|  |  |
| --- | --- |
| 大專新進人員訓練工作報告  　姓　　名： 文豐澤 單位：　台化工務部自控處  　報告類別：□輪班訓練　 ■基層實務工作訓練第 (三) 次報告  　訓練部門：　台化工務部自控處智能專案組  　起迄日期：　2023/03/07~2023/06/07    　工作項目：1. 台醋營運動態管理平台SQL資料庫  2. ARO2廠去庚烷塔優化人機介面建置  　報告項目：1.台醋營運動態管理平台SQL資料庫  2.鋼構最優化設計AI模組開發 | |
| 評 核 ( 評 語 ) | |
| (２) 一 級 主 管 | (１) 二 級 主 管 |
|  | 1.已完成AI學校技術領袖班訓練，  AI程式設計能力需再精進。  2.經由「台醋營運動態管理平台SQL  資料庫」，學習SQL程式設計，已  有初步了解，程式能力需再加強。  職: |
| (４) 經 營 主 管 | (３) 經 理 室 |
|  |  |

表號：P0002703 規格：A4

註：一.報告內容應包括訓練期間之 1.工作心得感想 2.所得之工作觀

念及精神 3.自我檢討 4.建議意見。

二.訓練部門主管應詳細評核「訓練工作報告」內容，並批註意見。

**目錄**

[一、前言 1](#_Toc137539198)

[二、台醋營運動態管理平台SQL資料庫 1](#_Toc137539199)

[2-1 系統架構 1](#_Toc137539200)

[2-2 資料庫介紹 1](#_Toc137539201)

[2-3 資料轉移程式設計 1](#_Toc137539202)

[三、鋼構最優化設計AI模組開發 1](#_Toc137539203)

[3-1專題介紹 1](#_Toc137539204)

[3-2資料處理 1](#_Toc137539205)

[3-3資料探索分析 1](#_Toc137539206)

[3-4建立模型與評估 1](#_Toc137539207)

[3-5決策支援 1](#_Toc137539208)

[四、工作心得 1](#_Toc137539209)

1. 前言

在現代的工作環境中，我們每天面臨著各種挑戰和機會。科技的迅猛發展、全球化的影響、組織架構的演變以及市場競爭的加劇，都使得工作變得複雜而多變。因此為了不被時代所淘汰，不論是個人還是組織，都需要持續學習、成長和創新，以因應這個複雜多變的世界。

職這段時間主要的工作項目為「ARO2廠去庚烷塔優化人機介面建置」、「台醋營運動態管理平台SQL資料庫」、人工智慧學校上課與其專題「鋼構最優化設計AI模組開發」。因「ARO2廠去庚烷塔優化人機介面建置」案已於上篇報告中詳細說明，此篇報告將只會著重於另外兩項。

首先，職將深入介紹「台醋營運動態管理平台SQL資料庫」的建置。這個專案要求將台醋公司儲存在企業ERP系統中的「銷售統計表」、「外銷受訂清單」、「內銷受訂清單」、「庫存明細表」、「出貨統計表(事業部別)」、「出貨統計表(客戶別)」共6項，透過撰寫的程式，定期進行資料擷取，並對其內容進行篩選，最後再儲存於SQL資料庫中，供台醋營運動態管理平台使用。

其次，職將分享人工智慧學校技術領袖班課程的期末專題「鋼構最優化設計AI模組開發」。這個專題是由營建設計處提出，目標為透過AI學習過去的設計案，在僅提供規模、鋼構位置、鋼構使用種類等基本資訊，快速的建議出新案的鋼構設計。本案共5人進行專題開發，職主要負責一部分的資料前處理與後面的決策支援、限制式與推薦邏輯的撰寫。

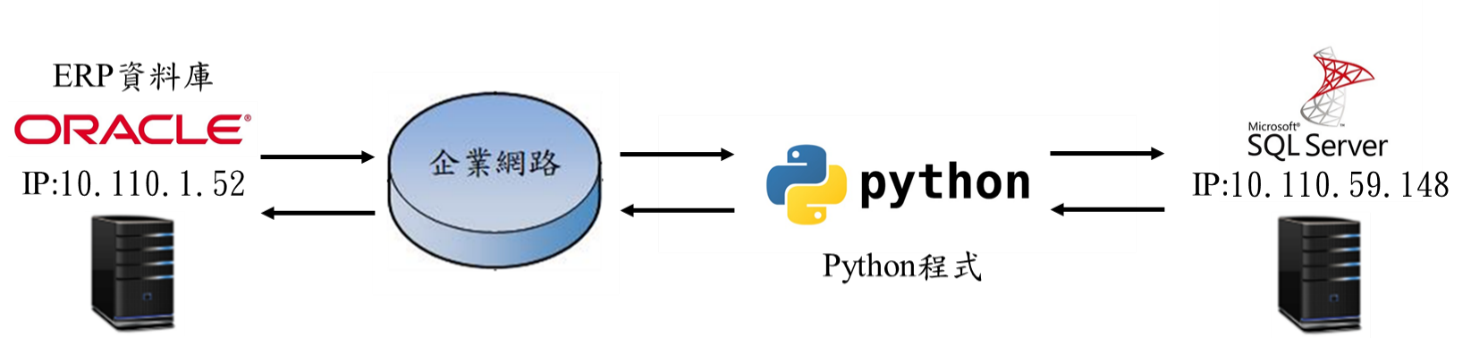
本報告將以這兩個項目為核心，進行說明。

1. 台醋營運動態管理平台SQL資料庫

本案主要對業主所提出的ERP報表資料需求，申請ERP資料讀取權限，再設計程式進行資料下載，並儲存在指定的資料庫。下方將會對本案進行詳細介紹。

2-1 系統架構

本案的系統架構如圖一所示。資料來源為ERP系統的Oracle server。此系統由資訊部負責管理，所以需透過填寫管理制度改善意見反應單，提出所需的資料內容，再由資訊部人員整理彙總成供讀取的資料表形式，讓申請者使用。此舉的目的是為了保護ERP資料庫的一致性，避免需求者對ERP系統的資料進行修改異動，造成非必要的問題。



