資料分析技術、機器學習演算法與應用

**課程內容：**

* 資料分析技術
  + 使用Python with Jupyter Notebook來建立BI報表
  + 使用套件
    - Numpy
      * <https://numpy.org/>
    - Pandas
      * <https://pandas.pydata.org/>
    - Matplotlib
      * <https://matplotlib.org/>
    - Seaborn
      * <https://seaborn.pydata.org/>
* 機器學習演算法與應用
  + 使用Python來建立機器學習的預測模型
  + 使用套件
    - scikit-learn
      * <https://scikit-learn.org/>

**參考書目：**

* 資料分析與視覺化
  + 教科書
    - Python 資料科學學習手冊
      * https://www.tenlong.com.tw/products/9786263246843
      * 電子版書籍
      * https://github.com/jakevdp/PythonDataScienceHandbook
  + 參考文件
    - Python 資料可視化之美
      * https://www.tenlong.com.tw/products/9789865501457
    - https://matplotlib.org/stable/tutorials/index
    - https://seaborn.pydata.org/tutorial.html
    - https://pandas.pydata.org/docs/user\_guide/index.html
* 機器學習原理與實作
  + 教科書
    - 精通機器學習 Hands-On Machine Learning
      * https://www.tenlong.com.tw/products/9781098125974
      * https://www.tenlong.com.tw/products/9786263246676
  + 參考文件
    - https://scikit-learn.org/stable/user\_guide.html

**個人專題：**

* 專題目標：使用BI工具來分析資料並建立報表
* 個人獨力完成
* 可以使用Tableau或是Python
* 資料來源：教師提供、自行準備、Kaggle上收集
* 2/4 個人專題(專題題目確認、問題解答)
* 2/14 個人專題報告

**團隊專題：**

* 專題目標：分析並建立機器學習的預測模型
* 5~6人一組
* 可以使用Vertex AI 或是 Python
* 資料來源：教師提供、自行準備、Kaggle上收集
* 專題題目需要教師核可
* 2/24(半日) 團隊專題主題確認
* 3/8,22 戴佳樺老師帶團隊專題
* 3/10(半日) 團隊專題進度檢查
* 3/27 團隊專題報告

**專題注意事項：**

* 資料必須是表格式資料 (能夠用一張EXCEL表格來儲存)
* 不要找時序性資料、股票預測資料
* 可以到 <https://www.kaggle.com/> 去找資料
  + <https://www.kaggle.com/datasets/toramky/automobile-dataset>
  + <https://www.kaggle.com/datasets/janiobachmann/bank-marketing-dataset>
  + <https://www.kaggle.com/competitions/bike-sharing-demand>
  + <https://www.kaggle.com/datasets/HRAnalyticRepository/absenteeism-dataset>
  + <https://www.kaggle.com/c/GiveMeSomeCredit>
  + <https://www.kaggle.com/datasets/redwankarimsony/heart-disease-data>
  + <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>
  + <https://www.kaggle.com/datasets/yasserh/housing-prices-dataset>
  + <https://www.kaggle.com/datasets/kaggle/us-baby-names>
  + PS: 最好自己找資料，自己找的資料分數會比較高。如果真的找不到再使用上面的資料。

**成績：**

* 個人專題
* 團體專題
* 資料分析技術LAB實驗作業
* 機器學習原理與實作LAB實驗作業
* 課堂抽問 (加分)