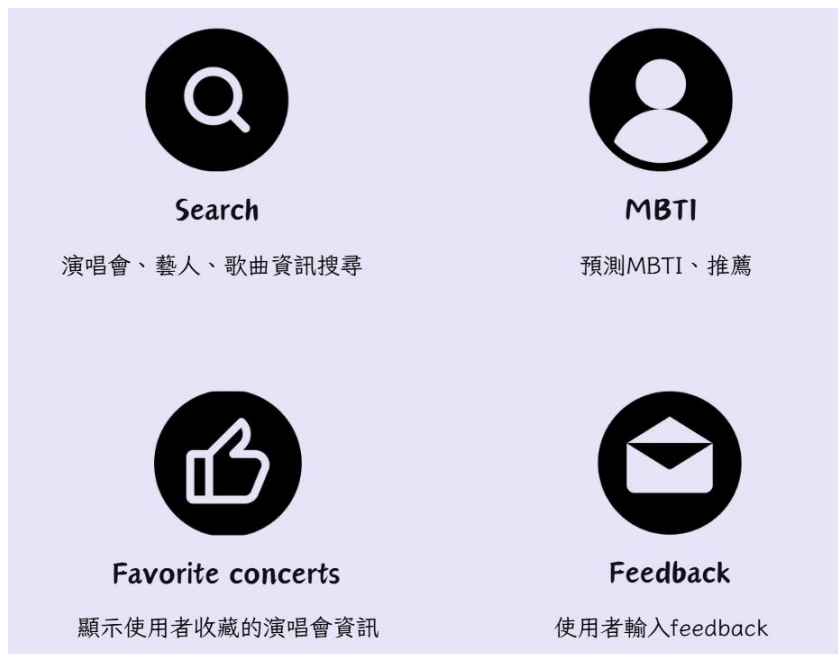
由圖五可得知對角線上的數字代表模型正確預測的比例。(例：對於 ENFJ 類型，有 23.94% 的樣本被正確地預測為 ENFJ)，非對角線的數字代表模型錯誤預測的比例。這些數字代表的意義是有多少實際為某一類別的樣本被預測為其他類別。(例：有 21.78% 實際為 ENFJ 的樣本被錯誤預測為 ESFJ)，由這張圖我們還可以得知對於 ENTJ 類型，有 43.98% 的樣本被正確預測為此類別是 16 個類別中最高的。但在 ENTJ 類別中被錯誤預測為 ESTJ 的比例是 25.47%，也是 16 個類別中最高的。結合以上被錯誤預測的類別中我們發現模型在分類 NS 類型與 PJ 類型時比較會發生錯誤預測的狀況。



圖五 十六分類模型評估指標-混淆矩陣

(四)機器人介面的設計和功能實作

由圖六可得知主選單主要功能分為四大類，包括 MBTI、Search 功能、收藏清單、及意見回饋，以下會根據 MBTI 功能、Search 功能、收藏演唱會功能以及資料庫設計詳細說明實作流程。



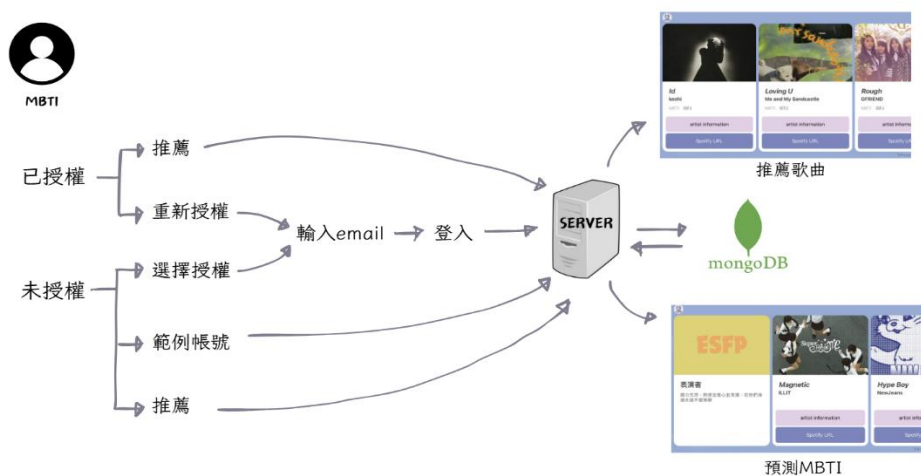
圖六 主選單

一、 MBTI 功能實作流程:

我們透過 Azure 實作 MBTI 分類模型，使用來自 Kaggle 的資料集，在 server 端需將使用者常聽的歌曲轉換成各項音樂特性的聚合資訊，再送入模型中進行預測。