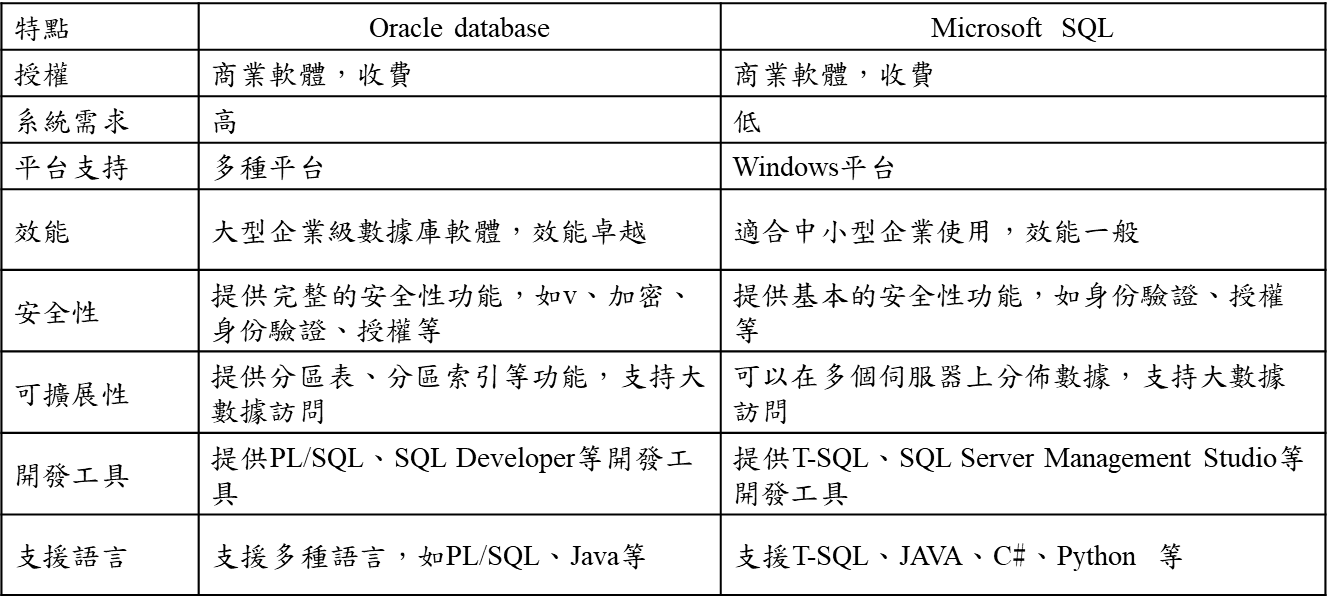
圖一、系統架構圖

使用Python撰寫的資料下載程式，透過企業網路將ERP資料庫裡的資料下載下來，並轉存至台醋營運動態管理平台SQL資料庫。後續台醋設計營運管理平台時，可直接讀取此資料庫的資料。

2-2 資料庫介紹

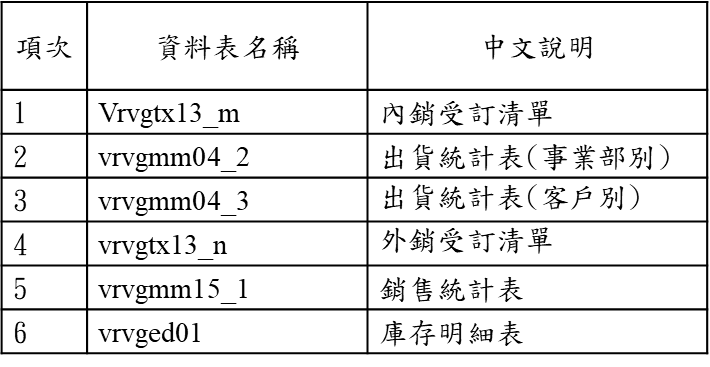
Oracle Database和Microsoft SQL Server是兩個廣泛使用的關聯式資料庫管理系統（RDBMS）。雖然它們都提供類似的功能，但在某些方面有一些明顯的差異。以下是Oracle Database和Microsoft SQL Server之間的幾個比較要點：

1. 平台支援： Oracle Database是一個跨平台的資料庫系統，可以在多個操作系統上運行，包括Windows、Linux、UNIX等。Microsoft SQL Server主要運行在Microsoft Windows作業系統上。
2. 功能和特性： Oracle Database和Microsoft SQL Server在功能和特性方面都非常強大。Oracle Database提供了廣泛的高級功能，如分區表、物化視圖等。Microsoft SQL Server則提供了許多企業級功能，如數據倉庫、數據同步、分析服務、整合服務等。
3. 可伸縮性和性能： Oracle Database被廣泛認為在大型企業環境中具有優秀的可伸縮性和性能。它可以處理大量的數據和用戶同時訪問的需求。Microsoft SQL Server在小型和中型企業環境中表現出色，但在處理極端大型數據集和高併發負載時，可能需要更多的配置和調整。
4. 成本和授權模式： Oracle Database通常被認為是較昂貴的選擇，特別是對於大型企業和高端用戶。它採用了複雜的授權模式，需要支付許可費用和維護費用。相比之下，Microsoft SQL Server在價格上更具競爭力，特別是對於中小型企業和那些已經使用Microsoft產品的用戶。



表一、Oracle database 與 Microsoft SQL簡易比較表

本案在MSSQL的HAC資料庫中，建置了對應需求的六個資料表，資料表名稱為「Vrvgtx13\_m」、「vrvgmm04\_2」、「vrvgmm04\_3」、「vrvgtx13\_n」、「vrvgmm15\_1」、「vrvged01」分別對應「內銷受訂清單」、「出貨統計表(事業部別)」、「出貨統計表(客戶別)」、「外銷受訂清單」、「銷受統計表」、「庫存明細表」共6項(如表二)。



表二、資料表對應表

資料表的設計，其資料行的名稱、資料型態、允許NULL等設定，則是沿用資訊部原本的資料表設定。本案同時對每個資料行屬性添加描述，因其原本的名稱皆為簡短的英文代號，在描述中為其備註相對應的中文說明，詳細對照表請參考表三。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 項次 | 欄位名稱 | 資料庫名稱 | 項次 | 欄位名稱 | 資料庫名稱 |
| 1 | 事業部 | Div | 10 | 本月交運-外銷金額 | cmexpamt |
| 2 | 資料日 | Dtdat | 11 | 本月交運-合計數量 | cmtotqty |
| 3 | 月份 | Ym | 12 | 本月交運-合計金額 | cmtotamt |
| 4 | 客戶名稱 | CuNo | 13 | 本年累計交運-內銷數量 | yyacdmeqty |
| 5 | 產品中類 | Pdpdm | 14 | 本年累計交運-內銷金額 | yyacdmeamt |
| 6 | 單位 | Un | 15 | 本年累計交運-外銷數量 | yyacexpqty |
| 7 | 本月交運-內銷數量 | Cmdmeqty | 16 | 本年累計交運-外銷金額 | yyacexpamt |
| 8 | 本月交運-內銷金額 | cmdmeamt | 17 | 本年累計交運-合計數量 | yyactotqty |
| 9 | 本月交運-外銷數量 | cmexpqty | 18 | 本年累計交運-合計金額 | yyactotamt |

表三、資料表名稱對應表

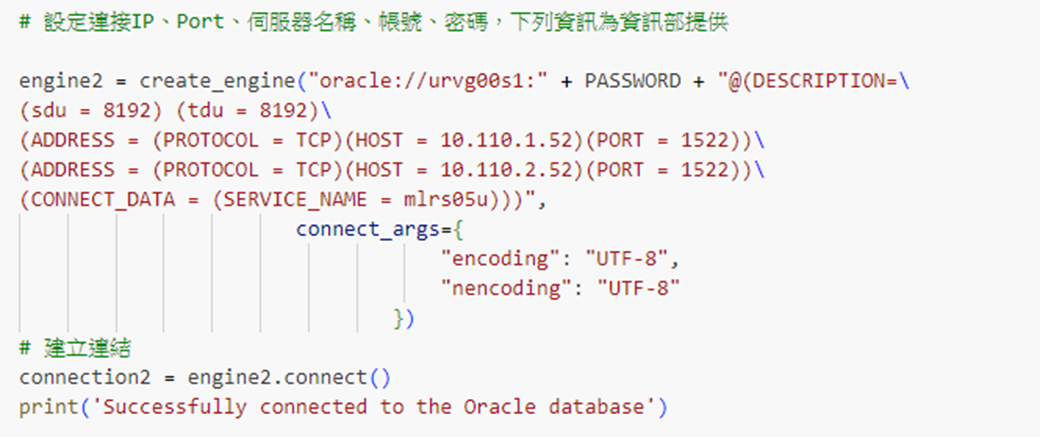
2-3 資料轉移程式設計

本案這次所申請之報表總計6項，分別為營業管理之「銷售統計表」、「外銷受訂清單」、「內銷受訂清單」、「庫存明細表」、「出貨統計表(事業部別)」、「出貨統計表(客戶別)」共6項。

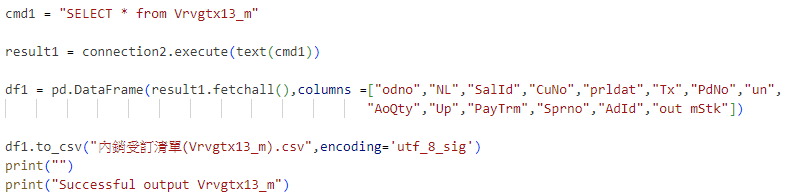
本案資料轉移程式使用Python語言進行設計，運用SQLAlchemy框架連接資料庫。SQLAlchemy 是一個流行的 Python ORM（對象關係映射）工具，它提供了強大的資料庫操作和映射功能，使開發人員能夠使用 Python 程式語言來處理關聯式資料庫。

SQLAlchemy 的主要特點如下：

1. ORM 功能：SQLAlchemy 提供了一個強大的 ORM 層，ORM（對象關係映射）是一種軟體開發技術，用於在應用程式和關聯式資料庫之間建立映射，將物件導向程式語言中的物件和資料庫中的資料表進行對應。開發人員可以定義資料表的結構、關聯和約束，並使用 Python 物件進行資料庫操作，從而簡化了資料庫操作的複雜性。

以下為本案使用此功能的部分程式碼:

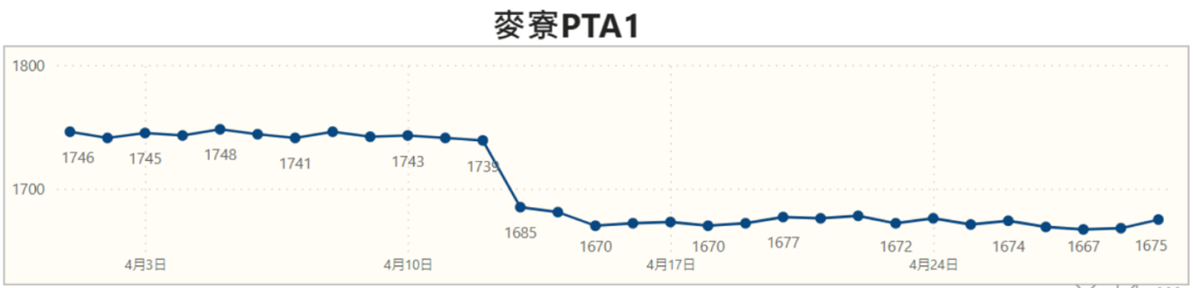
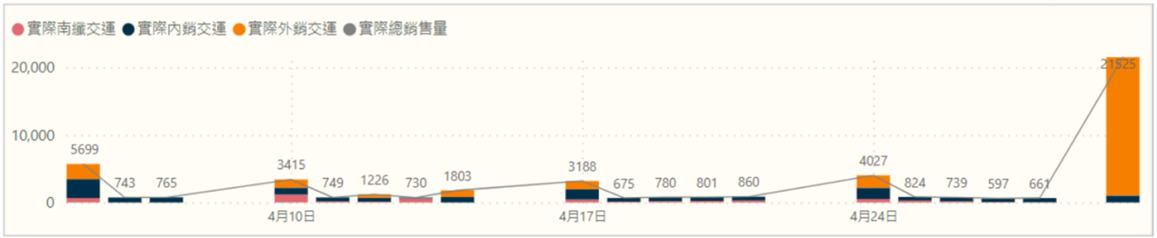
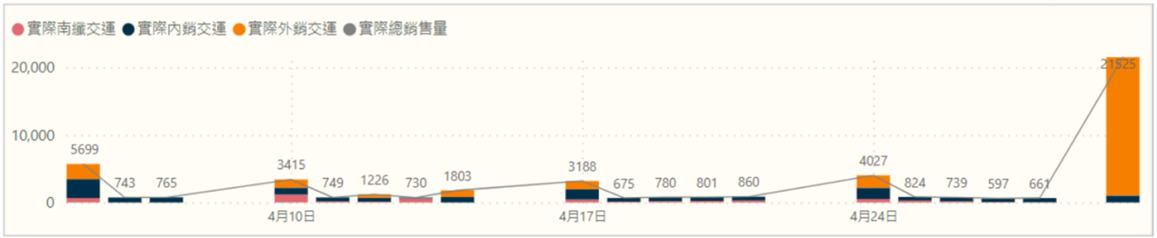
圖二、創建engine之程式碼

上圖的程式碼，為了創建Engine，提供必要之配置詳細資料，例如資料庫URL，其中包括資料庫類型、主機、端口、用戶名、密碼和資料庫名稱。在engine創建完後，就可以運行SELECT查詢並使用fetchall()方法獲取結果(圖三)，或是使用sessionmaker函式創建Session對象，進行SELECT(查詢)、 INSERT(新增)、 UPDATE(更新) 、DELETE(刪除)等操作。

圖三、利用fetchall獲取資料並輸出成CSV檔

1. 跨平台支援：SQLAlchemy 支援多種常見的關聯式資料庫系統，包括 MySQL、PostgreSQL、SQLite 和 Oracle 等。這使得開發人員能夠輕鬆地切換和遷移不同的資料庫系統，同時保持程式碼的相對穩定。
2. 資料庫抽象化：SQLAlchemy 提供了資料庫抽象層，隱藏了不同資料庫系統的細節，使開發人員能夠以相同的方式操作不同的資料庫。這提供了更大的靈活性和可移植性，並簡化了在不同資料庫之間切換的過程。
3. 自動關聯載入和延遲載入：SQLAlchemy 支援自動關聯載入和延遲載入的功能。這意味著當訪問關聯的物件時，它們會自動從資料庫中載入，從而減少了手動處理關聯載入的工作。

轉移程式的主要功能為將ERP系統的資料寫入台醋營運動態管理平台SQL資料庫，因此業主設定程式每日早上九點啟動更新資料庫。比較特別的是六個表單中的「庫存明細表」，此表單在ERP資料庫中本就是即時資料，更新頻率較為頻繁且會覆蓋舊資料，因此需額外寫入一欄「資料庫索取資料時間」，其他五個表單，ERP皆是晚上十二點進行更新，寫入前一天的資料。以上因素造成「庫存明細表」所殘留的資訊十分稀少，並且會被覆蓋掉，現今從ERP擷取出來後，將開始儲存歷史資訊，給營運平台使用。

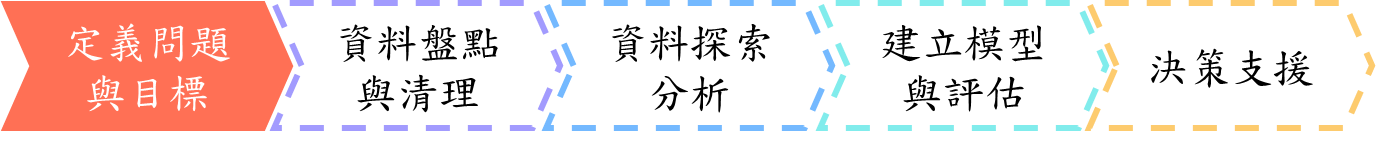
資料彙總到SQL資料庫後，後續台醋可使用Power BI Desktop或Python等程式連接到此資料庫，並依使用者需求進行資料視覺化設計製作成動態圖表，完成資料分析畫面供相關人員使用。下圖四為內銷、外銷等銷售資訊製作成折線圖的示意圖，可透過觀察趨勢變化，進行每日的銷售比較。