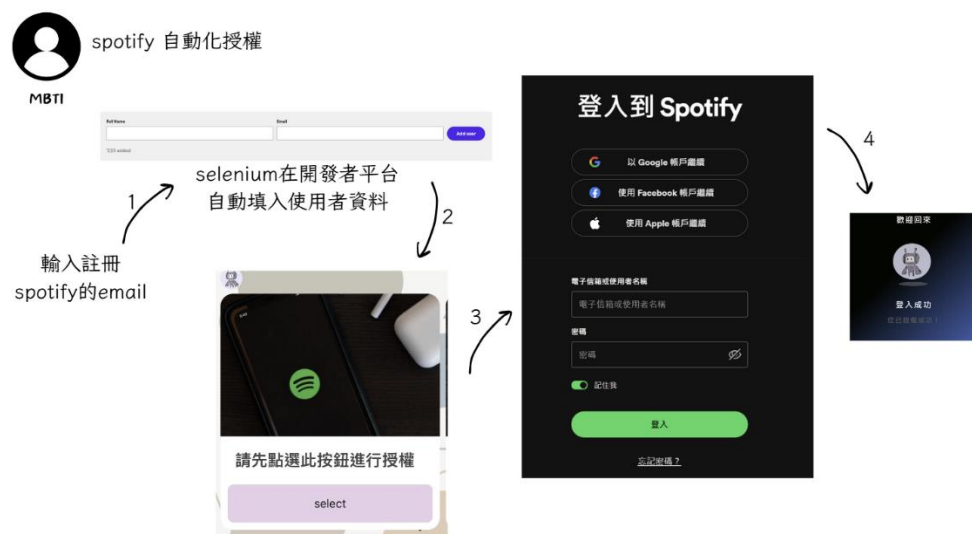
如圖七所示，是否授權會決定使用者可不可以進行 MBTI 預測，授權會透過 Spotify api 取得使用者的歌曲資訊，並將資訊保存在資料庫中。已授權的使用者可以選擇重新授權進行資料更新，因為隨時間不同喜歡的歌曲可能有改變，進而影響使用者的 MBTI。

另外，已授權及未授權皆可使用推薦功能，差別是授權過並預測出 MBTI 的使用者會依照該使用者之 MBTI 進行推薦(推薦該使用者的 MBTI 之其他使用者喜歡的歌曲)，而未授權的使用者則是會從其他已授權的使用者喜歡的歌曲隨機挑選進行推薦。



圖七 MBTI 實作流程圖

如圖八，由於 MBTI 的預測模型需要使用者的 Spotify 資訊，因此我們建立了 Spotify 自動化授權流程。使用者需要在聊天室輸入註冊 Spotify 的 email，後台會自動至 Spotify 開發者網頁填寫使用者訊息，並回傳給使用者登入的連結。使用者登入到 Spotify 後進行授權，server 方可接收到向 Spotify 請求的使用者資訊(即喜歡的前十首歌曲)。預測流程若正確進行會展示授權成功頁面，反之則顯示錯誤訊息。

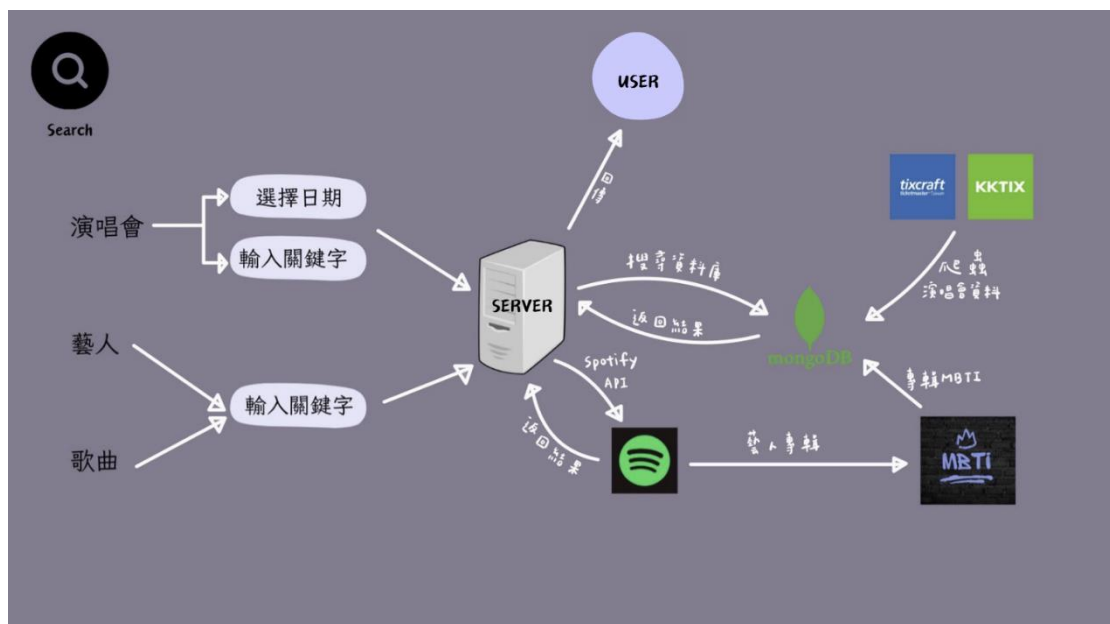


圖八 Spotify 自動化授權實作流程圖

二、 Search 功能實作流程:

可參照圖九，搜尋功能分為三類：演唱會、藝人及歌曲。演唱會的資訊透過 beautifulsoup4 爬蟲套件抓取 Tixcraft 以及 KKTIX 售票網站上的演唱會資訊，並且設定 server 每天會自動抓取一次，使得演唱會資料可以即時的更新。使用者可以透過日期或關鍵字進行搜尋，搜尋前會將資料庫中的演唱會資訊進行二元分詞，能更有效及更準確地進行關鍵字搜尋。日期則是使用 line 提供的日期選擇器，server 收到日期後再去資料庫查詢該日期對應的演唱會資訊。