圖四、銷售量折線圖(示意圖)

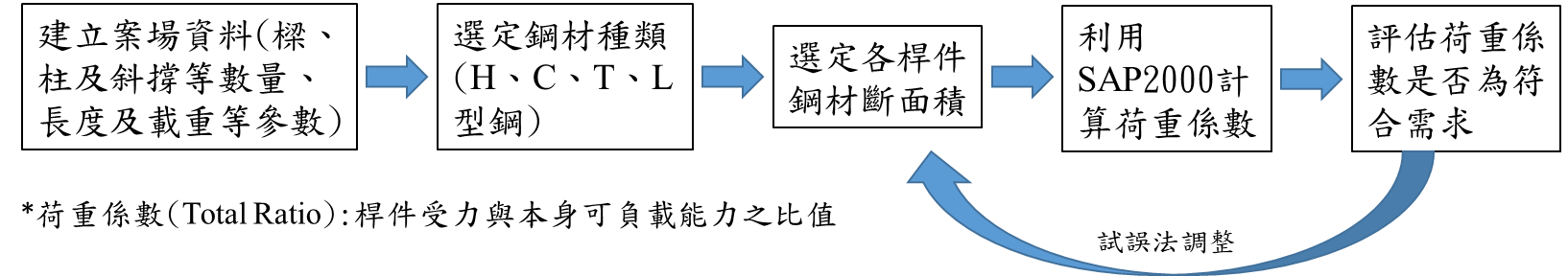
1. 鋼構最優化設計AI模組開發

台灣人工智慧學校為培育具人工智慧技術思維，與實戰經驗的人才，開設技術領袖培訓班。此課程主要聚焦在：「Python與統計」、「資料處理」、「各種機器學習」與「各種神經網路」的原理及應用，最後透過期末專題實作，實際撰寫一AI專題。下方將會介紹本次所製作的專題:「鋼構最優化設計AI模組開發」。

3-1專題介紹

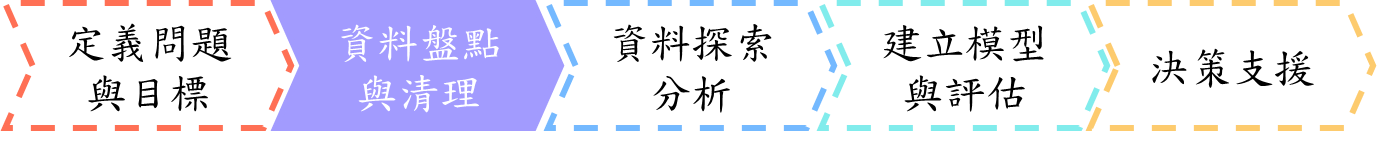


在過往的鋼結構設計中，設計人員需採用試誤方法，搭配自身專業知識與專業分析軟體SAP2000，耗費大量工時尋找最經濟的斷面組合。本專題希望透過建立AI模型，結合專業軟體SAP2000的驗證，節省試誤法所需耗費的時間。圖五為營建設計流程。

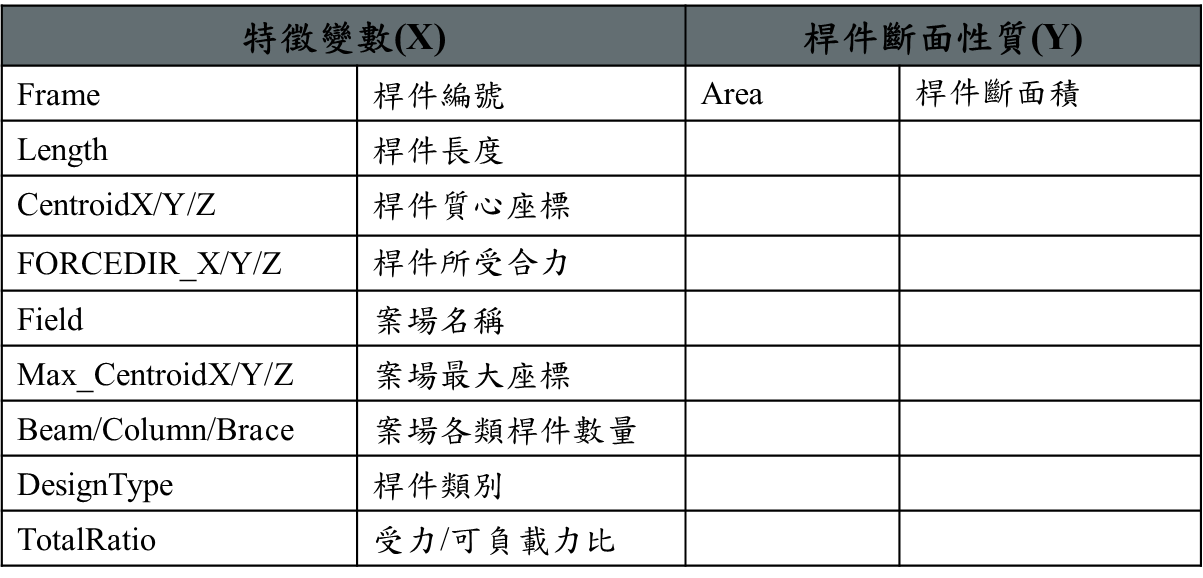
圖五、營建設計流程

本專題為一小組共五人合作進行開發，職在這案中主要負責一部分的資料前處理及後續的決策支援，下方將對整案進行介紹。

**3-2資料處理**

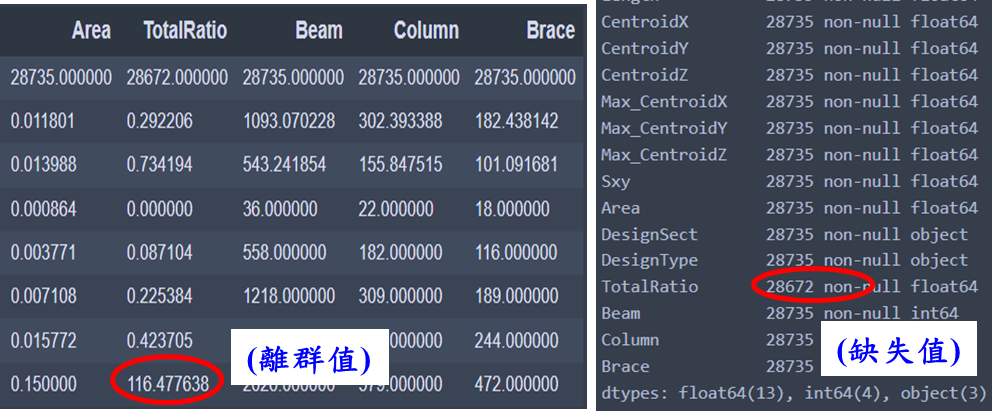


本次資料由營建設計處提供27個案場，分別有14組管架鋼構與13組鋼構廠房，其皆為ARO3廠興建的設計廠房，總體數據有28,735筆，每筆有115個變數。在這些資料中，先與營建設計處進行討論，依靠專業設計人員的經驗，篩選出認為最有影響的變數作為特徵，並對某些變數進行特徵變換的動作，最後取用17個變數作為特徵變數，與一個預測目標(如圖六，圖六已對特徵值進行合併，如:CentroidX/Y/Z)。



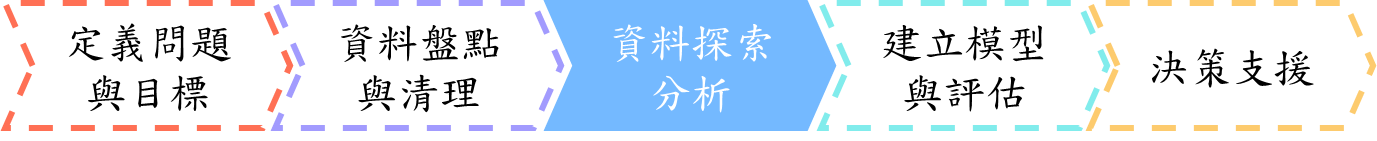
圖六、最終取用特徵值

特徵選擇完畢後，進行資料初步分析。拋棄含有空值的樣本，並將三倍標準差以外離群值進行排除，共排除985筆資料，剩餘27,750筆資料。



圖七、離群值判斷

**3-3資料探索分析**



聽取完營建設計人員的意見後，對17個變數進行視覺化分析，觀察數據的各種表現。

1. 離群值得檢查

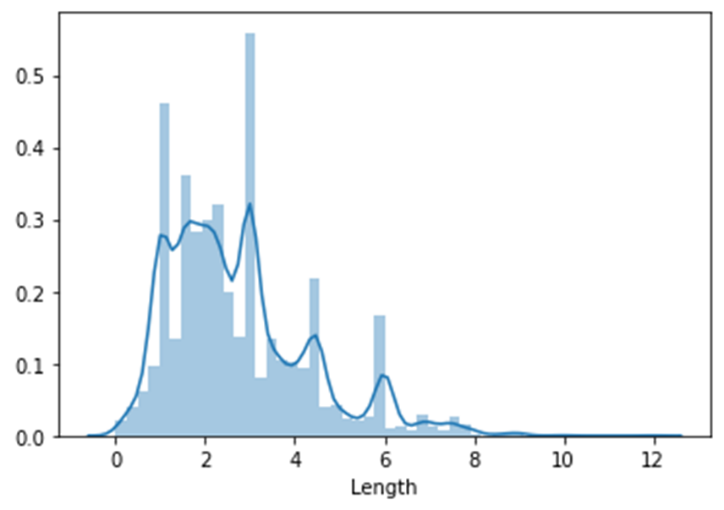
先對離群值進行觀察，以TotalRatio為例。從分佈圖(圖八)可明顯發現，唯有一點落在與其他點相距甚遠的位置。按TotalRatio的定義，值應落在0-1這個範圍內才是合格設計，超出此範圍也不太可能會大於10，而數據顯示此離群值竟來到了116。向營建設計人員確認後，將其視為離群值進行排除。



圖八、TotalRatio離群值視覺化

1. 桿件長度的分佈

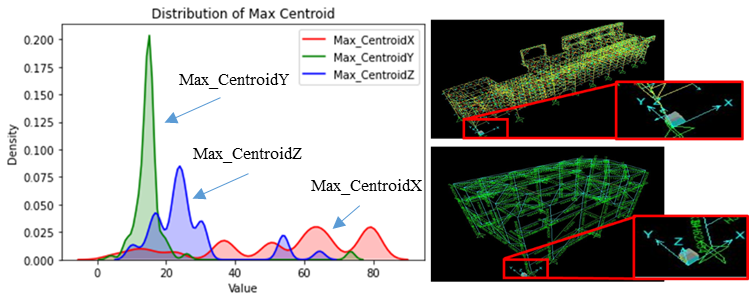
觀察Length桿件長度的長條圖(圖九)，從此圖可發現桿件的常用的長度大致落在1~3米之間，以設計者的角度判斷此為合理的規劃方式。



圖九、桿件長度分佈

1. 案廠規模分析

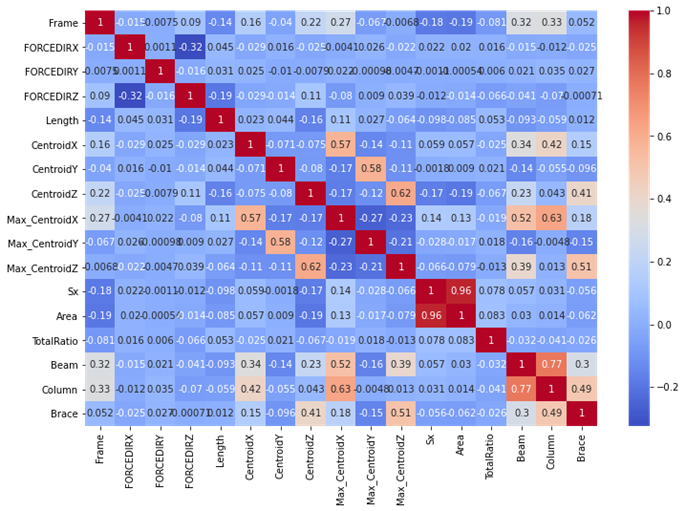
通過視覺化的表現，發現X座標(Max\_CentroidX)的值(Value)其最大值在三個方向所有資料中是最大的，在與營建所提供的設計模型圖進行比較後，發現設計人員通常將X設定為長邊Y為短邊的結論。 (如:圖十)



圖十、案廠規模分析

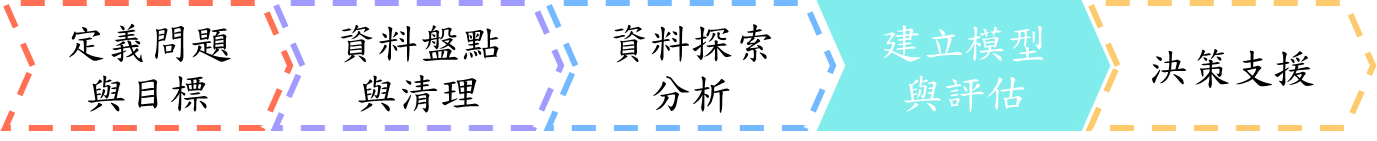
1. 相關係數分析

資料集經前面的數據處理，利用皮爾森相關係數，計算各個特徵值之間的相關性 (圖十一)。取用相關性大於0.4的特徵值，共有9個特徵值符合，與設計人員一開始提供之17個特徵做比較，發現這九個特徵值原本就有被納入原始17個特徵值。經與設計人員討論，另外原本有選擇的8個特徵相關性不高但不代表不重要，而是可能沒有直接相關，因此先不進行排除，經嘗試建模後，根據訓練結果再決定是要排除還是進行其他數據處理。