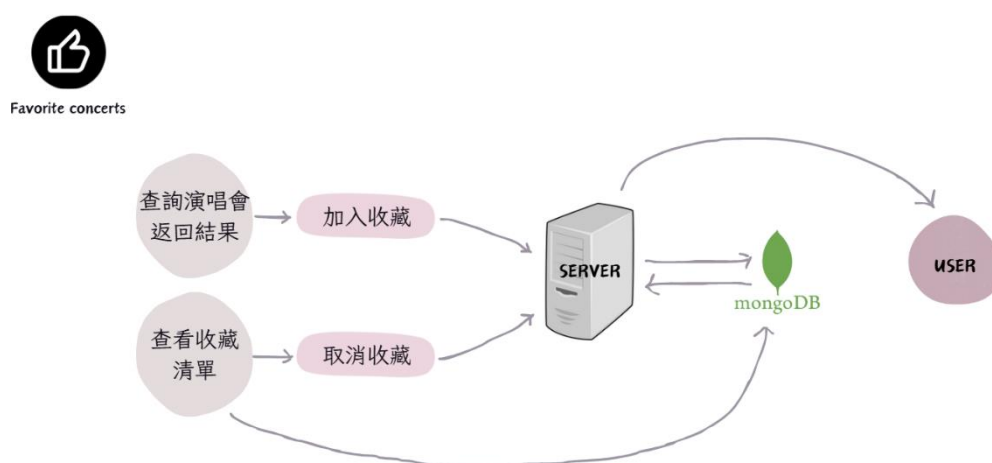
藝人及歌曲資訊則是串接 Spotify api，server 端會將使用者輸入的關鍵字，即時透過 Spotify api 獲得結果，並將回傳給 server 的資訊整合成卡片式訊息並回傳給使用者。



圖九 Search 實作流程圖

三、 收藏演唱會實作流程:

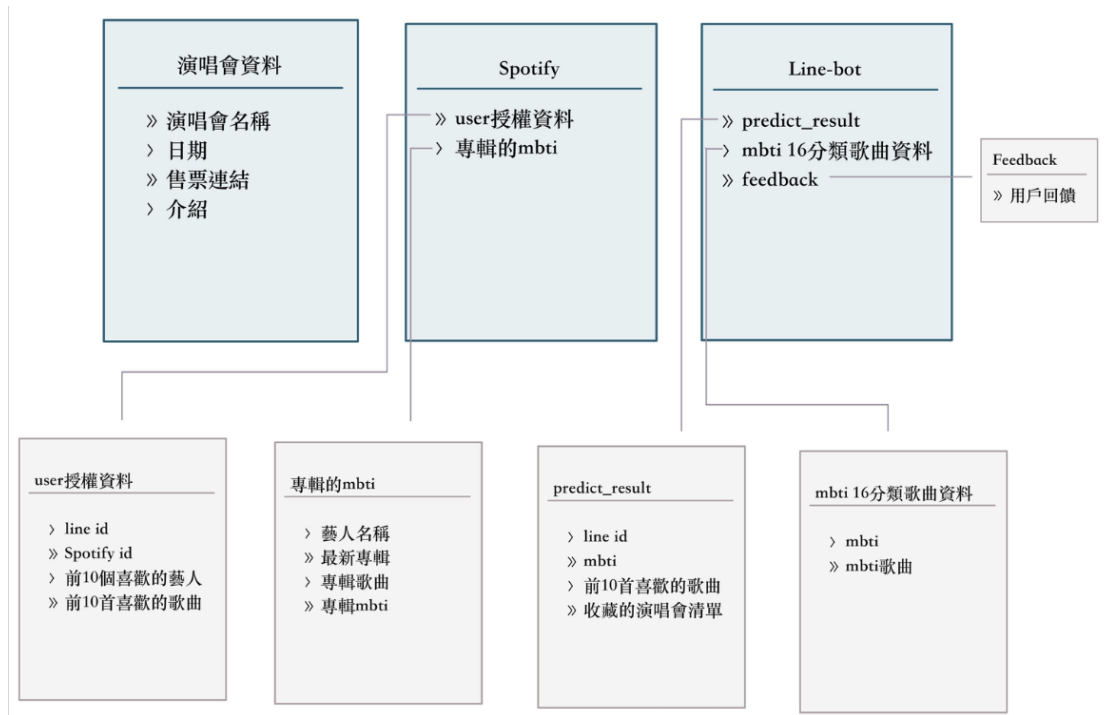
可參考附圖十，在查詢到演唱會資訊後若想把演唱會加入收藏清單，使用者只需按下演唱會的”Add to favorites”按鈕，server 會根據使用者的選擇將該演唱會加入資料庫裡使用者個人的收藏清單中。若選擇查看收藏清單，會去資料庫中取得對應使用者的收藏清單，並回傳給使用者。在收藏清單的功能中，使用者也可將特定演唱會資料從收藏清單中移除，即”Delete favorite”。



圖十 收藏演唱會實作流程圖

四、 資料庫設計:

如圖十一，資料庫架構主要分為三大類：演唱會資料、Spotify、Line-bot。演唱會資料庫中儲存我們從 Tixcraft 以及 KKTIX 爬取的演唱會資訊，包括演唱會名稱、開始結束日期、售票連結及演唱會介紹。spotify 資料庫中會儲存 user 的授權資料以及儲存使用者搜尋過哪些藝人的紀錄，後台會每天根據紀錄計算該藝人的專輯對應的音樂型 mbti 並新增在後台的資料中。Line-bot 資料庫中儲存 Line 使用者的 id、預測該使用者的音樂型 MBTI、使用者喜歡的前十首歌曲以及使用者的收藏演唱會清單。此外，推薦功能會使用到 MBTI 16 分類歌曲資料，有新的使用者預測 MBTI 時，server 會將該使用者喜歡的前十首歌曲存入此資料庫，因此此資料庫中會有 16 分類的 MBTI 及其對應 MBTI 類別的使用者喜歡的歌曲。