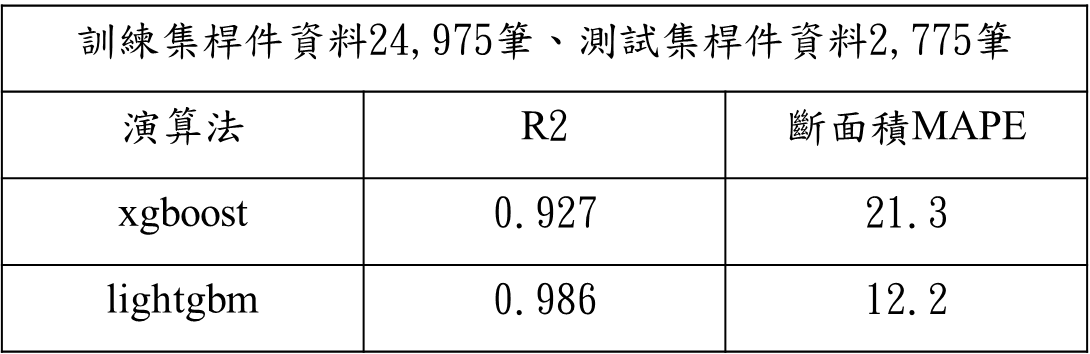


圖十一、相關係數分析

**3-4建立模型與評估**



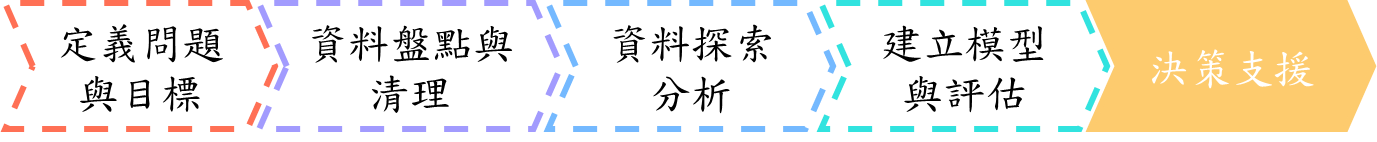
數據集經歷以上的前處理後，剩餘27,750筆資料。將其分為訓練集24,975筆與測試集2,775筆，進行對預測目標「鋼構斷面面積」預測的模組訓練。本案使用XGBoost與LightGBM作為訓練方法，XGBoost通過level-wise分裂策略和正則化技術，在預測性能和模型解釋性之間取得平衡，LightGBM則使用leaf-wise分裂策略和優化技術，快速處理大型數據集。這兩者在各種問題和數據集上都能取得出色的預測結果，因此將它們作為這次的訓練方法，並從中選取各自最好的測試結果做比較，如表四。



表四、測試結果

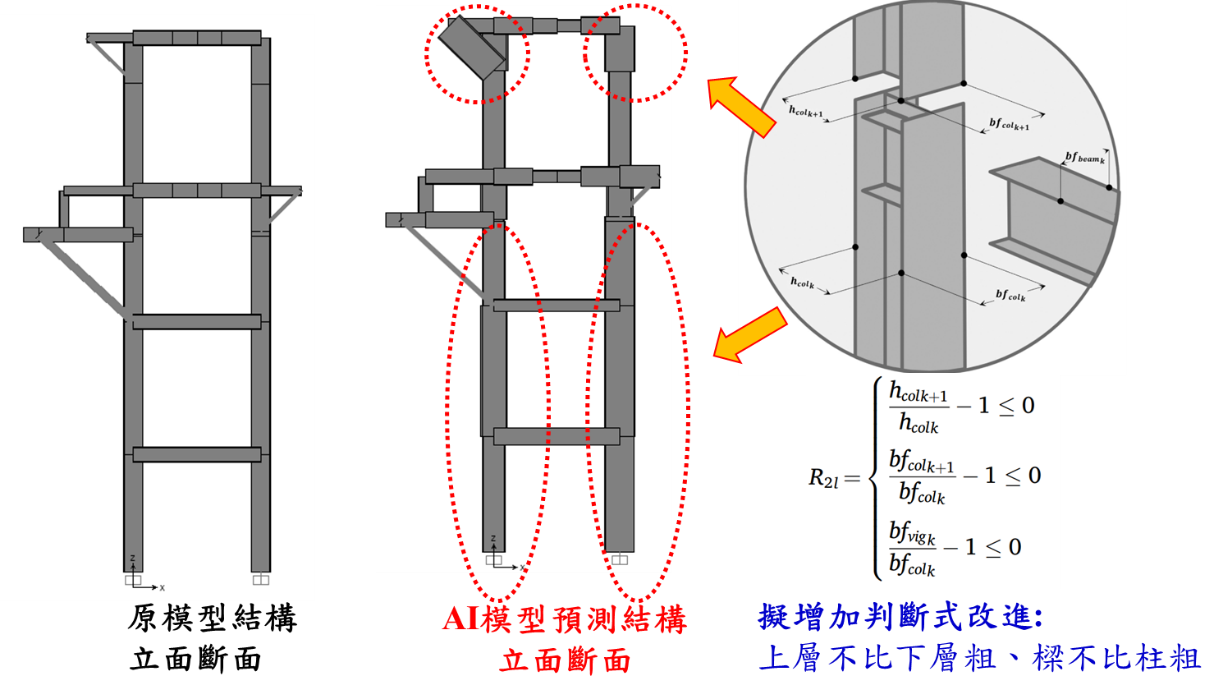
兩者的測試結果進行比較，由lightgbm的測試結果較優勝出，因此使用由lightgbm訓練得出的結果作為預測模型。此處預測出來的是每根鋼構其期望之斷面面積，還需進行限制式的加入與查表動作，將在下方的決策支援進行說明。

**3-5決策支援**



模組訓練完以後，我們發現AI所給出的建議斷面面積對應到現有的鋼構資料庫後，實際設計出來的建築使用的鋼構多變複雜(如圖十二)，鋼構使用種類多達十幾種，這並不符合實際的設計理念。經與營建設計處討論後，決定將不影響主結構應力之結構進行排除，減少不必要的雜訊，並由設計人員提供限制式:

1. 桿件上層不比下層粗、樑不比柱粗。
2. 設計人員一開始提供什麼樣的鋼型，預測完後的選擇也是從相同的鋼型目錄進行選擇。
3. 同層的柱使用相同的斷面，樑、斜撐比照辦理。



圖十二、實際與AI預測選擇

除了增加限制式，我們請營建對於可選擇的鋼構進行篩選，只留下常用的選擇，這個步驟使可選擇的鋼構數量從110種刪減至57種。另外對要做預測的案場先行設定每根要使用的鋼構，查表時只從對應的鋼構類別進行選擇，這使得可選擇對象進一步的縮小(例如:這根柱要選擇H型鋼，那AI選擇時也只會從H型鋼的類別中進行挑選)。同時對鋼構的選擇與查表建立相對應的流程:

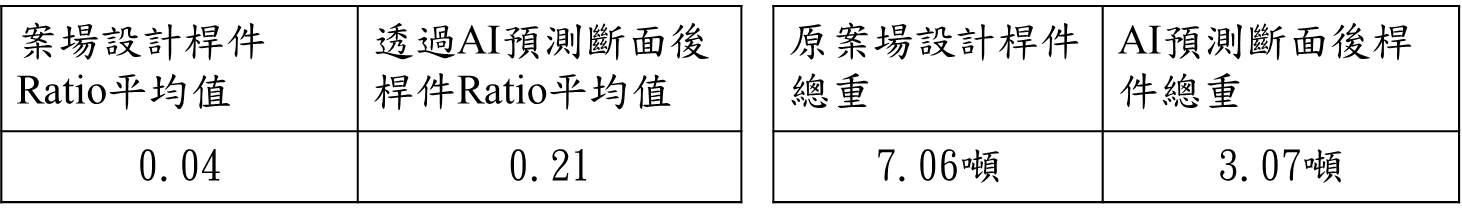
1.從要預測的案件中，利用桿件節點座標抓取各桿件Z方向最大高度，定義出分層方式。

2.接著先處理柱，最低層的柱統一選擇由模組預測的最小鋼構。其餘層數與自身下面一層做比較，若長/寬度大於下層，則選用下面那層的鋼構；若符合限制條件，則使用同層最小的鋼構。

3. 再來處理樑，定義方式為寬度不可大於柱的寬度，若樑寬大於柱寬，則改選用下層使用過的樑

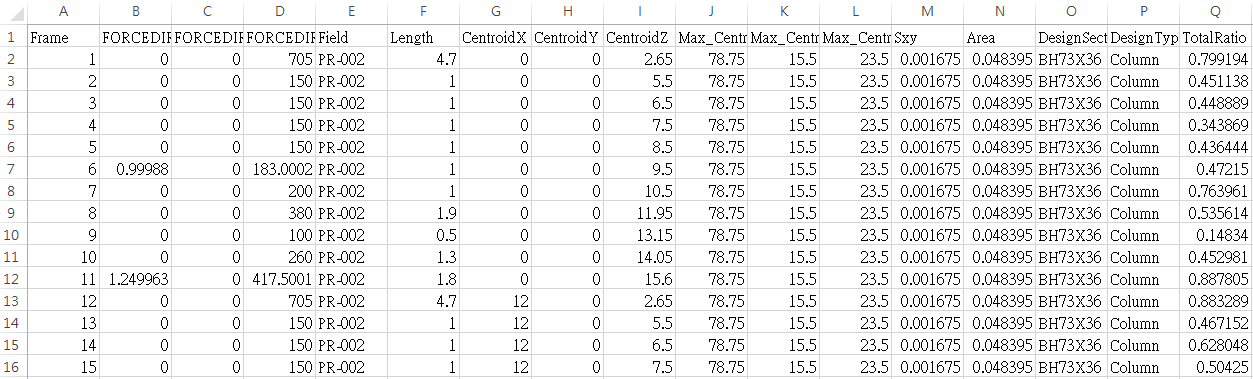
4. 最後處理斜撐:統一使用預測後最小的鋼構

經過以上條件，AI模擬結果有了顯著的改善。首先鋼構不再粗細交雜不一，所使用的鋼構種類也減少至合理範圍；再來經專業分析軟體SAP2000的分析，預測出來的結構皆符合各式法規與受力等條件；最後與原本設計案的設計相比，其總重也成功減少，實現鋼構選用上更為經濟且符合規範的目標。(表五呈現原始設計與經AI預測之差異)

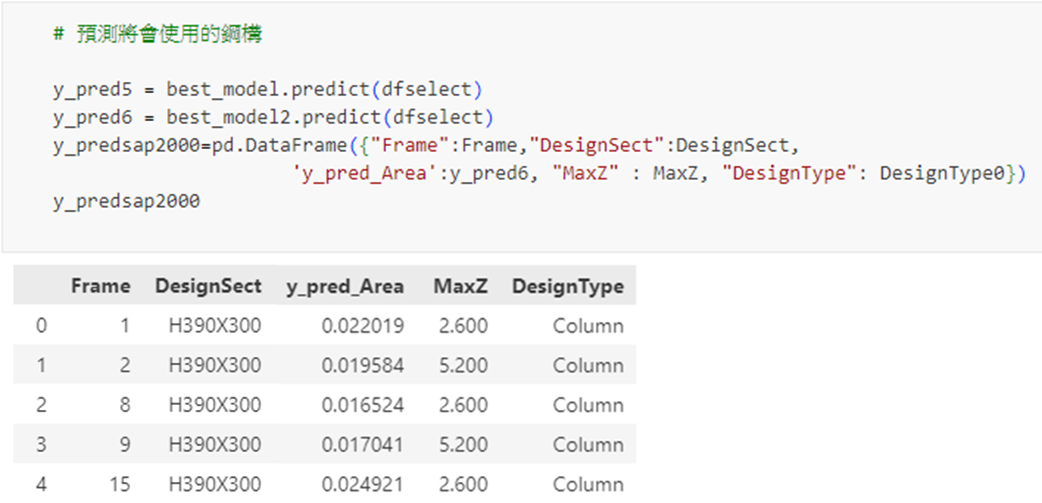


表五、再評估結果

本專題並未建置圖形化介面，所有資料的處理及預測都是直接運行程式碼，下列介紹鋼構設計優化的操作流程。首先，營建設計人員需將要設計的案場進行初步的規劃，包括桿件的放置位置、桿件使用的鋼型、總桿件的數量等等，並利用SAP2000分析這份設計圖導出一份EXCEL分析報表。根據此報表，將預測模型會使用的17個特徵值資料取出，整合至一個新的CSV檔中(圖十三)。

圖十三、特徵值整合

將這份整合過的檔案，配合訓練完的模型，對每根桿件進行斷面面積預測，並將預測結果與桿件基本資訊進行配對。圖十四為其程式碼與預測出來的成果，Frame : 指的是桿件編號；DesignSect : 原本設計使用的鋼型； y\_pred\_Area : 預測出來的斷面面積；MaxZ : 桿件高度；DesignType : 桿件類型。

圖十四、斷面面積預測結果

斷面面積預測結果出來後，開始對其進行限制式與其他條件的添加。在開始之前，先自訂一函數split\_designsect，此函數透過正則表達式對鋼構名稱進行分解，將名稱拆解為:鋼型、H(高度)、B(寬度)、翼板厚度(t1、t2)五項。