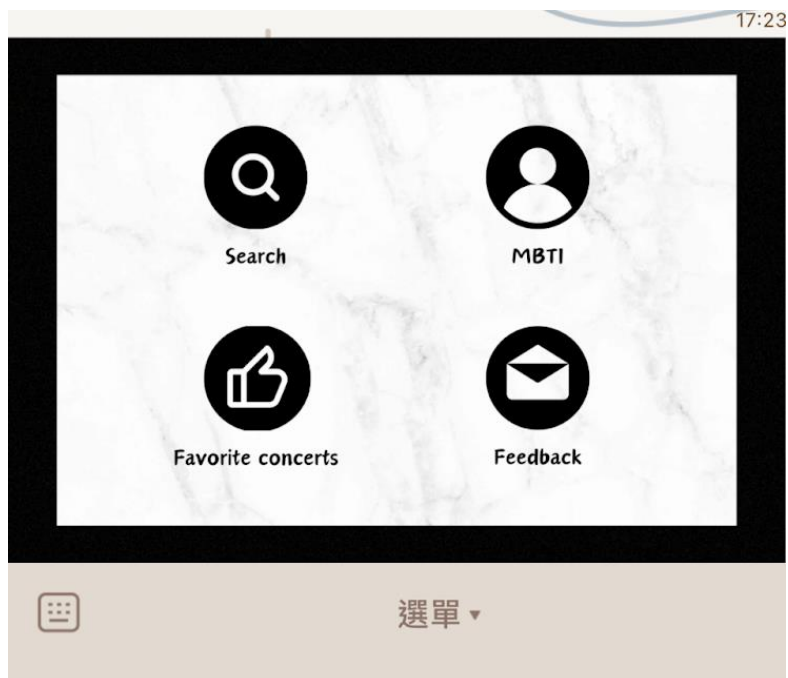
圖十一 資料庫架構圖

肆、專題成果

本章節將介紹系統最終實現的四大核心功能，旨在提升使用者的互動體驗與應用效能。系統設計目標在於提供一個便捷且多功能的平台，讓使用者能輕鬆進行音樂查詢、收藏喜愛的演唱會、預測音樂型 MBTI，並依據音樂偏好推薦適合的歌曲。以下將分為五大類進行成果展示，包含主選單、分類搜尋功能、演唱會收藏功能、預測 MBTI 功能及推薦功能。

(一)主選單

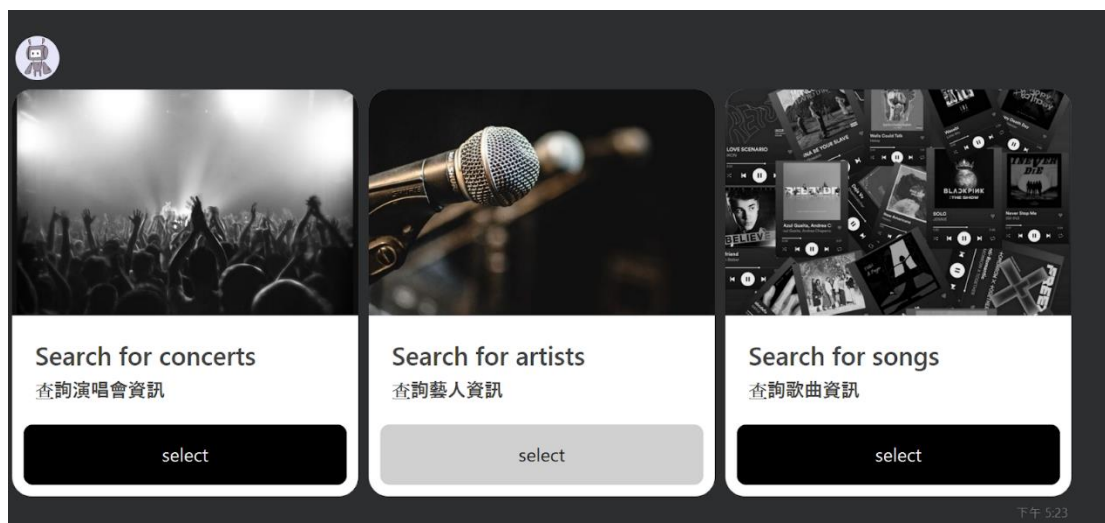
目前機器人在 LINE 介面採用的主選單如圖十二所示。



圖十二 主選單

(二) 分類搜尋功能

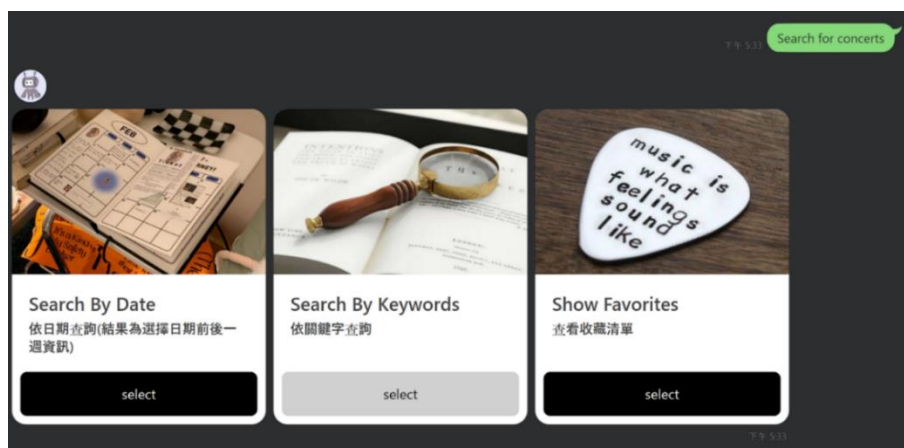
如圖十三所示，分類搜尋功能分為以下三類，以下將分為演唱會搜尋、藝人與相似藝人搜尋及歌曲搜尋進行展示。