圖十五、split\_designsect程式碼

將整個鋼構類別拆成三個部分:「柱」、「樑」、「斜撐」。



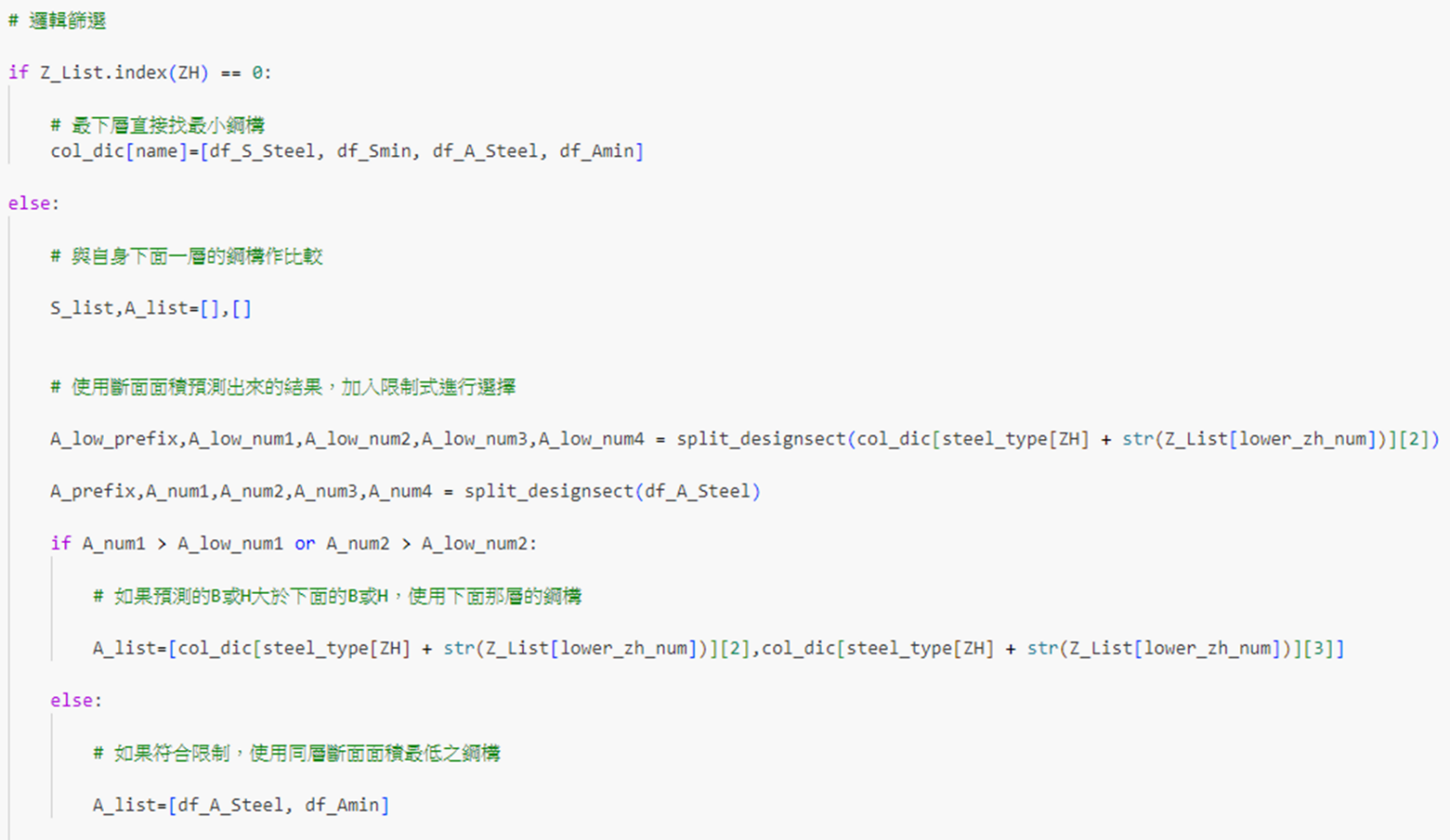
圖十六、將資料拆分三個類別

首先對「柱」進行挑選，因為所有限制式皆會被柱的選擇所影響。在「柱」這個類別中，依據高度再次進行分類，從高度最低的開始處理。同高度下，找出預測出來的最小斷面面積對應之鋼構並套用給同層所有柱使用，如果此層非最底層，必須再將下面那層使用的柱與這層預計要使用的柱進行比較。



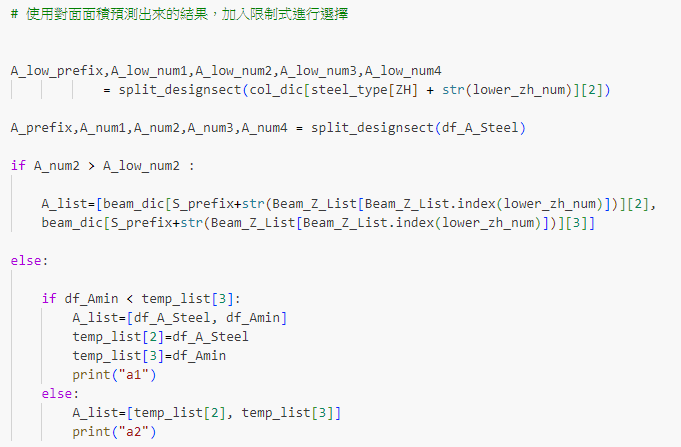
圖十七、鋼構選擇

比較方式為使用split\_designsect分解兩者的鋼構名稱，上層柱的H(高度)與B(寬度)要皆小於等於下層柱的H(高度)與B(寬度)，換句話說就是上層柱不能比下層柱粗，如果不符合，則沿用下層所使用的柱。



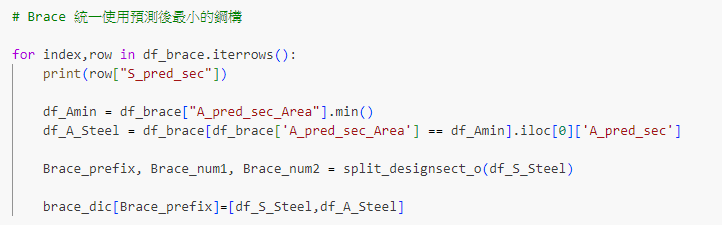
圖十八、柱限制式條件

接著處理「樑」，與柱一樣分層進行處理，不同的地方在於限制條件，需與連接樑的下方柱，比較B(寬度)，樑的B(寬度)不能大於柱的B(寬度)。



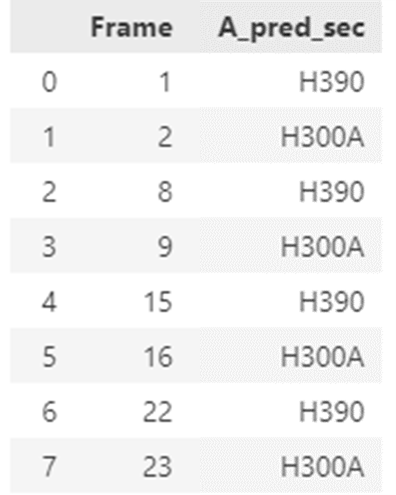
圖十九、樑限制式條件

最後「斜撐」統一使用斷面面積最小之鋼型，如果有其他需求，再由設計人員自行替換。



圖二十、斜撐的選擇

將上述整理完的資訊進行彙整，會得到一桿件編號對應一個要使用的鋼型，如圖二十一。



圖二十一、鋼型預測結果

為了方便導入SAP2000進行驗證，將鋼型預測結果以CSV檔輸出，依照桿件編號替換原始設計檔案的鋼構名稱，這樣就不需人為在SAP2000的圖形化介面對桿件逐一進行替換，加快驗證速度。

現今的預測對象只限定在較為簡單的結構，後續工作將導入基因演算法與透過API連結SAP2000進行更多組合的配對，讓此AI能夠有更精準的預測與預測更為複雜的結構物。

1. 工作心得

「台醋營運動態管理平台SQL資料庫」這個案子，對我來說是一項非常重要的經驗。因為資料來源資料庫屬於資訊部管理，因此必須要提前進行申請。申請的過程中，有許多需要注意的地方，申請表的格式是否符合、業主的需求是否都有填寫、向業主確認資料的索取時間範圍、參考過去的案子思考業主可能漏掉什麼或是可能需要什麼等等。光是為了向資訊部申請資料，就來來回回與業主溝通好幾回合。實際送出申請單後，還需經歷漫長的簽核流程，這個流程大約耗時約一個多禮拜。當資訊部正式開始這項委託，時間已經過去約一個月了。如果沒有同事的經驗與警告，直到開案才開始申請作業，一定造成時間上的壓迫。這使我理解到，有些事情需事先進行作業，保留一定的工作時間彈性，避免如有緊急事件，案子將無法如期完成。

從2/11開始至5/27止，參加台灣人工智慧學校技術領袖培訓全域班第二期課程。通過期末專題，我學習到如何配合團隊撰寫AI程式，程式的流程、變數的設定、功能的規劃、程式的結合等等，都是這次專題曾經碰到過的絆腳石。除此之外也學習到如何與業主溝通互相的需求，業主不一定熟悉AI領域，我們也不保證能理解業主的專業，因此頻繁的溝通以及討論成了非常重要的一件事，透過長期的磨合，尋找各自的需求，才是有效完成案件的關鍵。

於6/1開始的「寧波熱電廠線上燃料煤熱值預測」案，將把人工智慧學校課程中所學的知識應用於其中，期許AI模型設計的品質提升與產出優良的預測結果。