圖十三 分類搜尋介面

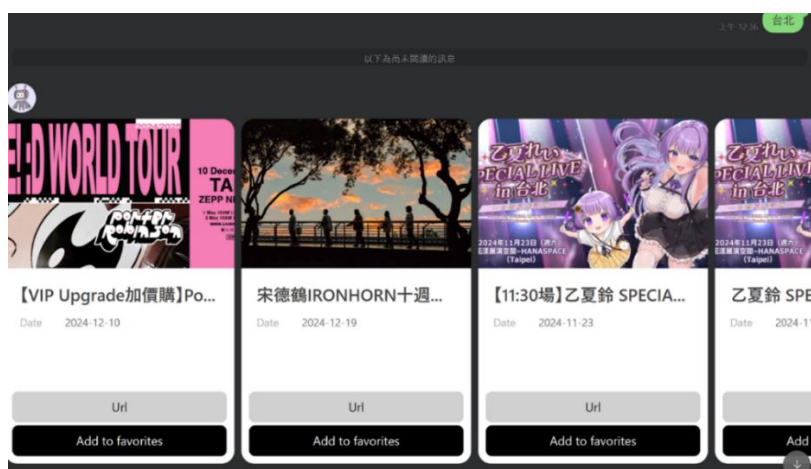
一、 演唱會搜尋:

如圖十四所示，使用者可以透過日期或者關鍵字搜尋，日期搜尋顯示的結果為選擇日期的前後一週的演唱會資訊，關鍵字搜尋可輸入特定地區(例:台北)或演唱會名稱。



圖十四 演唱會搜尋介面

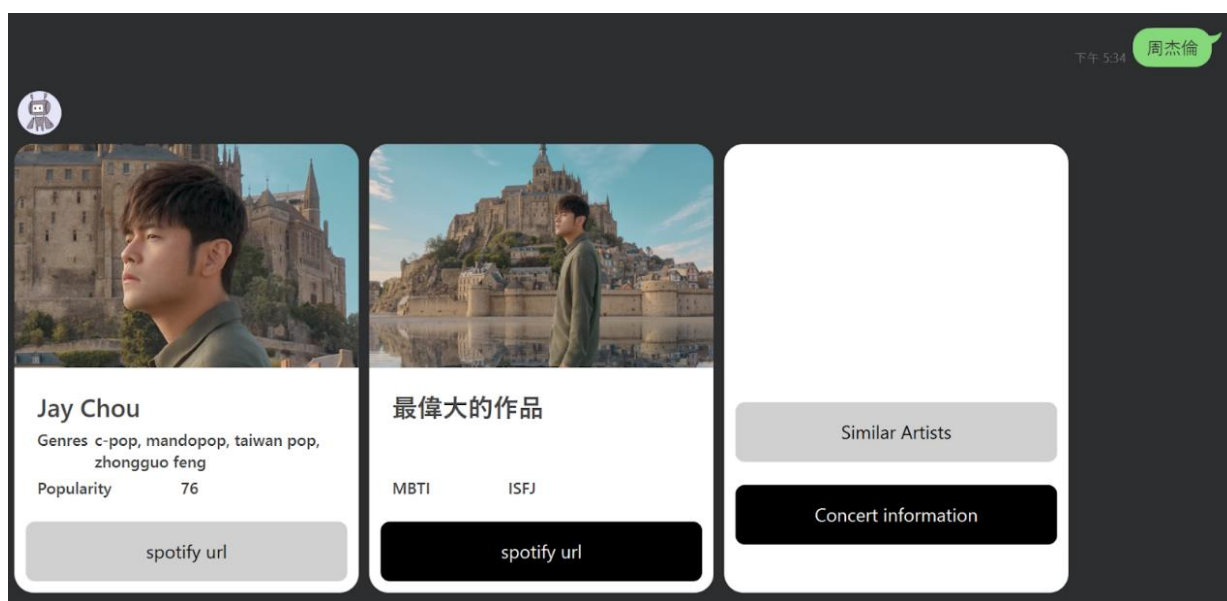
如圖十五所示，搜尋結果會顯示該演唱會售票網址的連結以及加入收藏的按鍵，透過結合兩大售票網站的售票訊息，省去查詢演唱會售票網站的時間，同時提供收藏功能可將想記住的演唱會資訊加入收藏，有效節省重複搜尋的時間。



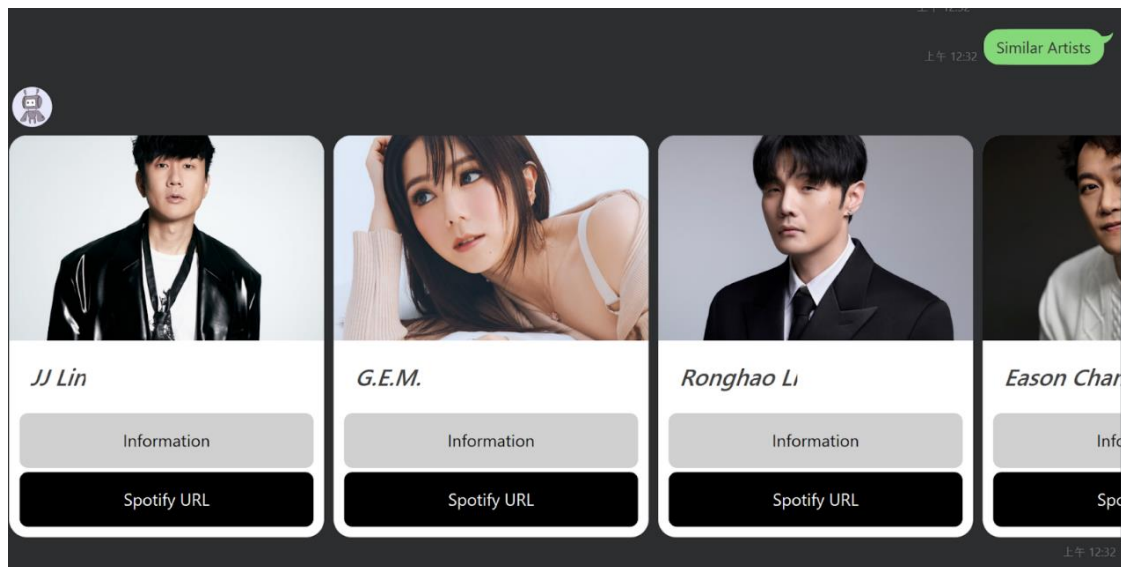
圖十五 演唱會搜尋結果

二、 藝人、相似藝人搜尋:

如圖十六所示，輸入想查詢的藝人名稱，第一頁的結果會顯示該藝人的曲風、受歡迎程度(範圍為 0-100)以及該藝人的 Spotify 頁面。第二頁則是該藝人的最新專輯，以及該專輯之音樂型 MBTI，下方連結可跳轉到該專輯的 Spotify 頁面。最後一頁可查詢該藝人的相似藝人或者搜尋相關演出訊息，如圖十七所示。相似藝人會顯示 10 項供使用者選擇，可透過此功能發掘其他可能有興趣的藝人，並進一步了解該藝人的詳細資訊。



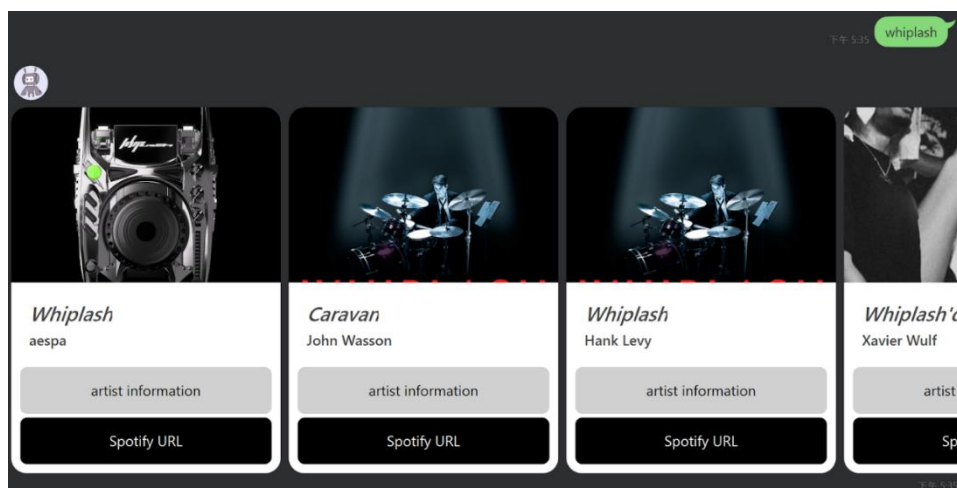
圖十六 藝人搜尋結果



圖十七 相似藝人搜尋結果

三、 歌曲搜尋：

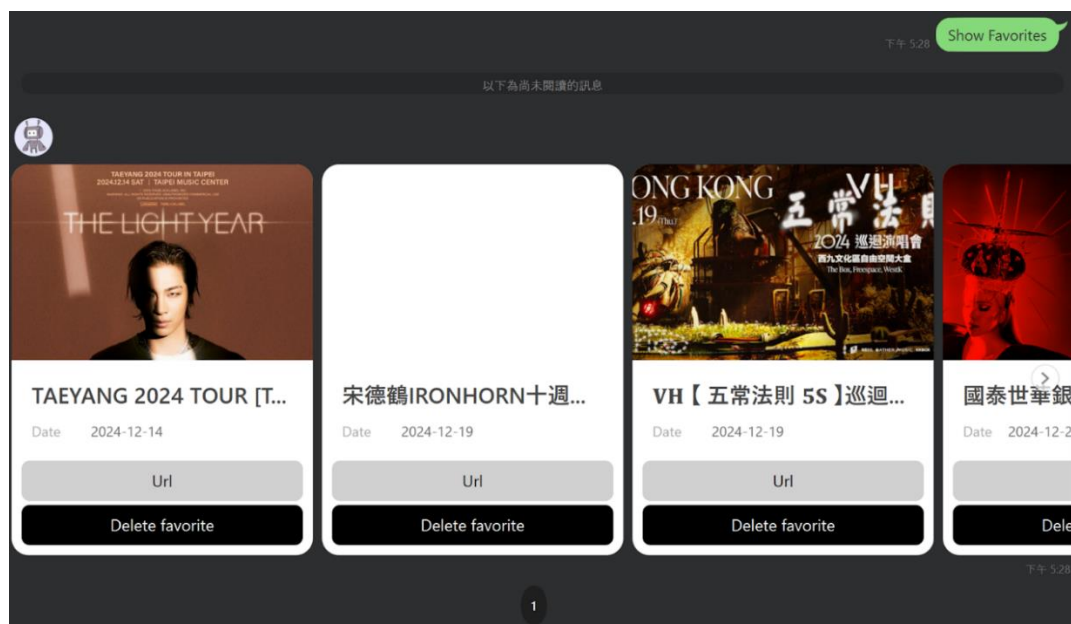
如圖十八所示，搜尋結果會顯示前五項與輸入關鍵字最關聯的搜尋結果。可以透過下方按鈕查詢該歌曲的藝人資訊，即上方所敘述查詢藝人之功能，下方連結可跳轉到 Spotify 的歌曲播放畫面。



圖十八 歌曲搜尋結果

(三)演唱會收藏功能

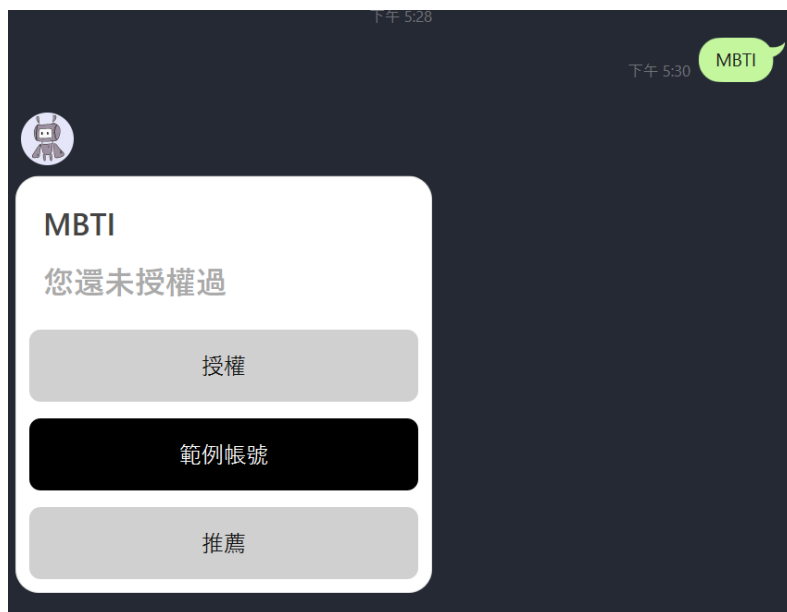
如圖十九所示，下方連結按鍵可跳轉到該演唱會售票網址，另外可以不限個數的新增或刪除演唱會，收藏的演唱會過期將會自動刪除，必免錯誤的資訊。



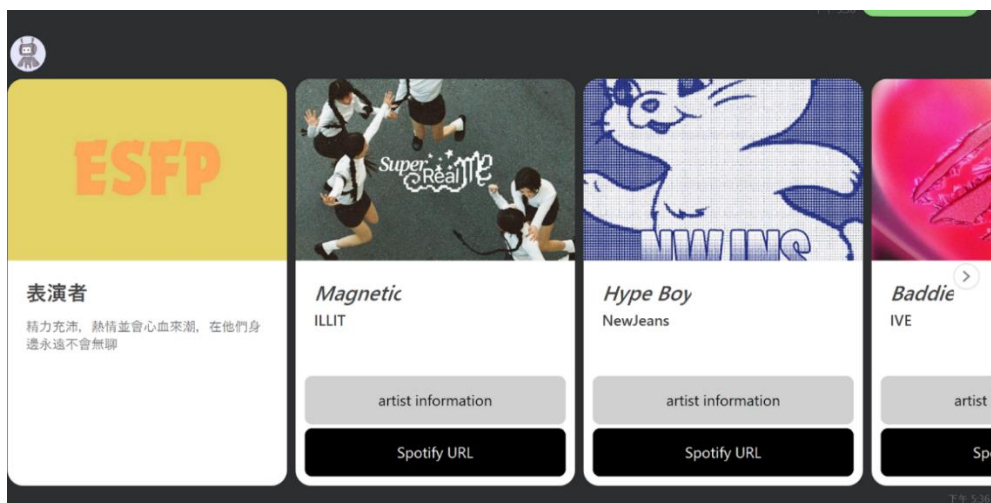
圖十九 演唱會收藏功能

(四)預測 MBTI 功能

MBTI 功能如圖二十所示，點選授權按鍵會經過 Spotify 授權流程取得使用者常聽的歌曲，並且將歌曲資訊經過處理後透過我們的分類模型進行預測。預測結果如圖二十一所示，會顯示音樂型 MBTI 人格以及該人格的介紹後面則是使用者最常聽的前五首歌。



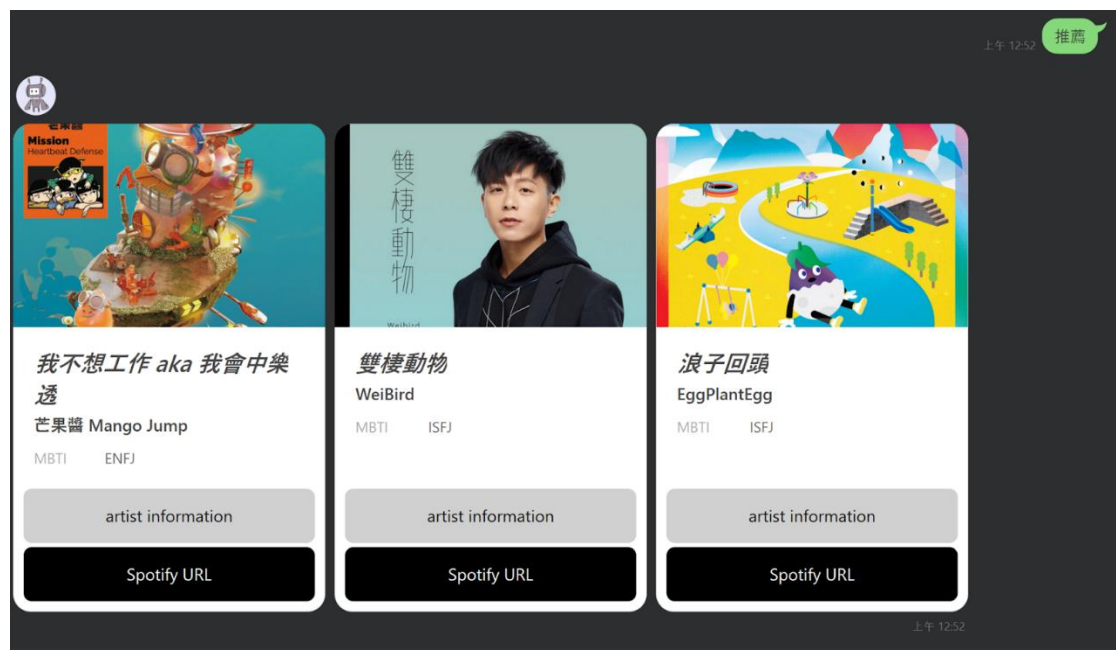
圖二十 預測 MBTI 介面



圖二十一 預測 MBTI 結果

(五)推薦功能

推薦功能的資料來源是將進行過預測的使用者喜歡的歌曲，依照使用者的音樂型 MBTI 分成 16 類並存入資料庫中，進行推薦時會推薦與該使用者有相同 MBTI 的其他使用者喜歡的歌曲。但若該 MBTI 類型的歌曲資料數不足則會推薦其他相似 MBTI 喜歡的歌曲。如圖二十二所示，推薦功能會顯示三首推薦歌曲，下方按鍵可查詢該藝人資訊以及該歌曲 Spotify 的聆聽頁面。



圖二十二 推薦功能結果

伍、結論與未來方向

結論:

目前的機器人系統已實現多項功能，致力於提升使用者的音樂探索與互動體驗。首先，使用者可以根據自己的音樂偏好，預測出屬於自己的音樂型 MBTI，進一步了解自己在音樂上的個性特質。系統也會根據這一 MBTI 分類，推薦與使用者相似偏好的歌曲，幫助他們發掘更多喜愛的音樂。此外，機器人整合了多個售票網站的資訊，使使用者能迅速找到心儀演唱會的購票連結。對於特別喜歡的演唱會，使用者還可以將其加入收藏，避免日後需要重複搜尋的麻煩。最後，系統提供藝人及歌曲的綜合資訊查詢功能，讓使用者能夠輕鬆獲得完整的音樂資料，提升整體使用體驗。

未來方向:

預測模型的精確度稍為有些不足，希望可以將此部分完善並且提供給使用者更加準確的預測結果，希望能增加使用者互動的部分，可以將相同音樂型 MBTI 的使用者串連並讓大家可以交流彼此喜歡的音樂以及分享演唱會資訊，讓此機器人的功能變得更加完善與增加社交性。

組員分工:

全部功能均由徐梓薰與謝旻軒共同討論與製作

陸、參考資料

- [1] xtrnglc. *Spotify MBTI Playlists*. Kaggle, 2023,
<https://www.kaggle.com/datasets/xtrnglc/spotify-mbti-playlists>.
- [2] NERIS Analytics Limited. 免費系統測試 16Personalities,
<https://www.16personalities.com/tw/%E6%80%A7%E6%A0%BC%E6%B8%AC%E8%A9%A6>
- [3] Fisher, Ronald A. *Statistical Methods for Research Workers*. Oliver and Boyd, 1925.
- [4] 拓元股份有限公司. "Tixcraft 拓元售票系統客服中心." Tixcraft, n.d., <https://help.tixcraft.com/hc/zh-tw>. Accessed 12 Nov. 2024.
- [5] KKTIX. "KKTIX 活動平台." *KKTIX*, n.d., <https://kktix.com/>. Accessed 12 Nov. 2024.
- [6] Spotify. "Spotify for Developers." *Spotify*, n.d.,
<https://developer.spotify.com/>. Accessed 12 Nov. 2024.
- [7] LINE Corporation. "Flex Message Simulator." *LINE Developers*, n.d., <https://developers.line.biz/flex-simulator/>. Accessed 12 Nov. 2024.
- [8] LINE Corporation. "LINE Business ID." *LINE*, n.d.,
<https://manager.line.biz/>. Accessed 12 Nov. 2024.
- [9] Microsoft .Microsoft learn ,n.d.,<https://learn.microsoft.com/zh-tw/training/azure/>. Accessed 14 Nov. 2024.
- [10] Introduction to Azure Machine Learning, n.d.,
<https://learn.microsoft.com/en-us/training/modules/intro-to-azure-ml/> . Accessed 14 Nov. 2024.